Exploratory Representations for Geographic Information Retrieved from the Internet

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde (Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Awase Khirni Syed

aus

Republic of India

Promotionskomitee

Prof.Dr. Robert Weibel (Vorsitz)

Dr. Ross Stuart Purves

Prof.Dr. Menno-Jan Kraak (ITC, Enschede, NL)

Zürich 2008

To

My Family

Summary

Among the many information sources, the internet has evolved as the one that is most widely used. Much of the information on the internet involves some geographical context. Extracting this geographical context from internet documents has therefore drawn interest from geographic information science and computer science to exploit this potential resource. This information seeking and extraction process, however, is made made harder by the fact that in contrast to conventional databases, internet documents are essentially unstructured, incomplete, heterogeneous and dynamic in nature.

Existing search engines or information retrieval systems are only partially capable of extracting geographical context in the same manner as other content. Conventional search engines are ill equipped to interpret geographic prepositions expressed by the user's such as "Hotels in Hyderabad". The geographic preposition "in" would generally be treated as a stop word and discarded from the search. Likewise, most of the existing and conventional geographic information systems do not use the geographic information available on the World Wide Web.

The thesis establishes the need for a geographic search engine as a key element of a semantic web, since geography is perhaps one of the most familiar and common means by which we express our information requests. We start off with a comprehensive framework for the extraction of geographic information, investigate various exploratory representations of geographic information retrieved from the internet and finally draw empirical measures to evaluate the effectiveness and efficiency of these representations in supporting the exploratory information seeking process.

The thesis starts by motivating this work within the evolution of the internet and the World Wide Web as a source of geographical information, reviewing some of the recent efforts towards geographical information retrieval, as well as the limitations of current web search engines. Subsequently we introduce the basics of information retrieval, information seeking models, visualization techniques, interaction models, and cognition. This is followed by an empirical study to analyse the requirements for constructing a spatially aware information seeking system, using existing search engines and web mapping services widely used today. We then go on to discuss the visualization options used in the design of the multi modal user interface of the prototype GIR system developed by researchers of the EU/FP5 project SPIRIT, which this thesis project was part of. The following chapter then presents an overview of the SPIRIT prototype.

In an effort to reinforce theoretical foundations discussed in the earlier half of the thesis, an empirical study was conducted to evaluate the exploratory representations adopted in the SPIRIT prototype. This prototype supports the extraction and visual representation of geographic information from web documents. While the emphasis of these representations has been on spatialization of geo-referentiable information retrieved from the internet, the most elementary display of the results is in the form of a textual listing, supported by relevant footprints (i.e. geographic extent) of the documents displayed on a map, which is referred to as map based representation. Documents are simply represented by their footprint's centroid, though they may have widely varying spatial extents (e.g. the footprint of Switzerland vs. the footprint of Zurich). Footprints have evolved as a medium of cataloguing the geographic extent of the information available on the web relating to a specific location. Besides the elementary map-based representation other visualization techniques have been

adopted in the SPIRIT prototype, including density surface representation, footprint based filtering and cartogram based representation. Apart from these, two other mock-up representations namely spatial acitivity map and isoline based representations have been proposed. Each of these aforementioned techniques serve varying functionality.

The empirical study with 50 test subjects revealed that the density surface representation is a widely accepted visual representation in comparison with other visual representations. Such inferences are supported by a stronger response value favouring density surface representation. This is due to their inherent characteristics that they present a comprehensive summary of the spread of the documents over a defined geographic region. Footprint based filtering, on the other hand, is the least understood and users tend to get confused most. However, there has been significant increase in approval in the users' response when provided with adequate training. This significant shift has been accounted by the fact that users lacked adequate understanding of the concepts relating to 'geographic extent of a footprint', while after training it has been easier to identify, locate, rank and understand the underlying relationships between footprints and documents. This is due to the ability to selectively prune through the results based on the geographic extent and the number of footprints. The learning effect was less pronounced for the cartogram representation.

As another finding, the mock-up designs for activity maps and for isoline representations received very positive judgments by the test users, despite the fact that by definition mock-ups are not functional. This finding seems to suggest that mock-up representations may not be sufficient to evaluate the effectiveness of visualization designs in an unbiased manner, though this is often done in cognitive experiments.

In summary, while none of the representations that were tested in this work seem to be capable of covering the full gamut of requirements of exploratory visualization of geographic information retrieval, it could be shown that they are in many ways complementary to the simple map based representations that are used in current geographic search engines such as Google Maps. Hence, they offer interesting alternatives for extending the current state of the art, hence improving the exploration and findability of geographic information on the internet.

Zusammenfassung

Unter den vielen Informationsquellen, die uns heute zur Verfügung stehen, hat sich das Internet (bzw. World Wide Web) in den letzten Jahren zur am weitesten verwendeten entwickelt. Ein grosser Teil der Information auf dem Internet involviert dabei einen geographischen Kontext. Die Nutzung dieser Informationressource durch Extraktion des geographischen Kontexts aus Internetdokumenten hat deshalb das Forschungsinteresse aus der geographischen Informationswissenschaft und der Informatik auf sich gezogen. Dieser Prozess der Informationssuche und -extraktion wird allerdings erschwert durch den Umstand, dass Internetdokumente im Gegensatz zu konventionellen Datenbanken typischerweise unstrukturierter, unvollständiger, heterogener und dynamischer Natur sind.

Existierende Suchmaschinen oder Information Retrieval Systeme sind nur ansatzweise in der Lage, geographischen Kontext in der gleichen Art zu extrahieren wie sie dies für andere Inhalte tun können. Konventionelle Suchmaschinen sind schlecht ausgerüstet, um geographische Präpositionen in Suchanfragen wie “Hotels in Hyderabad” zu interpretieren. Die geographische Präposition “in” beispielsweise wird allgemein als sog. Stopword behandelt und ignoriert. Ebenso nutzen die meisten existierenden Geographischen Informationssysteme die geographische Information auf dem World Wide Web nicht.

Diese Arbeit geht von der Annahme aus, dass eine geographische Suchmaschine ein Schlüsselelement des Semantischen Webs ist, da Geographie wohl eine der vertrautesten und allgemeinsten Arten ist, wie wir unsere Informationsanfragen ausdrücken. In dieser Arbeit entwickeln wir ein umfassendes Konzept der Extraktion geographischer Information, untersuchen verschiedene explorative Darstellungen für aus dem Internet extrahierte geographische Information und evaluieren danach durch empirische Untersuchungen die Wirksamkeit und die Effizienz, mit der diese Repräsentationen den explorativen Prozess der Informationssuche unterstützen können.

Die Dissertation startet in der Einleitung damit, diese Forschung zu motivieren durch die Evolution des Internets und des World Wide Web als Quelle geographischer Information und indem jüngere Bemühungen für Geographic Information Retrieval (GIR) sowie die Einschränkungen heutiger Web-Suchmaschinen diskutiert werden. Danach führen wir in die Grundlagen von Information Retrieval, Modellen der Informationssuche, Visualisierungsverfahren, Interaktionsmodellen und Kognition ein. Dies wird gefolgt von einem Bericht über eine empirische Studie, in der eine Analyse der Anforderungen an ein System für die räumlich bewusste Informationssuche durchgeführt wurde; die Studie bediente sich heute existierender Suchmaschinen und Diensten für Web Mapping. Wir fahren danach weiter mit einer Diskussion verschiedener Visualisierungsoptionen, die im Design der multimodalen Benutzerschnittstelle eines Prototypsystems für GIR berücksichtigt wurden, das von ForscherInnen des EU/FP5 Projekts SPIRIT entwickelt wurden. Das folgende Kapitel gibt sodann einen Überblick über den SPIRIT-Prototypen.

In der Absicht, die theoretischen Grundlagen und Annahmen des ersten Teils dieser Dissertation zu erhärten, wurde eine empirische Studie durchgeführt, um die explorativen Repräsentationen des SPIRIT-Prototypen zu evaluieren. Dieser Prototyp unterstützt die Extraktion und visuelle Repräsentation von geographischer Information aus Webdokumenten. Während der Akzent dieser Repräsentationen auf der räumlichen Visualisierung (Spatialization) georeferenzierbarer textueller

Information aus dem Internet liegt, ist die einfachste Darstellungsform für die Resultate in Form einer Textliste, unterstützt durch eine Darstellung der Footprints (d.h. räumliche Ausdehnung) der relevanten Dokumente auf einer Karte. Diese Darstellungsform wird kartenbasierte Repräsentation genannt. Die Dokumente werden dabei vereinfachend über den Schwerpunkt ihres Footprints abgebildet, obwohl die räumlichen Ausdehnungen stark unterschiedlich sein können (z.B. Ausdehnung der Schweiz vs. Ausdehnung der Stadt Zürich). Footprints haben sich zu einem Mittel der Katalogisierung des geographischen Bezugs der Information entwickelt, die auf dem Web für eine spezifische Lokalität vorhanden ist. Neben der Grundform der kartenbasierten Repräsentation wurden auch andere Visualisierungstechniken im SPIRIT-Prototypen implementiert, darunter Dichteoberflächen, Footprint-basierte Filterung und Kartogramme. Zusätzlich zu diesen wurden zwei weitere Visualisierungsformen vorgeschlagen und getestet, die allerdings nur als Mock-ups (Maquetten) realisiert wurden, sog. Spatial Activity Maps und Isolinien-Repräsentationen. Jede der oben erwähnten Visualisierungstechniken dient unterschiedlichen Funktionen.

Die empirische Studie mit 50 ProbandInnen zeigte, dass Dichteoberflächen breite Zustimmung finden im Vergleich mit anderen Repräsentationen. Diesen Schluss legen die klar besseren Zustimmungswerte für Dichteoberflächen nahe. Dies ist auf die inhärente Eigenschaft zurückzuführen, dass Dichteoberflächen einen guten Überblick über die Verteilung der Dokumente einer geographischen Region bieten. Andererseits ist die Footprint-basierte Filterung die am wenigsten gut verstandene Repräsentation, welche die Benutzer am meisten verwirrte. Allerdings war eine signifikante Zunahme der Zustimmung der Benutzer festzustellen, wenn diese eine adäquate Einführung in die Funktionsweise dieser Repräsentationsform erhalten hatten. Diese signifikante Veränderung konnte an der Tatsache festgemacht werden, dass die Benutzer das Konzept der 'geographischen Ausdehnung eines Footprints' ohne Training nicht verstehen konnten, während es für sie nach erfolgtem Training einfacher war, Beziehungen zwischen Footprints und Dokumenten zu identifizieren, zu lokalisieren, zu rangieren und zu verstehen. Dies wird möglich, indem mit dieser Darstellungsform die resultierenden Webdokumente selektiv gefiltert und ausgewählt werden können entsprechend ihrer geographischen Ausdehnung und der Zahl der Footprints. Der Lerneffekt war für die Kartogrammdarstellung weniger ausgeprägt.

Als weiteres Resultat dieser Untersuchung ergab sich, dass die Mock-up Designs für die Spatial Activity Maps und für die Isoliniendarstellungen sehr positive Bewertungen durch die BenutzerInnen erhielten, obwohl solche Mock-ups definitionsgemäss gar nicht funktional sind. Diese Beobachtung scheint darauf hinzuweisen, dass Mock-up Repräsentationen nicht genügend sein könnten, um die Wirksamkeit von Visualisierungsdesigns in einer unverfälschten Weise zu evaluieren, obwohl dies oft so gemacht wird in kognitiven Experimenten.

Zusammenfassend kann gesagt werden, dass zwar keine der untersuchten Repräsentationen in der Lage zu sein scheint, die volle Bandbreite der Anforderungen der explorativen Visualisierung im Geographic Information Retrieval abzudecken, dass aber auch gezeigt werden konnte, dass diese in mannigfaltiger Weise komplementär sind zu den elementaren kartenbasierten Repräsentationen, die in heutigen geographischen Suchmaschinen wie Google Maps verwendet werden. Sie offerieren so interessante Alternativen, um den Stand der Forschung zu erweitern und damit die Exploration und Auffindbarkeit von geographischer Information im Internet zu verbessern.

Acknowledgements

This thesis was made possible by the help and support of a number of people. I would like to start by thanking the Swiss Federal Office of Education and Science for funding this research under the EU-IST 5th Framework Programme (within the project SPIRIT, IST-2001-35047) and particularly Prof. Robert Weibel for hosting me at the GIS Division, Department of Geography, University of Zurich.

My deepest gratitude to my supervisors Prof. Dr. Robert Weibel and Dr. Ross S. Purves who have been of great help, and their scientific guidance and review have tremendously improved the quality of the work presented here.

I would like to express my gratitude to the external member of my PhD committee, Prof. Dr. Menno-Jan Kraak of the ITC, Enschede (NL), as well the external reviewer of the thesis, Prof. Dr. Monika Sester of the University of Hannover, who have made every effort to give in their constructive arguments for enhancing the thesis.

I would also like to acknowledge the contributions of the SPIRIT consortium who have contributed in more ways than one to the results presented in this thesis. All the other staff in the department helped me in many ways. My special thanks to Ms. Elisabeth Cottier who has made my life easier in dealing with the administrative processes.

To my colleagues at work, who have constantly encouraged me and have been patient to review and enhance my work, especially Dr. Bisheng Yang, Daniel Wirz, Dr. Marco Hugentobler, Dr. Nikos Koutsias, Dr. David Caduff, Dr. Urs-Jakob Rüetschi, Dr. Charis Keller, Dr. Corinna Heye.

I would like to acknowledge the help from Prof. Dr. Sara Fabrikant, University of Zurich; Dr. Daniela Petrelli, University of Sheffield; Dr. Bénédicte Bucher, Dr. Anne Ruas, and Dr. Sandrine Balley, IGN France; and Prof. Dr. Werner Stahel at the Statistics Seminar, ETH Zurich.

My thanks for the patience and effort of all the participants in experimental studies who have contributed to this.

The support of the Library staff at the University of Zurich and ETH Zurich for giving me access to scientific resources inside and outside University of Zurich. Many thanks to the IT team for providing me with the adequate computing resources.

My sincere thanks to all the people who have contributed either directly or indirectly and who have strongly supported and motivated me to pursue my research objectives.

Last but not the least to Prof. Dr. Faisal Irshad Khan, Prof. Dr. Rajiv Gupta, Dr. K. R. V. "Raja" Subramanian, Dr. J. P. Mishra, Dr. B. R. Natarajan, Deshmukh, Leeladharesh, Dr. Rajkumar Raikar who have constantly motivated me and have guided me during my career to take the right decisions. They have constantly stood by me and have always been there. My gratitude to Prof. Dr. S. Sadagopan, Prof. Dr. S.S. Prabhu and IIIT, Bangalore for supporting me during the review process since September 21, 2006 by providing me a place to work in Bangalore, India.

In particular i am indebted to my parents, Syed Ameese Sadath and my wife Shagufta for their support, motivation and financially supporting me since September 2006.

Contents

Figures

Tables

1. Introduction 29

1.1 Search Engines 29

1.2 Motivation 31

1.3 Problem Statement 35

1.3.1 Research Questions 36

1.4 Organization of the thesis 36

1.5 Contributions to the SPIRIT project 38

2. Theoretical Background 39

2.1 Information 39

2.2 Information Seeking 40

2.2.1 Definitions of information seeking 40

2.2.2 Models of information seeking 41

2.2.3 Cognitive styles in information seeking 43

2.2.4 Relevance to ongoing research 44

2.3 Visualization 44

2.3.1 Definition of Visualization 44

2.3.2 Geovisualization 45

2.3.3 Visual Exploration 46

2.3.3.1 Principles of visual information seeking systems 49

2.3.3.2 Interaction/Interactivity 50

2.3.3.2.1 Interaction with data 52

2.3.3.2.2 Interaction with data representation 56

2.3.3.2.3 Interaction with temporal dimension 61

2.3.3.2.4 Contextualizing interactions 64

2.4 Taxonomy of user tasks for interaction 67

2.5 Visualizing text 68

2.5.1 Visualizing text document collections 68

2.5.2 Visualizing information hierarchies 71

2.6 Role of Cognition 75

2.6.1 Cognition and spatial cognition 75

2.6.2 Spatial awareness 76

2.7 Limitations of current research 78
2.7.1 General limitations in information seeking 78
2.7.2 Specific limitations in exploratory visual representation 79

3. Spatially aware information seeking (Requirements Gathering) 81
3.1 Introduction 81
3.2 Scenario development 82
3.2.1 Scenario analysis 89
3.3 Storyboards and simulated work tasks 92
3.3.1 Storyboards 92
3.3.2 Simulated work tasks 92
3.3.3 Environment and subjects 93
3.3.4 Analysis methods 94
3.3.5 Findings 95
3.4 Hierarchical task analysis 104
3.5 Defined requirements 105
3.6 Development of an information seeking model 107
3.7 Summary 109

4. Visualization approaches 111
4.1 Delivering background web mapping 112
4.1.1 Degree OpenGIS web servers 112
4.1.2 Data for background mapping 113
4.2 Web interface constraints 113
4.3 Ideation- Concepts for visualizing spatially aware information retrieved from the internet. 114
4.3.1 Query formulations 115
4.3.2 Query disambiguation 117
4.3.3 Query expansion 118
4.3.4 Result presentation 119
4.3.4.1 Map based representation 119
4.3.4.2 Ontology based traversal of the query space 120
4.3.4.3 Cartograms 120
4.3.4.4 Target Metaphor 121
4.3.4.5 Touchgraph and ThemeRiver Visualization of Spatial Relations 122
4.3.4.6 Statistical Approach 122
4.3.4.7 Density Surface representation 123
4.3.4.8 Footprint based representation 123
4.3.4.9 Spatial activity map 124
4.3.5 Search refinement 125
4.3.6 Feature Control 125
4.4 Characteristics of the results and the visualization requirements. 126

4.5 Summary 126

5. Prototype 129
5.1 Functional components 129
5.1.1 Pre-processing 130
5.1.1.1 Assigning spatial extents to web documents 130
5.1.1.2 Building Document Indexes 132
5.1.2 Run-time 135
5.1.2.1 User interface 135
5.1.2.2 Query disambiguation 135
5.1.2.3 Query expansion 136
5.1.3 Geographic ontology 138
5.1.4 Search engine 139
5.1.5 Ranking 140
5.1.6 Metadata 142
5.1.7 Results display 143
5.1.8 Broker component 147
5.2 Summary 147

6. Alternative visualizations 149
6.1 Density surface representations 150
6.1.1 Method 150
6.2 Footprint based filtering 152
6.2.1 Method 152
6.3 Cartograms 154
6.3.1 Method 155
6.4 Spatialy activity maps/plots 156
6.4.1 Symbolic integration 160
6.4.2 Clutter reduction 160
6.5 Isoline representations 160
6.5.1 Method 161
6.6 Summary 162

7. Evaluation Methodology 167
7.1 Underlying theories 167
7.2 Materials and Testing Environment 169
7.3 Focus Groups 169
7.4 Procedure 170
7.4.1 Questionnaire design 171
7.4.2 Evaluation Phase 173
7.5 Defining user tasks 173
7.6 Empirical criteria 179

7.7 Summary 180

8. Results 181
8.1 Preliminaries 181
8.1.1 Missing values 181
8.1.2 Participants profiles and sessions 182
8.1.3 Time taken to complete tasks 183
8.1.4 Self-learning exercise 183
8.2 Visualization Methods of SPIRIT Prototype 184
8.2.1 Repeated measures ANOVA 184
8.2.2 Simple Main Effects 185
8.2.2.1 Within Subjects 192
8.2.2.2 Between Subjects 197
8.3 Mock-up Visualization 216
8.3.1 Spatial Activity Maps/Plots 216
8.3.2 Isoline representation 218
8.3.3 Comparison of mean response for working systems and mock-ups 218
8.4 Combined relevance vs perceived relevance 221
8.5 Observations 222
8.5.1 Background & Expertise 222
8.5.2 Map Interactions 223
8.5.3 Respondents' Wishlist 223
8.5.4 Arguments and criticisms 224
8.5.5 Suggested improvement to existing SPIRIT system 225
8.6 Summary 225

9. Discussion 227
9.1 Visualization methods used 227
9.1.1 Outcome of the empirical studies 228
9.1.2 Preferred representations 232
9.2 Information seeking process 233
9.3 Cognition and conceptualization of geographic space 233
9.4 Mock-up representations 234
9.5 Impact of level of granularity and context on visualization of search results 234
9.6 Combined relevance vs. perceived relevance 235
9.7 Learnability of the SPIRIT prototype 235
9.8 Limitation of the experimental design 235
9.9 Summary 236

10. Conclusions 237
10.1 Insights 237
10.2 Limitations 240

10.3 Outlook 242
10.4 Closing remarks 243

References 245

Appendix A - Publications 297

Appendix B - Curriculum vitae 299

Appendix C - Internet Statistics 300

Appendix D - Early stage requirements gathering excercises 302

Appendix E - Snapshots of various interfaces widely used 310

Appendix F - Geographic data sources (Purves, 2003) 318

Appendix G - Experiment design 321

Appendix H - Factorial design to impart uniform learning effect 345

Appendix I - Real needs expressed by participants during evalution of visual representations developed for SPIRIT 346

Appendix J - SPIRIT prototype evaluation Questionnaire (Benedicte Bucher) 349

Appendix K - Additional background information: User preferences 355

Figures

1.1 Taxonomy of Web Metrics (Dhayani *et al.*, 2002. p.471, Fig.1) 33

2.1 Visual variables and their perceptual properties (from MacEachren, 1995, p. 279) 47

2.2 The range of functions in an idealized research sequence. (DiBiase, 1990; http://www.geovista.psu.edu/publications/others/dibiase90/swoopy.html) 49

2.3 Goals of use arrayed in the map use cube (MacEachren & Kraak, 1997; p.338) 50

2.4 A unified taxonomic framework for information visualization (Pfitzner *et al.*, 2001; p. 64) 51

2.5 Classification of information visualization techniques (Keim *et al.*; p. 26 in Dykes *et al.*, 2005) 52

2.6 A taxonomy of visualization operations (Goodell *et al.*, 2006) 53

2.7 An example of brushing and linking. When the user moves over the map or histogram the corresponding element is highlighted (Neumann, 2003; http://www.karto.ethz.ch/neumann/cartography/vienna) 55

2.8 Dumpster – An interactive online visualization that depicts a slice through the romantic lives of American teenagers. (Levin *et al.*, 2006; http://artport.whitney.org/commissions/thedumpster/) 56

2.9 Namevoyager (Wattenberg, 2005; http://www.babynamewizard.com/namevoyager/Inv0105.html) 57

2.10 An example of non-interactive highlighting (Corum, 2006; http://style.org/judgealito) 58

2.11 An illustation of simple filtering and highlighting by layer choice. Filtering information is achieved by by toggling on/off the layers that are not of interest to the user (source: http://www.carto.net/papers/svg/tuerlersee/) 58

2.12 Taxonomy of presentation for large data spaces (Leung and Apperley, 1994; p. 127) 61

2.13 U.S Federal Election (http://www.nytimes.com/packages/html/politics/2004_ELECTIONRESULTS_GRAPHIC/) 62

2.14 The population of Europe (country codes according to ISO 3611) (van Kreveld and Speckmann, 2004) 62

2.15 Geographical distribution of news stories. Newman and Gastner show (http://www.sciencenews.org/articles/20040828/bob8.asp) 63

2.16 % of votes by county cast for the leading candidate (Corum, 2004. http://www.style.org/mappingvotes/) 63

2.17 State delegate equivalents and the 2004 IOWA caucuses. Individual results per county are scaled in a checker board pattern. (Corum, 2004. http://www.style.org/iowacaucus/) 63

2.18 An illustration of America's richest – Hometowns of the richest 400 (http://www.forbes.com/2005/09/15/hometowns-networths-america-richest_05rich400_map.html) 65

2.19 Topix.net Conversation Map for the contiguous states of the USA (http://www.topix.net/blogs) 66

2.20 Percentages of votes cast by county for Schwarzenegger vs. Bustamante in the 2003 election for Governor of California [Source: http://www.style.org/mappingvotes/] 67

2.21 US Bureau of the Census, Tiger Map Server: An exemplar illustrating the combination of data layers. The maps illustrate the overlay of different data layers superimposed on the map in order to infer the results of interest for the user. (Source: http://tiger.census.gov/cgi-bin/mapbrowse-tbl) 69

2.22 Taxonomy of user tasks for interaction with multidimensional visualization system (redrawn from Pillat *et al.*, 2005) 71

2.23 State of the union parsing tool (http://www.style.org/stateoftheunion/parse/) 71

2.24 Galaxies visualization (Crow *et al.*, 1994) 72

2.25 Themescape visualization (Wise *et al.*, 1995) 72

2.26 A typical newsmap using the Themescape technique (developed by Cartia, Inc. http://www.mundi.net/maps/maps_015/#ref_2) 72

2.27 Treemap: a space filling approach to visualization of hierarchical information structures (Johnson & Shneiderman, 1991) 73

2.28 The Cone-Tree: a conical representation of hierarchy (Robertson *et al.*, 1991, 1994) 73

2.29 Hyperbolic Browser (Lamping & Rao, 1996) 74

2.30 CHEOPS: A compact explorer for complex hierarchies (http://www.icg.tu-graz.ac.at/~hofer/mmis97/thh/TreeMaps/Cheops.htm) 74

2.31 RINGS: Ringed interactive navigation graph system (http://www.cs.ucdavis.edu/~ma/papers/graph2.pdf) 75

2.32 Information slices: expanding the swing/com/sun/java/swing subdirectory into the right-hand disc (http://www.iicm.edu/liberation/iicm_papers/ivis98.pdf) 75

2.33 Touchgraph visualization (http://www.touchgraph.com/TGGoogleBrowser.html) 76

2.34 Textscape (Rossi, 1999; Fig. 2) 77

3.1 Example Scenario obtained in terms of region delinations of the Black Forest (Source: SPIRIT deliverable D3 7101) 91

3.2 Dimension of need of SPIRIT user (Balley, 2002 p.11 Fig 1) 95

3.3 A snap shot of the Starting Point , results presentations of geo-web service: SomewhereNear (http://www.somewherenear.com) (clockwise) 97

3.4 A snap shot of the Starting Point, results presentations of geo-web service: Maponus (http://www.maponus.com)(clockwise) 99

3.5 A snapshot of the Starting Point and result presentation of geo-web service: Mapquest (http://www.mapquest.com) (clockwise) 100

3.6 A snap shot of Starting point and result presentation of geo-web service: Multimap (http://www.multimap.com) 101

3.7 A snap shot of Starting point and result presentation of geo-web service: Map24 (http://www.map24.com) 101

3.8 An episode of Spatially Aware Information Seeking process influenced by the works of Bates (1986), Taylor(1968), Mizzaro(1998), Pirrolli(2000), Saracevic(1996). The model illustrates “an episode” based on berry picking model. 110

4.1 Free text querying: A design template(Syed *et al.*,2002) 117

4.2 Structured query interface: A design template(Syed *et al.*,2002) 118

4.3 Graphical query interface using web mapping. : A design template(Syed *et al.*,2002) 119

4.4 Map based presentation of the results: A design template (Syed *et al.*,2002) 120

4.5 Ontology based traversal of query space: A design template (Syed *et al.*,2002) 121

4.6 Results presentation using cartograms- note that the centre map is used to illustrate the concept, and that in implementation a topographic map would be displayed here: A design template (Syed *et al.*, 2002) 122

4.7 Visualization of query results using a target metaphor: A design template (Syed *et al.*, 2002) 123

4.8 Interface design for ThemeRiver visualization for results presentation (Syed *et al.*, 2002). 124

4.9 Query visualization using a statistical approach to illustrate spatial relevance vs distance from the centroid of expressed region of interest. Themes illustrate the semantic similarity of the documents. 125

4.10 An example of geographic references on web page referring to various geographic locations around the world along with addresses, telephone numbers and postal codes. Snap shot of Glen More lodge web site (source: http://www.glenmorelodge.org.uk/) 126

5.1 SPIRIT architecture showing run-time and pre-processing components and linkages (from Purves *et al.*, 2007, Figure 1). 132

5.2 Spatial Indexing illustrating fixed grid scheme (Vaid *et al.*, 2005) 134

5.3 Diagram of run-time information flow (Syed *et al.*, 2003; Fig. 14) 136

5.4 Workflow UML diagrams showing system interactions during search. a) For structured text query and b) for polygon or point based search (Purves and Yang, 2005; p. 5) 136

5.5 Structured query interface (also called Simple query interface) – starting point 137

5.6 User preference setup interface illustrating different preferences: Query disambiguation, Query expansion, Spatial indexing, Ranking method, and Number of Results to be presented. 138

5.7 Query Specification by Polygon – In this diagram a polygon is used to specify the area in which the search will be made, for example: Looking for castles within some distance of Cardiff (Purves *et al.*, 2005). 139

5.8 Place name disambiguation interface 140

5.9 Role of the geographic ontology 141

5.10 Plot of BM25 (textual) against spatial document scores and associated ranking for a distributed method (van Kreveld *et al.*, 2004, Arampatzis *et al.*, 2006) 144

5.11 Results presentation of the structured query interface. Each of the elements of this interface are described in pink boxes (Purves *et al.*, 2005). 145

5.12 Brushing and linking results for multiple documents with identical footprints (Purves *et al.*, 2005). 146

5.13 The graphical interface for the presentation of results for a query such as "Castles in Scotland". The region of Scotland was selected by the user by defining the search polygon. Highlighting of document footprints that share the same location is achieved by drawing a 'spider' graph of circles (in red). 147

5.14 Background mapping for the United Kingdom (Purves *et al.*, 2005). 148

5.15 Tele Atlas Data set illustrating the map background for Edinburgh at the highest possible resolution (Purves *et al.*, 2005). 149

6.1 An example of a density surface representation for "castles" in a user specified region of Scotland, United Kingdom 154

6.2 An illustration of footprint based filtering technique for a query "castles in Scotland", highlighting footprints within a document illustrated as document footprint (candidate region) 155

6.3 An illustration of footprint based filtering technique demonstrating interactive filtering for a user's selection of a few documents for a specific footprint (geographic extent); for the query "castles in Scotland". 156

6.4 An illustration of cartogram based representation for a user's selection of a specific region for the query "castles in Scotland". 157

6.5 An illustration of spatial activity plot (a mock-up representation using real datasets) depicting spatial hierarchy on the Y-axis and distance from the centroid of expressed region of interest (candidate region) for a query "resorts in Scotland, United Kingdom". 158

6.6 An illustration of spatial activity map using real datasets with symbols depicting the contextual details of each document for a query "resorts in Scotland, United Kingdom". 159

6.7 An example of spatial hierarchy and level of detail for the United Kingdom obtained from SABE dataset. a) Region (level-1) b). Province (Scotland- level 2) c). County (level-3), d). Ward (level-4). 160

6.8 A graphical illustration of spatial hierarchy of the Seamless Administrative Boundaries of Europe data sets demonstrating the political administrative district hierarchies. 160

6.9 An illustration of isoline representation: a mock-up visualization using real data sets obtained from the SPIRIT engine. 163

7.1 Evaluation framework 172

8.1 Interaction plot depicting the visualization profiles against question type and user response (GLM mean response) 188

8.2 Interaction plot depicting the question type against visualization methods and user response (GLM mean response) for each empirical measure 189

8.3 Interaction plot for mean response between visualization tasks for unsupervised session 202

8.4 Interaction plot for mean response between visualization tasks for supervised session 203

8.5 Interaction plot for mean responses between visualization tasks for different scenarios (Mountaneering in Scotland and Hotels near Portree) 210

8.6 Interaction plot for mean response between visualization tasks for comparing map, density, footprint and cartogram representations for a query cottages in “western Scotland”. 212

8.7 Interaction plot for mean response between visualization tasks comparing map, density , footprint and cartogram representations for a query “biking south of london”. 213

8.8 Interaction plot comparing working system (map, density, footprint and cartogram based representations) with mock-up visualization (spatial acitivity map and isoline representation) based on common criteria 221

8.9 Interaction plot comparing working system (map, density, footprint and cartogram) with mock-up visualization (spatial acitivity map and isoline representation) based on common criteria. 222

Tables

1.1 Statistics of geo-queries in 2001 Excite log (Sanderson & Kohler, 2004) 35

1.2 Contribution by the author for the SPIRIT project 40

2.1 Forms of information seeking (Bates, 1986) 42

2.2 Modes of Scanning (Choo *et al.*, 2000; Fig. 1) 43

2.3 Information seeking behaviours and web moves (Choo *et al.*, 2000; Fig.2) 44

2.4 A preliminary typology of interactivity (Crampton, 2002; p. 90) 54

2.5 Cartogram classification (Dent, 1998) 62

2.6 Variables of the temporal dimension in animated representations expressed in concepts introduced by DiBiase *et al.*(1992) and MacEachren (1994); Blok (2005, p. 64, Table 4.1) 64

2.7 Summary of rule and areas of major impact on utility of multiple views (Baldanado *et al.*, 2000; p. 112, Table 1) 68

3.1 Model for user-centered IR interaction and interface design (based on the model by Allen, 1996, p.24) (Hansen, 1997, p.8, Table 2) 84

3.2 Summary of the scenarios generated by focus groups (Balley *et al.*, 2002, p.8) 85

3.3 Key elements of scenarios (Balley *et al.*, 2002, p.14) 92

3.4 Conceptual correspondences between mathematical space and experiential space (Couclelis, 1998, Table 2, p.215) 93

3.5 Elements of the information landscape and associated informational functions. (Couclelis, 1998) 93

3.6 Chosen use-case scenarios for the assessment of early stage requirements of existing search engines and geowebservices (Balley *et al.*, 2002) 95

3.7 Representative users from various domains recruited for Wizard of Oz studies (Bucher *et al.*, 2004 p.12 Table 2) 96

3.8 Types of data collected, data collection methods and analysis methods. 97

3.9 Geoweb services feature comparison 98

3.10 Comparison of results presentation of selected search engines. 101

3.11 User's expertise and knowledge with respect to usage of computers [Q: As a computer user, do you consider yourself as?] 102

3.12 Search at a work place [Q: How often do you search for information for work purposes?] 102

3.13 Language preference and search [Q: How often do you use another language than your native one?] 102

3.14 Search expertise [Q: As a searcher do you consider yourself as?] 103

3.15 Search Engine usage [Q: Which web search engines do you use?] 103

3.16 Information sources [Q: What are different information sources the users are aware of , for the scenarios presented and would choose?] 103

3.17 Search strategies and styles [Q: If you do not get the result you expect, you normally:?] 103

3.18 Geo-web services [Q: Have you ever used geographical information services on the web?] 104

3.19 The context of usage of the geo-web services illustrated in Table 3.16 [if you have used one of the services above, in which context?] 104

3.20 Preferred starting points of interface [Q: What are preferred starting points in the interface?] 104

3.21 Keywords identified by the users during the early stage requirements assessment of existing search engines and geoweb services (Simulated Work Tasks) (ref. Table 3.6) 105

3.22 User Requirements List (Balley *et al.*, 2002) 107

3.23 Broadening of SPIRIT Requirements (Bucher *et al.,* 2004 p.18-20) 108

4.1 Task Analysis for the SPIRIT interface (Syed *et al.*, 2002) 116

4.2 Analysis of Patterns in Search Refinement Process (Lau and Horvitz, 1998) 127

4.3 Visualization requirements and the characteristics of the results (Yang *et al.*, 2006; Purves and Yang, 2005) 128

5.1 Summary of footprints (UIDs) identified by geoparsing for the SPIRIT collection (from Purves *et al.*, 2007; Table 1) 133

5.2 Spatial Indexing Schemes (Vaid *et al.*, 2005) 135

5.3 Requirements for the SPIRIT ontology (Fu *et al.*, 2003, p.4) 140

5.4 Formulae used for footprint similarity score and geographic prepositions used in the query (van Kreveld *et al.*, 2004) 143

5.5 Methods adopted to combine spatial and BM25 scores (van Kreveld *et al.*, 2005) 144

5.6 Scales for mapping in different countries (nominal resolution 72 dpi) (Purves and Yang, 2005) 147

6.1 First and second order effects (O' Sullivana nd Unwin, 2003) 152

6.2 Justification for the use of density estimation for collapsed geographic extents (footprint centroids) point data sets obtained from the SPIRIT engine. 154

6.3 Justification for the use of footprint based representation demonstrating the spatial extent fo the footprints 156

6.4 Justification for the use of Dorling's Cartogram 157

6.5 Justification for the use of spatial activity maps/plots 161

6.6 Justification for the use of isoline representation 164

6.7 Comparison of strengths and weaknesses of various representations adopted 165

7.1 Borlund's model of evaluating interactive information retrieval (Borlund, 2003) 170

7.2 Task abstraction level and scope (Winckler *et al.,* 2004 p.167) 170

7.3 Description of the participants involved in the design of questionnaire & (I,II) pilot studies. 173

7.4 Real work task scenarios envisioned during the experiment design phase 174

7.5 Comprehensive task structure of the SPIRIT system grouped by components and functionality. Subsequent mapping on to the information and interaction typologies proposed by Crampton (2002) 176

7.6 Empirical criteria adopted for comparing four (4) different visual representations (Map, Density, Footprint, Cartogram) 181

8.1 Participants involved in SPIRIT experimental studies (Refer to CD: Descriptive Statistics) 184

8.2 Self learning phase descriptive statistics for simple structured interface for a query “hotel in Scotland” and “hotels near Glasgow” 185

8.3 Advantages and Disadvantages of repeated measures (Davis, 2002 p.2) 186

8.4 Pair-wise comparison between experimental task * visualization methods using General Linear Model Repeated Measures 189

8.5 Empirical criteria- ease of use vs different visualizations (General Linear Model Repeated Measures-WithinSubjects) 194

8.6 Empirical criteria- information representation and understandability for different visualizations (General Linear Model Repeated Measures-WithinSubjects) 195

8.7 Empirical criteria- legends and annotations as supporting aides vs different visualization. (General Linear Model Repeated Measures-WithinSubjects) 196

8.8 Empirical criteria- Colour usage and appropriateness for different visualizations(General Linear Model Repeated Measures-WithinSubjects) 196

8.9 Empirical criteria- viewpoint or relativity for different visualizations (General Linear Model Repeated Measures-WithinSubjects) 197

8.10 Empirical criteria- locate/associate/findability for different visualizations (General Linear Model Repeated Measures-WithinSubjects) 198

8.11 Empirical criteria- counter check for ease of use and understandability for different visualizations(General Linear Model Repeated Measures-WithinSubjects) 198

8.12 Empirical criteria- relevant task accomplishment for different visualizations.(General Linear Model Repeated Measures-WithinSubjects) 199

8.13 Descriptive statistics of different visualizations to locate geographic scope of documents 200

8.14 Mean responses for ease of use using GLM (Between Subjects) 201

8.15 Mean responses for information presentation and usage with GLM (Between Subjects) 204

8.16 Mean responses for legends as supporting aides using GLM (Between Subjects) 204

8.17 Mean responses for colour usage and appropriateness for different visualization using GLM (Between Subjects) 205

8.18 Mean responses for viewpoint or relativity for different visualization using GLM (Between Subjects) 206

8.19 Mean responses for counterchecks for ease of use and understandability for different visualization using GLM (Between Subjects) 206

8.20 Mean responses for findability of documents sharing the same location using GLM (Between Subjects) 207

8.21 Mean responses for effectiveness of visualization to aid users in identifying most relevant documents using GLM (Between Subjects) 208

8.22 Mean responses for usefulness of different visualizations (Spatial Qualifiers) 209

8.23 Mean responses for information presentation and understandability of different visualizations (Spatial Qualifiers) 211

8.24 Mean responses for legends and annotations as supporting aides for different visualizations (Spatial Qualifiers) 211

8.25 Mean responses for colour usage and appropriateness for different visualizations (Spatial Qualifiers) 214

8.26 Mean responses for viewpoint and relativity for different visualizations (Spatial Qualifiers) 215

8.27 Mean responses for findability of documents which shared the same location for different visualizations (Spatial Qualifiers) 216

8.28 Mean response for findability of documents which shared the same location for different visualizations (Spatial Qualifiers) 217

8.29 Mean response for effectiveness of different visualizations in their ability to identify relevant documents (Spatial Qualifiers) 218

8.30 Descriptive statistics for spatial activity plots or maps. 219

8.31 Descriptive statistics for isoline representation. 220

8.32 Qualitative feedback questionnaire to draw statistical inference between system generated relevance and perceived relevance for the user 223

Chapter 1

Introduction

"What to do with too much information is the great riddle of our time"

- Theodore Zeldin. An Intimate History of Humanity (Stafford & Webb, 2004)

During its course of evolution the internet has seen exponential growth and has become increasingly complex and has scaled to new dimensions. Among the many information sources, the '*internet*' – and more specifically the World Wide Web (WWW, or 'the web') – has evolved as the most widely used information source today. Green (2000, p.124) states that "*despite its uniform interface and seamless linked integration, the web is not a single coherent element*". There are two distinct elements of the web: the '*visible*' and the '*invisible*', which are sometimes referred to as the '*static*' and '*dynamic*`. "*The static web pages provide the same generic information to everyone, while dynamically generated web pages provide unique information, customized to the user's specific requirements*" (Green, 2000. p.124). Neilsen (1995) has placed the estimates of internet growth to around 333 percentage annually, while Murray (2000) based on the continuum model has estimated that 7.3 million unique pages are added per day. However, the study results do not include ftp sites or secure sites (HTTPS). Furthermore, the study reveals that of the more than 350 million links considered about 10.43 percent generated broken link error messages and less than one percent timed out. Google claims to have indexed more than 8 billion pages on November 10th 2004 (Source: http://www.caslon.com.au/metricsguide2.htm#pages).

Understanding the heterogeneous nature of the internet can lead to a characterization as being unstructured (vaguely structured), incomplete, inter linked, disparate, diverse, and dynamic. Bearing in mind the complexities in gleaning through information on the internet, numerous strategies have evolved to present relevant information. Thus, information retrieval engines, widely known as '*search engines*', have become widely adopted. In the next section, we explore through the evolution of search engines.

1.1 Search Engines

Early search engines primarily focused on retrieving the information based on keywords from an indexed database of web pages. Green (2000, p.126) states that the primary components of a search engine are:

1. a spider that examines web sites;
2. an index/database of web site listings;
3. interrogation/retrieval software.

`spiders` are "*programs that search the web for new web pages, index words and /or links on those pages, and match the indexed words with the URL of the pages on which they appear*" (Green, 2000, p.126). The core component of these search engines is "*what the user interrogates*" (Green, 2000, p.126). Based on the user's information request any search engine is equipped with tools that facilitate interrogation of a database to match words or terms which are indexed by the spiders. The location and frequency of these search terms against a list of matching web sites defines the ranking/relevance of each of the corresponding web pages with respect to it. Different algorithms and techniques are applied to analyse the locations and frequencies to '*rank*' each of these web pages retrieved by the search engine.

The rank of a web page (document) can be defined as a mechanism to determine the importance of any web page based on the user's information request. It has been widely argued in the literature that the ranking of a web page (document) does not necessarily reflect the potential relevance as perceived by the users. Research regarding relevance dates back to the late 1950's and 1960's (Mizzaro, 1997) where relevance has been considered as a potential property not only of utterances but also of other observable phenomena, thoughts, memories and conclusions of inferences. Mizzaro (1997) presents an exhaustive literature review outlining definitions and criteria adopted in the past three decades to qualify and quantify the measure of relevance in the information sciences. The relevance history has been placed in three periods (before 1958, 1959-1976, 1977-present). A careful analysis is presented subsequently based on seven aspects: "*methodological foundations, different kinds of relevance, beyond-topical criteria adopted by users, mode of expression of the relevance judgements, dynamic nature of relevance, type of documents representation adopted, and agreement among different judges*" (Mizzaro, 1997. p.811). For the sake of brevity, we strike at those aspects of relevance which have been a result of demonstrating the capability to retrieve and represent the information requested by the user. Barry (1994, p. 154, Table 1) has defined a set of criteria which govern the relevance of a specific document, including the usefulness of the documents' information content, the user's experience and background, the user's belief and preferences, the information environment, the sources of the documents, obtainability of the document, and the user's situation.

Often varied statistics and comparisons (Bar-Ilan, 1998/1999), Henzinger *et al.* 1999) have been used based on *size*, *freshness*[1], *vaguely structured context within the document* (Henzinger *et al.,* 2002), "*popularity or connectedness*" (Henzinger *et al.,* 2002, Lawrence & Giles, 2000), *short-lived inconsistent search behaviour*[2] over *search engine overlap*[3] (Jacsó, 2005, Sherman, 2004). *distribution* (Lawrence & Giles, 2000), *change overtime, coverage and recency* (Lawrence & Giles, 2000. Eastman & Jansen, 2003). *Unique hits*[4] and *dead links*[5] have been presented by different

---

[1] http://www.searchengineshowdown.com/stats/freshness.shtml

[2] http://www.searchengineshowdown.com/features/google/inconsistent.shtml
http://www.searchengineshowdown.com/features/av/inconsistent.shtml
http://www.searchengineshowdown.com/features/hotbot/inconsistent.shtml
http://www.searchengineshowdown.com/features/nlight/inconsistent.shtml
http://www.searchengineshowdown.com/stats/

[3] Notess (1999) cites that the importance of trying multiple search engines, highlighting missing information, if one uses long phrases and that presents better results when broken into smaller phrases. http://www.searchengineshowdown.com/stats/overlap.shtml

[4] http://www.searchengineshowdown.com/stats/unique.shtml

[5] http://www.searchengineshowdown.com/stats/dead.shtml

industries working in the information retrieval business.

However, despite the proposal of such differentiated criteria in the literature, for over a decade, '*size*' has been a surrogate of '*relevance*' for comparative benchmarks between different search engines. A relative size showdown has placed Google as forerunner [Notess, 2005, Sullivan *et al.,* 2004]. However, the notion of size as a bench mark for comparison for search engines has been widely criticized by technology journalists and editors of Search Engine Watch (Sullivan *et al.,* 2004). Furthermore, they argue that "*search engine size figures are useful but by no means should they be taken as a surrogate for relevancy figure*".

According to Schwartz (1998) search engine services can be primarily categorized into classified lists and query-based engines. According to Schwartz (1998,p. 974), classified lists (or directories) such as Yahoo, MSN, Lycos, Snap, AOL, and Search Britannica "*present an array of resource links in systematically arranged categories, often complex hierarchies*" and "*usually allow query-based search of category labels and resource titles*". Conversely, query-based engines such as Google, Altavista, HotBot, Excite, Teoma and Go "*run search algorithms based on user input text expressions*" and "*provide browsable categories as well, but it is generally obvious that a search engine is primarily of one kind of the other*".

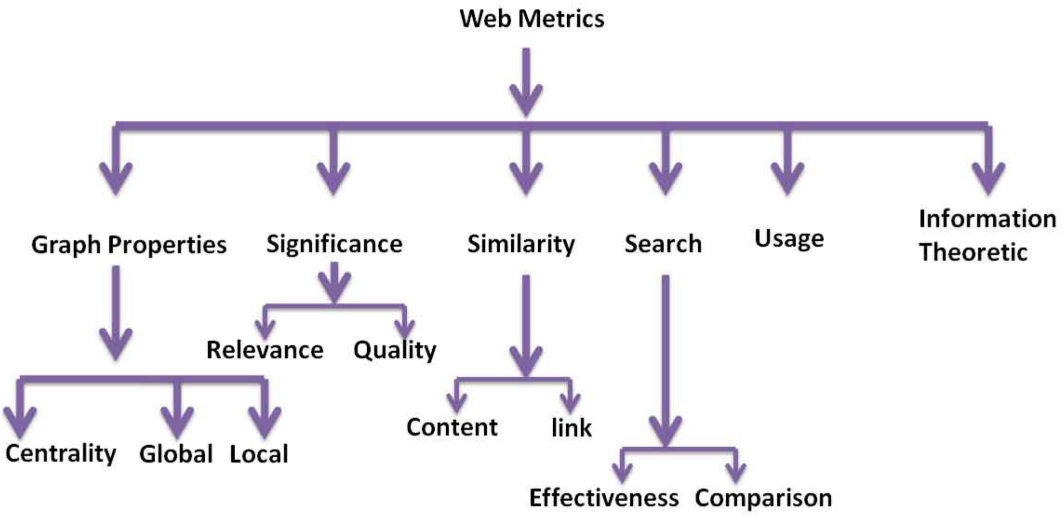


Figure 1.1 : Taxonomy of Web Metrics (Dhyani *et al.*,2002. p.471, Fig.1)

Dhyani *et al.,* (2002) made a similar effort in measuring all aspects of the information and identified the taxonomy of metrics shown in Fig 1.1, which originates from diverse areas such as classical informetrics, library science, information retrieval, sociology, hypertext and econometrics. These metrics presented are "*invariably proposed within the context of techniques to improve the quality and usefulness of measurable objects, in this case, information on the web*" (Dhyani *et al.,* 2002., p. 470). These search services are predominantly governed by the set of features and functionalities each of the search engines deliver. A comparative chart of search engines features is presented by the technology journalist Notess (2005) of Search Engine Showdown.

1.2 Motivation

The internet has mobilised significant research interests within cartography, geographic information science, and computer science over the last decade. Information on the internet in essence is in the form of text and graphics. During its evolution varied formats such as text and graphics have been embedded within the internet to describe and illustrate concepts. Berners-Lee (1999) conceptualized the internet (web) as an information space that facilitates not only human information exchange, but also participation of machines and users within the communication medium. In addition to this, it is rich with information relating to a geographic location, despite being abstract. The internet is also embedded with information relating to the occurrence of specific events with respect to a geographic location. One such illustration are news articles which detail the occurrence of specific events with reference to a geographic location (Herzog, 2003).

The potentials of the web as an interface for mapping had been envisioned in the late 1980s (Egenhofer & Frank, 1988). DiBiase *et al.* (1993) illustrated that the provision of a map-based graphical interface to geographic databases facilitates the ease of presentation of dense information as compared to text-based interfaces. Laurini & Milleret-Raffort (1990) state that using an object-based hyper document introduces spatial referencing to all the components. So-called hypermaps have been considered as a medium to navigate through huge amounts of information. The mid 1990s saw an upsurge in free web mapping services, including the U. S. Bureau of Census Tiger Mapping Service, Multimap (1996), Map24 (1996), and MapQuest (1996). These services focused primarily on meeting the the demands of the tourism industry. Their primary features include door-to-door travel directions, road maps of the world, street level maps, and tourism related services such as booking of holiday cottages, restaurants, or train tickets.

Subsequently, evolution took place in the Alexandria Digital Library (Smith, 1996; Chen *et al.,* 1997) established in 1990s to address the problems of traditional map libraries. By 1999, it had evolved to facilitate services based on an extensive collection of geo-referenced information objects. This included capabilities to access information from items in the library spread across the internet and by description of geographic locations on Earth, as well as by other characteristics of the information (Smith *et al.,* 2001).

Ding *et al.* (2000, p. 326) discussed the interest within the geographical community to identify information embedded within the web pages and cited that "*unfortunately, current web search engines largely ignore the geographical scope of the web resources*". In 2000, there was a tentative version of "somewherenear" (http://www.somewherenear.com), which was designed as a geographic search engine facilitating users to find things by proximity to other things, based on distance from the information available on the internet. This search engine focused on retrieving resources associated to a specific place or post code (or zip code) found in associations with yellow page services and some digital map web sites.

Jones *et al.* (2002) identified that "*the northern light geo-search tool from Vicinity (http://www.northernlight/geosearch.html) includes a facility for geographical information retrieval. It allows the user to enter part or all of an address, along with a category of interest and a search radius. It finds other places within the specified radius, with the aid of a digital map*". A similar retrieval facility is found on www.somewherenear.com, as mentioned above. A major shortcoming of this approach (i.e. linking yellow pages to web documents) is that these geo-search tools cannot recognize alternative for the same place, whether they are modern and historical variants, informal names or names of contained places. Furthermore, their user interfaces are limited in the options offered to specify and refine a search and to visualize the results, reflecting their lack of knowledge of the semantics of geographical terminology and the associated spatial structures.

A study carried out by McCurley (2001) to estimate the percentage of web documents referring to geographic locations has established that 10% of the web documents referred to a US zip code or telephone number. Gravano *et al.,* (2003) carried out experimental studies to classify web pages (and resources in general) according to their geographical locality either local or global, independent of whether they contain place names, based on the prevalence and diversity of place names in search results for the queries. Search engine query logs are rich with information relating to geographic location, formulated by the users. Sanderson & Kohler (2004) analysed an Excite Query Log from 2001 (1,025,910 queries) to investigate the extent and variation of web queries containing geographic terms. They found that 18.6% of the queries contained a geographic term and 14.6% had a place name (refer to Table 1.1). In 2005, Himmelstein estimated that at least 20% of the web pages included one or more easily recognisable and unambiguous geographic identifier, such as a postal address. Very recently, Zhang *et al.,* (2006) identified that 12.7% of four million queries sampled contained a place name, and argued that it is the proponent of the user's information request.

Table 1.1: Statistics of geo-queries in 2001 Excite log (Sanderson & Kohler, 2004)

Variables	**No of Queries**	**% of the geo-queries**	**% of full sample**
Queries with place names	369	79.5	14.8%
Queries with other geo-terms	189	40.7	7.6%
Queries with any geo-terms	464	100	18.6%

Despite the prevalence of geographic context in web documents, existing search engine facilities are poorly adapted to help users find information that relates to a specific '*geographic location*', and neither are they equipped with tools to visualize this information in/onto the geographic dimension. Hence, an ambitious quest for the exploitation of these potential internet resources has been on the agenda in cartography and geographical information science during recent years. The dynamism, scale and diversity of the data coupled with its incomplete and unstructured nature, pose a still stronger challenge to perceive this information while adding a geographic dimension to it. Notedly, technology journalist Levy (2004) quotes of the potential challenges of "*adding a geographical dimension to an existing application not only increases its utility but sometimes produces a level of information that's downright scary*". Indeed by adding a geographic dimension to this unstructured, disparate, diverse and dynamic information obtainable from the internet has offered a multitude of possibilities and also solutions to users interested to learn about specific information with respect to a geographic location.

The complexities and the challenges posed by the addition of the geographic dimension to the information available from the internet led to the demand for tools that process and disseminate information to the users. Egenhofer (2002, p.1) envisioned the need for "*better retrieval methods by incorporating the data semantics and exploiting the semantics during the search process*" with a "*geospatial perspective so that the particularities of geospatial meaning are captured appropriately*". Furthermore, Egenhofer (2002, p.1) identified the need for the creation of "*multiple spatial and terminological ontologies, each with formal semantics*" which facilitates the ease of representation with respect to computation and human understanding. In addition to this Egenhofer argues that it would "*lead to the creation of new frameworks for geospatial information retrieval based on the semantics of spatial and terminological ontologies*". For the creation of such frameworks, would necessitate the "*evaluation of retrieval results based on the match between the semantics expressed information need and the available semantics of the information resources and search systems*". However, in the years following Egenhofer's paper little attention has been paid to retrieve information based on geographic context and in generating adaptive and easy to understand visual representations of the same.

Thus, in this work we aim to exploit contemporary knowledge from diverse fields such as information retrieval, spatial cognition, cartography, information visualization and usability studies to deliver a comprehensive discussion of the problem of geospatial representation and exploratory visualization of the results of spatially aware information retrieval. This work was embedded in the EU/FP5 research project SPIRIT (http://www.geo-spirit.org), in which six European groups collaborated to develop technology and methods for geographic information retrieval.

A further source of inspiration is drawn from Egenhofer's (2002) vision for the creation of a framework for drawing geospatial perspectives from unstructured, disparate, dynamic, diverse and sometimes incomplete information from the internet. Montello (2002, p. 283) pointed out that for drawing geospatial perspective, '*maps*' are cognitive devices that ease human understanding, but has cautioned that these '*maps*' are usually of subjective design and their influence would stimulate new methods and new perspectives. In tune with these arguments, Benelli *et al.* (2001, p. 22) also stated that "*maps are cognitive artifacts that represent not only the characteristics of the information space but also the use people make of the space*".

Foundations towards mobilizing such a framework have been seen in the works of Woodruff & Plaunt (1994), Buyukkokten *et al.* (1999), which primarily focused on exploiting geographical location information of web pages. Their work offers initial approaches to compute the geographical location of all entities. Woodruff and Plaunt (1994) proposed an algorithm for automated extraction of geo-positional coordinate index terms from text to support geo-referenced document indexing and retrieval. Presenting a case for the need for such manual indexing in the literature by Hearst & Plaunt (1993) and Holmes (1990). Notedly, it has also been identified that Aronoff (1989) presented a notion of geo-referenced data which is commonly characterized as having physical dimensions and a spatial location, and a geo-referenced document is indexed according to such data. In addition, the inherent difficulties involved in text-based georeferencing are presented. These being:

- *neologism*[6] (Farrar and Lerud, 1982).
- *spatial and naming variation* (Farrar and Lerud, 1982)
- *name changes* (Farrar and Lerud, 1982 and Griffiths, 1989)
- *spatial boundary changes* (Griffiths, 1989)
- *lack of uniqueness* (Griffiths, 1989 and Holmes, 1990)

Finally, an algorithm is presented by Woodruff & Plaunt (1994) which extracts geographic references from text by identifying and matching with a thesaurus, entries of geographic place names and other geospatial perspectives (2D Polygons), by weighting value which reflects its geographic nature and other properties (such as location and density). These weighing values are further mapped onto 3D topographies whose geographic scope is determined by the threshold for the evaluation of topographies and areas that are beyond some threshold, depicted as results.

Watters & Amoudi (2002) presented an experimental system with location-based ranking of search engine results according to physical location and distances; these results were re-ranked based on distance between the query and the scope of the retrieved document determined by geo-referencing the URL of a page to longitude & latitude. Lee *et al.* (2003) implemented an experimental geographic web search facility by employing a map interface for the specification of the area to be searched, for the city of Kyoto, usually referred to as KyotoSEARCH. In 2003, Rauch presented a novel approach to compute the geographical focus for each web page which is used to restrict results to a specific region and display results on a map. A confidence based framework has been adopted to model the probability that a given name refers to a given place, which is often termed as disambiguation. Furthermore, Rauch (2003) attempts to understand grammatical expressions which define some kind of relative positioning, during geographic place name extraction for e.g. "5 km north of downtown Zurich".

Similar attempts have been made by Markowetz *et al.* (2004) to develop a geographic web search engine similar to SPIRIT, where the geographical scopes of the web pages are based on the

---

[6] neologism refers to a word or term or phrase which has been recently coined or created, referring to new concepts, to synthesize a pre-existing concept or to make older terminology sound more contemporary.

multiple sources of evidence including the WHOIS directory and text content of the web pages. A geographic footprint is created with a set of grid cells recording the degree of spatial relevance for each web page. However, their approach does not include interpretation of spatial relations that relate the subject of interest to a particular place, other than distance and no experimental studies have been conducted. Amitay *et al.* (2004) presented a system similar to Markowetz *et al.* (2004) and Rauch (2003) to identify the geographical focus of the web page called Web-a-Where. Unlike Markowetz *et al.* (2004), Web-a-Where extracts and grounds geographic references found in the web pages and uses these to compute a focus for the page – a locality the page is assumed to refer to as a whole. The geographic locality here is referred to as source and target based on the physical location of the server hosting of the page and the latter on the coverage of the page's content. Furthermore, they demonstrated that by integrating it into an existing data mining framework to geotag web pages gathered from a web crawl, they can enable a geographic search. Wang *et al.* (2005) and Silva *et al.* (2006) adopted a different approach for identifying the geographical scope of the document while the former categorized geographic references into three types: provider location (physical location of the web resource derived from yellow page addresses), content location (refers to the location that the content of a web resource describes), and serving location (scope of the web page through analysis and linking or referencing to other web resources) to reflect the observation that various types of location can co-exist in a single source. The latter approach, called GREASE (Geographic REasoning in Search Engines) makes use of a graph in the ranking method to assign a single geographic scope to a web page.

Larson (1996) defines geographic information retrieval (GIR) as an area of specialization, concerned with providing access to geo-referenced information sources. It includes all of the areas that have traditionally formed the core of information retrieval (IR) research with an emphasis, or addition, of spatially and geographically oriented indexing and retrieval. Resonating with Larson, Jones and Purves (2004) define geographic information retrieval as "*the provision of facilities to retrieve and relevance rank documents or other resources from an unstructured or partially structured collection on the basis of queries specifying both theme and geographic scope*".

Some of these geographically enabled graphic representations may be foreign to the users or more seriously, appear analogous but have different meaning; creating opportunities to demonstrate new techniques and representations for spatially aware information retrieved, from the internet.

1.3 Problem Statement

Cartography and information visualization have been evolving and changing rapidly with the usage of technological tools. There has been a growing interest to equip users with tools to extract geographic information from unstructured textual sources such as the internet, and the SPIRIT project aimed to fill this gap by designing and implementing a geographic information retrieval engine (see http://www.geo-spirit.org).

More specifically, situated within the context of SPIRIT as well as other earlier work, this research aims to deliver a comprehensive solution to the problem of generating geospatial perspectives and visualisations of spatially aware information retrieved from the internet. The emphasis is placed on capturing the geospatial meaning of the relevant context and matching it with user's information request. Within the context of this research, we perceive spatial awareness of geographic space, wherein human activities are rooted. In turn, the information entity which describes the set of human activities or features associated to it in geographic space can be characterized as the spatial context.

Within the context of cartography and geographic information science, this research critically engages in

- understanding the spatially aware information seeking process to develop new theories and to underpin our development of visualization methods;

- understanding the metaphors adopted during the spatially aware information seeking process;

- proposing techniques to visualize information retrieved from the internet and combine geographic and thematic relevance in appropriate ways;

- and finally validating the effectiveness of these visualization methods by devising empirical measures, in order to infer their strengths and weaknesses.

1.3.1 Research Questions

1. To what extent different information seeking models help us in defining the process of spatially aware information seeking ?
2. How does our understanding of spatially aware information seeking process guide us in designing exploratory representations ?
3. Which of the exploratory representations assist users to identify geographic information retrieved from unstructured textual source such as the internet ?
4. What metrics assess the efficacy of the visualization techniques adopted to represent geographic information retrieved from unstructured textual source such as the internet ?

The focus is on developing exploratory 2D visual representations for geographic information extracted from unstructured textual web sources. However, this work does not deal with the temporal dimension of the geographic information retrieved from the internet, which could multiply the complexity of the problem. This research develops theories and models associated with the geographic information seeking and retrieval process, sometimes termed as spatially aware information seeking and retrieval process. Foundations of this research are drawn from different disciplines that can help us think about approaches to visualize the geographic information retrieved from internet sources.

1.4 Organization of the thesis

The thesis is organized into 10 chapters, arguing the necessity to carry out this research and subsequently building from the literature, investigating the search process by carrying out experimental studies and drawing a requirement specification to construct a theoretical model and finally validate the effectiveness of the system implemented by the SPIRIT Consortium, particularly its visual representations of search results.

Chapter 1 explores the evolution of the internet and search engines as well as the importance of the concept of relevance. During the course of this chapter, the necessity for extracting geographic information from unstructured textual sources has been identified and various approaches demonstrated in the literature are discussed, laying down the necessity for a spatially aware information retrieval engine and the necessity to develop new theories to understand the process and underpin our development of on-demand cartographic representation for geographic information retrieved from the internet.

Chapter 2 presents the scope of this research and develops the theoretical and technical background. First, theories central to spatially aware information seeking are reviewed followed by a contemplation of their relationships to information seeking models, information visualization, geographic information science and cartography. Each of which are dealt with in detail by the keywords identified. Second, a range of contemporary visualization techniques are reviewed. Subsequently, typologies of different geovisualization tasks are identified. Third, we explore various mapping interfaces widely adopted today. Finally, we summarize the potential of various visualizations and their implications within the context of this research.

Chapter 3 introduces what the spatially aware information seeking process is, and develops a conceptual model. It identifies how a spatially aware information seeking model differs from traditional information seeking models. First, theoretical models are presented by illustrating a clear understanding of user tasks and activities involved, underlying interactions and feedback with the system, role of cognition and affordance by carrying out early stage experimental studies for requirements gathering. Then, the assessment of the theoretical model is carried out by means of a reassessment exercise.

Chapter 4 defines and characterizes different visualization variables to present possible scenarios with a design perspective. Data characteristics and constraints associated with the problem of representing geographic information on the web are discussed. Finally, various concepts for visualizing spatially aware information retrieved from the internet are presented.

Chapter 5 describes the transfer of the conceptual model of spatially aware information seeking and the core elements of geovisual representations into 'The Prototype', developed by the members of the SPIRIT Consortium. The architecture for spatially aware information retrieval is presented along with user interaction models. The prototype illustrated two different interfaces, the so-called '*structured interface*' that enforces a structured text query and the explorable, visualization-based interface, termed the '*graphical interface*'. Both of these interfaces involve the three phases query formulation, query refinement/expansion, and result presentation.

Chapter 6 discusses various visualization approaches adopted such as density surface representations, footprint based representations, cartograms, spatial activity maps and isoline representations. Visualization elements associated with each of the visualizations are identified and their characteristic features are presented in detail.

Chapter 7 describes the user experiments which are modelled based on the factorial experiment strategy to draw meaningful metrics of the different visual representations proposed in chapter 5. First, user experiments are designed bearing in mind the focus groups. Secondly, pre-test experiments and a reassessment of experiment design are carried out. Finally, pilot tests are carried out to enhance the experiments and define the framework and the parameters involved to extract usable metrics for each of the visualizations.

Chapter 8 discusses the resulting metrics adopted to comparing the usability of different visualizations. First, descriptive statistics are presented as pair-wise comparisons of different visual representations. Second, significance tests are carried out and discussed regarding the advantages and disadvantages of different visualizations. Finally, we summarize the findings of the experimental studies.

Chapter 9 runs through discussing the critical findings of the experimental studies presented in chapter 8 and tries to draw meaningful inferences from the statistical insights and experiments carried out not only in Chapter 8 but also during the course of this research, which are illustrated in the earlier chapters of this thesis.

Chapter 10 presents the achievements and contributions of this research and presents careful insights, suggested improvements and criticism, and an outlook on future research.

1.5 Contributions to the SPIRIT project

This thesis was embedded within the EU Vth Framework Programme sponsored project SPIRIT, and in view of my research objectives, the thesis explored various theoretical models of information seeking, extended them to the spatially-aware case and sought to evaluate and test them. The contributions by the author for the SPIRIT project are outlined in Table 1.2.

1.2 Contribution by the author for the SPIRIT project

1.	Requirements analysis and design.
2.	Design of exploratory representations for spatially aware information retrieved from the internet.
3.	Contributions to interface design and development.
4.	Experimental framework for the evaluation of geographic information retrieved from the internet.
5.	Analysis of the results gathered from the experimental studies.

Chapter 2

Theoretical Background

“Findability isn’t limited to its content, nor is it limited to the web. It is about designing systems that help people find what they need”

-Peter Morville, The Age of Findability, Boxes and Arrows

This chapter seeks to lay the theoretical foundations for the key terms of this thesis. The foundations of this work lie in information retrieval, information seeking models, cognition, geographic information sciences, cartography, as well as knowledge discovery and usage. Each of these areas is reviewed keeping in mind their relationship with the research questions of this thesis. Therefore, this chapter sets out discussing key aspects and summarizes their potential contributions to deliver a comprehensive solution to the problem of extracting and representing spatially aware information retrieved from the internet.

2.1 Information

Information in essence is a “*series of correspondences observed within a finite set of variational concepts or components*” (Bertin, 1983. p.16). This could be widely argued as a field specific definition of the term “*information*”. Losee (1998, p.260) has argued for a discipline independent definition of the term `*information*’; “*which is always informative about something, being a component of the output or result of the process. This* ‘*aboutness*’ *or representation is the result of a process or a function producing the representation of the input, which might, in turn, be a function producing the representation of the input, which might, in turn, be the output of another function and represent its input, and so on and so forth*”. In addition to this, Losee (1998, p.260) concludes that “*information may be understood as the value attached or instantiated to a characteristic or variable returned by a function or produced by a process. The value returned by a function is informative about the input to the process and about the process itself*”.

Within the context of this research, information can be defined as the meaning, knowledge instructions, communication, representation and mental stimulus that triggers a finite set of correspondences by the users. It is a state in which an individual shows understanding or realization or perception. In psychology this could be well termed as being responsive to a specific condition based on perception or realization. The quest for ‘*information*’ is not limited to a specific source. Speech, text, graphic images etc. are in essence different forms of information. We tend to search for necessary information within various sources. One such source is the internet, which has evolved today as a key potential source for seeking information. In the subsequent sections, we present an in-depth review of information seeking.

2.2 Information seeking

Information on the internet in essence is in the form of text and graphics. During its evolution varied formats have been embedded within the internet. This section sets out discussing key aspects with respect to seeking information on the internet (the source) and is organized as follows:

- Definitions of the information seeking process.(2.2.1)
- Differing models of the information seeking process. (2.2.2)
- Cognitive styles of users in the information seeking process. (2.2.3)

2.2.1 Definitions of information seeking

Prior to the internet revolution, various media such as speech and paper had been used to communicate specific information. With the advent of technological tools, there has been rapid transformation of these media. Thus, numerous sources of information are at the user's disposal and finding information has no longer been limited to a specific information source, and the internet has evolved into a key potential source of acquiring information. This act of acquiring '*information*', which isn't limited to the content nor to the internet can be termed as '*information seeking process*'. Recalling Losee's definition of information as a process, it should be noted that, "*information is not the process itself. The input to the process is not information about the process, although it clearly may be information about another process. The output is also not information by itself – the values in the output are information only in the sense that they are information about the process and the input, that is, information in the context of the process and its input.*" (Losee, 1998, p. 261).

Numerous authors (Schutz & Luckmann, 1973; Belkin *et al.,* 1982; Norman, 1988; Kuhlthau, 1991; Marchionini, 1995; Shneiderman *et al.,* 1997; Wilson, 1999; Hearst, 2001) have significantly contributed over the past four decades to developing theories and models based on user perspectives and behaviour for information seeking. As discussed in Chapter 1, the evolution of the internet has facilitated creation of tools to find relevant information, what we understand as '*information retrieval*'. Navigation within the information spaces on the internet has facilitated browsing or exploring through web documents, thanks to the hypertext protocol.

Characterizing the information seeking behaviour on the basis of interaction or as a search paradigm Bates (1986, p.92-93) states searching is an active and directed form of information seeking where users commit various actions with the intent of acquiring information and directed towards specific goals of interest, while browsing is active but undirected form of information seeking where users commit various actions and goals which are directed towards best match/exact match. Users adopting a passive behaviour of information seeking tend to watch for specific patterns within the information which is subjective (i.e. directed) towards their needs. Some users adopt a passive form of information seeking which is more undirected in nature and tend to learn or be receptive (i.e. aware) of the information contained in and out of experiences on interactions during the process of information seeking. These interaction paradigms are summarized in Table 2.1.

Table 2.1: Forms of information seeking (Bates, 1986)

	Directed	Undirected
Active	Searching for	Browsing
Passive	Watching for	Aware

2.2.2 Models of information seeking

Jansen *et al.* (1998) and Mann (2002) state that there are significant differences between information seeking on the internet and classical information retrieval. In classical information retrieval systems, the frequency of reformulation of search terms is relatively high when compared to the information searching process. Despite such differences, we consider classical information retrieval and information search processes as a single process.

Subsequently, Bates (1989) excogitated a model that is of great theoretical and practical importance to date, '*the berry-picking model*'. Wilson (1981, 1999), Mann (2002) and Järvelin *et al.,* (2003) propose different information seeking and retrieval models based on task, processes and behaviour that serve different research purposes. Knowledge of the different modes of information seeking behaviour aids our understanding of the search process and to what extent different models represent reality.

Choo *et al.* (1998, 2000) perceive information seeking behaviour as modes of scanning usage (illustrated in Table 2.2) and web moves (illustrated in Table 2.3). A number of observational studies carried out simultaneously to validate the theoretical models illustrated above and to determine the true nature of the process by Ellis (1989), O'Day & Jeffries (1993), Borgman (1996) had supported the '**berry-picking model**'. The model illustrates that the user's actions are unprophetical in nature and tend to shift their objectives or goal posts based on acquired information and inferences that come into play. Unfurling interest across various domains a novel approach was presented by Pirolli and Card (1995) based on the optimal foraging theory in biology and anthropology "*which analyses the adaptive value of food-foraging strategies*". This theory focuses more on the value of information gained against the costs of performing activity in human-computer interaction tasks. This theory is called the '*information foraging theory*'.

Table 2.2 Modes of Scanning (Choo *et al.*, 2000;. Fig. 1)

Scanning Modes	Information Need	Information Seeking	Information Use
Undirected Viewing	General areas of interest; specific need to be revealed	**'Sweeping':** Scan broadly a diversity of sources, taking advantage of what's easily accessible	**'Browsing':** Serendipitous discovery
Conditioned Viewing	Able to recognize topics of interest	**'Discriminating':** Browse in pre-selected sources on pre-specified topics of interest.	**'Learning':** Increase knowledge about topics of interest
Informal search	Able to formulate simple queries	**'Satisfying':** Search is focused on area or topic, but a good-enough search is satisfactory.	**'Selecting':** Increase knowledge on area with narrow boundaries.
Formal Search	Able to specify targets in detail	**'Optimizing':** Systematic gathering of information about an entity, following some method or procedure	**'Retrieving':** Formal use of information for decision, policy making.

Based on these studies and the current understanding, information seeking can be characterized as a set of structured or unstructured actions performed by the users based on their "*individual interest (content) or situational activity (context)*" (Turnbull, 2003). The whole process constitutes a set of information seeking `*episodes*', that is, gathering of '*bits and pieces*' of information towards a final goal. Supporting this, many authors (e.g. Xie, 2000; Spink *et al.*, 2002) state that these structured or unstructured actions are dependent on the user's cognitive state, personal knowledge (Heinstrom, 2003; Blickle, 1996), prior experiences, selection of the information source (Dervin & Nilan, 1986), language familiarity with the problem definition, interactions during the course of information seeking (Bellardo, 1985), and judgement/intuitions based on the outcome and their personal interests, amongst other factors.

The linear models proposed by Marchionni (1992, 1995), Shneiderman (1998) and Hearst (1999) inadequately represent the exploratory nature of information seeking, which is non-linear. Studies carried out by Bates (1989), O'Day *et al.,* (1993), Borgman (1996), Hendry and Harper (1997), Cousins *et al.,* (1997), Foster (2004) support these arguments. In addition to this, user actions – interactions with the system – haven't been fairly presented in the earlier, linear models.

Table 2.3 Information seeking behaviours and web moves (Choo et al., 2000;. Fig.2)

	Literature search moves (Ellis *et al.* 1989, 1993, 1997)	**Anticipated web moves**
Starting	Identifying sources of interest	Identifying web sites/ pages containing or pointing to information of interest
Chaining	Following up references found in given material	Following links on starting pages to other content related sites
Browsing	Scanning table of contents or headings	Scanning top-level pages: lists, headings, site maps.
Differentiating	Assessing or restricting information according to their usefulness	Selecting useful pages and sites by book marking, printing, copying and pasting etc.; choosing differentiated pre-selected sites.
Monitoring	Receiving regular reports or summaries from selected sources	Receiving site updates using e.g.: push, agents, or profiles; revisiting, favourite sites.
Extracting	Systematically working a source to identify material of interest.	Systematically search a local site to extract information of interest at that site.

Xie's (2000) work sheds more light on interaction facilitating understanding of the information seeking trends of the users, where users engage in multiple type of information seeking strategies when using information retrieval (IR) systems. The study presents an indepth analysis at the micro-level of user goals when interacting with different types of libraries and identifies four types of shifts (*i.e. planned shifts, oppurtunistic shifts, alternative shifts, assisted shifts*) in interactive intentions and three type of information-seeking strategies (*They are: current search goal shifts, interactive intention shifts and information seeking strategy shifts*). The study recommends "*interactive approach*" which takes account of both "*planned model*" and "*situated action*". This approach has a number of advantages:

- It helps in the segregation of user tasks involved in each episode to develop a task typology (Bystorm & Järvelin, 1995; Bystorm, 1999).
- It facilitates understanding of the cognitive metaphor adopted by the user to accomplish the objective (Fabritius, 1998).
- It reinforces the mapping of functionalities of the system onto user interactions, for example for system feedbacks (Spink, 1997).

2.2.3 Cognitive styles in information seeking

Among other factors that influence the information seeking process are perception and cognition. They not only play a vital role in mobilizing the structure of the content but also the way it is being represented. Ingwersen (1992, 1994, 1996) has identified the "*potential value of matching the multidimensional variety of representations inherently existing in or extracted from information objects and from the cognitive space of a user*". This tendency consistently displayed by the individuals to adopt a particular type of information processing strategy has been termed '*cognitive styles*' (Ford *et al.*, 2002). These cognitive styles do not necessarily remain the same for all users. Some tend to adopt a holistic approach, while others adopt a serialistic approach.

Pask and Scott (1972, 1973) and Pask (1976a; 1976b; 1976c; 1979; 1988) state that holistic users tend to adopt a comprehensive understanding to learning, examining interrelationships between several topics early in the learning process and concentrate on building a capacious overview into which details are subsequently fitted. On the contrary, serialists adopt a predominantly local learning approach examining one thing at a time. Drawing a clear distinction from Pask & Scott (1972, 1973), Witkin *et al.* (1977) state that the users tend to adopt a '*spectator*' approach to learning rather than a hypothesis testing approach, which is termed '*field dependence*' or '*participatory approach*' where the user is adept at structuring and analysing the hypothesis (Witkin *et al.*, 1977).

2.2.4 Relevance to ongoing research

Having explored through various linear and non-linear models and supporting empirical and behavioural studies investigating the nature of the information-seeking process we have learnt that the '*berry-picking model*' (Bates, 1989) appropriately reflects the true nature of the process. Hence, our foundations for spatially aware information retrieval on the internet are built based on this model. However, as these beliefs should be validated by further empirical studies, a series of experiments should be designed, in order to build not only from the existing literature, but to draw a better understanding of the process with specific respect to spatially aware information seeking. Subsequent chapters in this thesis will therefore address this issue further.

In an effort to equip an individual with necessary interactive and exploratory tools to glean through huge amounts of information retrieved from the internet, we explore different disciplines that help us think about extracting and representing the information retrieved. One effective tool that equips an individual with these capabilities to sense, stimulate and coordinate different concepts in the mind is through *visualization*. The subsequent sections deal with those disciplines that can help us think about visualization and also prove to hold strong relevance to our specific research problem.

2.3 Visualization

Visualization in essence is the graphic depiction/association to what is to be perceived and has been a medium of communication and understanding. "*Graphic representation constitutes one of the basic sign-systems conceived by the human mind for the purpose of storing, understanding and communicating essential information*" (Bertin , 1983, p.2). This medium of communication, termed visual instruct is diverse in nature and characterised spatially, temporally, objectively or subjectively with limitations.

Sifting through the literature, the course of the research builds from what visualization is, to exploring various methods/techniques widely being practised and their relevance to the ongoing research. The section is set out as follows:

- Definition of visualization (2.3.1)
- Geovisualization (2.3.2)
- Visual exploration (2.3.3)
    - Principles of visual information seeking (2.3.3.1)

- Interaction typologies (2.3.3.2)

2.3.1 Definition of Visualization

Graphics owes its special significance to its double function as a storage mechanism and a research instrument. A rational and efficient tool when the properties of visual perception are competently utilized, graphics is one of the major '*languages*' applicable to the information process (Bertin, 1983).

Visualization can be defined as

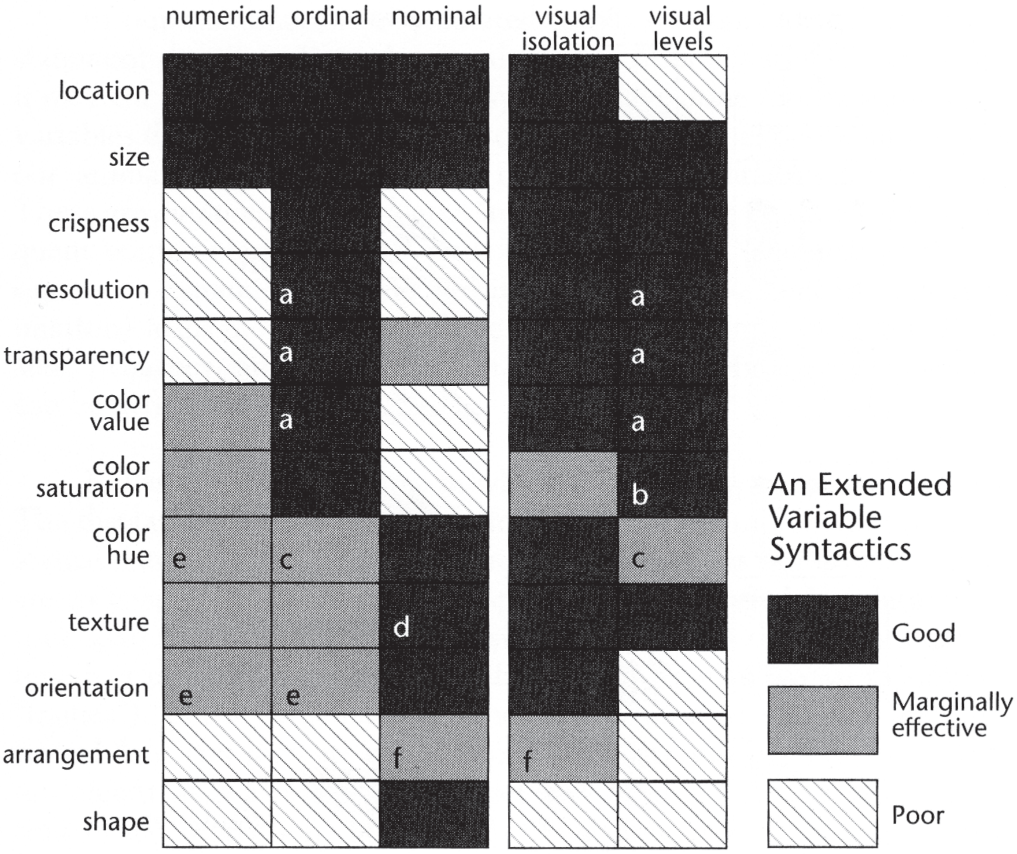
- (*a classical definition*) " *a formation of mental visual images, the act or process of interpreting in visual terms or putting into visual form*". (McCormick *et al.*, 1987)
- "*the use of computer imaging technology as a tool for comprehending data obtained by simulation or physical measurement*". (Haber and McNabb, 1990)
- "*the process of representing information synoptically for the purpose of recognizing, communicating and interpreting pattern and structure. Its domain encompasses the computational, cognitive and mechanical aspects of generating, organizing, manipulating and comprehending such representation. Representations may be rendered symbolically, graphically or iconically and are most often differentiated from other forms of expression (textual, verbal or formulaic) by virtue of their synoptic format and with qualities traditionally described by the term "Gestalt"* (Buttenfield and Mackaness, 1991, p.432).
- "*the binding (or mapping) of data to representations that can be perceived. The types of binding could be visual, auditory, tactile, etc. or a combination of these*" (Foley *et al.*, 1994)

According to Andrienko & Andrienko (2006, p.166) to visualize means

- "*To recall or form mental images or pictures* (*intransitive*)";
- "*To form a mental image of* (*transitive*)";
- "*To make perceptible to the mind or imagination* (*transitive*)".

Bertin (1983) has developed a fundamental theory of the graphic sign system and the semiotics that can be used to build and understand visual representations in visualization. Bertin (1983, p. 53) distinguishes between four ways in which graphical signs can be arranged to create a visualization, and terms these '*groups of imposition*': diagrams, networks, maps, and symbols. Maps are specific to the presence of geodata.

The notion of information as a process primarily constitutes one or several pertinent correspondences between a set of finite variational concepts and the invariant. The central notion common to all pertinent correspondences is termed '*invariant*' by Bertin and the variational concepts are called '*components*' and '*visual variables*'. Furthermore, these *visual variables*' constitute elements/classes or categories/steps, that depict the qualitative and quantitative aspects of them (Bertin, 1983, p.4-6). Bertin (1983) identifies six visual variables at the disposal of a graphic sign system for expressing a pertinent correspondence, including *size, value, texture, colour, orientation* and *shape*. On a two dimensional plane these tend to vary in relation, leading to the four forms of imposition mentioned above. Other authors, primarily to adapt to the technological evolution that made dynamic, rather than static displays possible, later extended the original system of visual variables by Bertin. For instance, further contributions by MacEachren (1995) include *crispness (fuzziness) of object edges, resolution (spatial precision) and transparency*. Figure 2.1 provides an elaborate overview of perceptual properties of visual variables.



	numerical	ordinal	nominal	visual isolation	visual levels
location	■	■	■	■	▨
size	■	■	■	■	■
crispness	▨	■	▨	■	■
resolution	▨	■ a	▨	■	■ a
transparency	▨	■ a	▒	■	■ a
color value	▒	■ a	▨	■	■ a
color saturation	▒	■	▨	▒	■ b
color hue	▒ e	▒ c	■	■	▒ c
texture	▒	▒	■ d	■	■
orientation	▒ e	▒ e	■	■	▨
arrangement	▨	▨	▒ f	▒ f	▨
shape	▨	▨	■	▨	▨

An Extended Variable Syntactics

■ Good

▒ Marginally effective

▨ Poor

a - The clarity variables of crispness, transparency and resolution can be used for no more than two or three categories. These variables are untested, but assumed to be most useful for representation of uncertainty. They may prove to be most practical in an interactive setting in which an analyst is able to toggle them on and off when needed.

b - Purer, more saturated colours appear to be in the foreground, while dull, unsaturated colours fade into the background.

c - Hues must be carefully selected for an order of hierarchy to be apparent (e.g. the part-spectral sequence from yellow through orange or red) Hues interact with one another in sometime unpredictable ways, so it is often difficult to determine which hues will dominate others.

d - Pattern texture is good for only two, or perhaps three, identifiable categories.

e - Orientation provides limited ability to communicate numerical or ordered information—glyphs based on a clock face and geologic strike and dip symbols are successful examples.

f - Pattern arrangement is best as a redundant variable to make a visual difference between categories more obvious.

Figure 2.1: Visual variables and their perceptual properties (from MacEachren, 1995, p. 279)

2.3.2 Geovisualization

Visualization can be broadly characterized into information visualization and scientific visualization based on scope and usage. McCormick *et al.*, (1987) states that "*scientific visualization is concerned with data that is based on physical reality and may or may not have an explicit spatial component*", whereas information visualization (Card *et al.*, 1999) can be defined as "*the representation that is concerned with abstraction of data that is not physically based and not inherently geometrical in nature such as financial data or collection of documents*". Within the realm of scientific visualization, in the last decade, there has been growing interest in the explicit spatial component, the geovisualization domain.

Geovisualization can be defined as

- "*the map use domain, specifically for private use by individuals having the goal of revealing unknown or interesting geographical relationships through map use activities that have a higher degree of human interactions*". (Board & Taylor, 1977)
- "*the use of concrete visual representations; whether on paper or through computer displays or other media; to make spatial contexts and problems visible, so as to engage the most powerful of human information-processing abilities, those associated with vision*". (MacEachren *et al.*, 1992, p. 101).

- "*a science that combines GIS, cartography, scientific visualization, and exploratory data analysis that has considerable potential to provide theory, methods, and tools for the visual exploration, analysis, synthesis and presentation of data that contains geographic information* (MacEachren and Kraak, 1997; MacEachren *et al.*, 1999)
- "*the usage of visual geospatial displays to explore (methods of exploration) data and through that exploration to generate hypothesis, develop problem solutions and construct knowledge*" (MacEachren, 2001).

Hence, geovisualization can be understood as the integration of cartographic communication, cognition and the usage of latest computing techniques to amplify cognition in relation to a specific problem. Gahegan (1999) identifies some barriers that constrain the effective use of geovisualization. They are as follows:

- *"Rendering speed, technological barriers affecting the interactive display and manipulation of large datasets;*
- *Visual combination effects – problems to be overcome when displaying many layers, themes or variables simultaneously;*
- *The orientation of the user into a visualized scene or virtual world and measuring the effectiveness of the scene or the world as the problem-solving tool".*

2.3.3 Visual Exploration

The complexity in modelling and understanding the visual exploration process has excited interest by scholars to approach it with different perspectives. With users in action, correspondences or interactions are governed based on the metaphor adopted by them. These pertinent correspondences of users with the variational concept and the invariant can be termed '*interaction*' (Bertin, 1983, p.5). By correspondence, we mean the input and the response or feedback obtained from the system (which is a black box to the user). Often these correspondences are reciprocate in nature between the user and system, which are termed feedback. Feedback when delivered by the system is termed system feedback and conversely denoted as user feedback when given by the users.

We have discussed above some primary components of the visual exploration system, while DiBiase (1990), Taylor (1991), and MacEachren *et al.,* (1992) expressed the need for a process based approach to illustrate the stages involved within a graphic exploratory system (as shown in Fig 2.2). DiBiase (1990) identifies these four stages as:

- *Exploration of data to reveal pertinent questions*
- *Confirmation of apparent relationships in the data in the light of formal hypothesis.*
- *Synthesis or generation of findings.*
- *Presentations which evolve in parallel with progression from private to the public realm.*

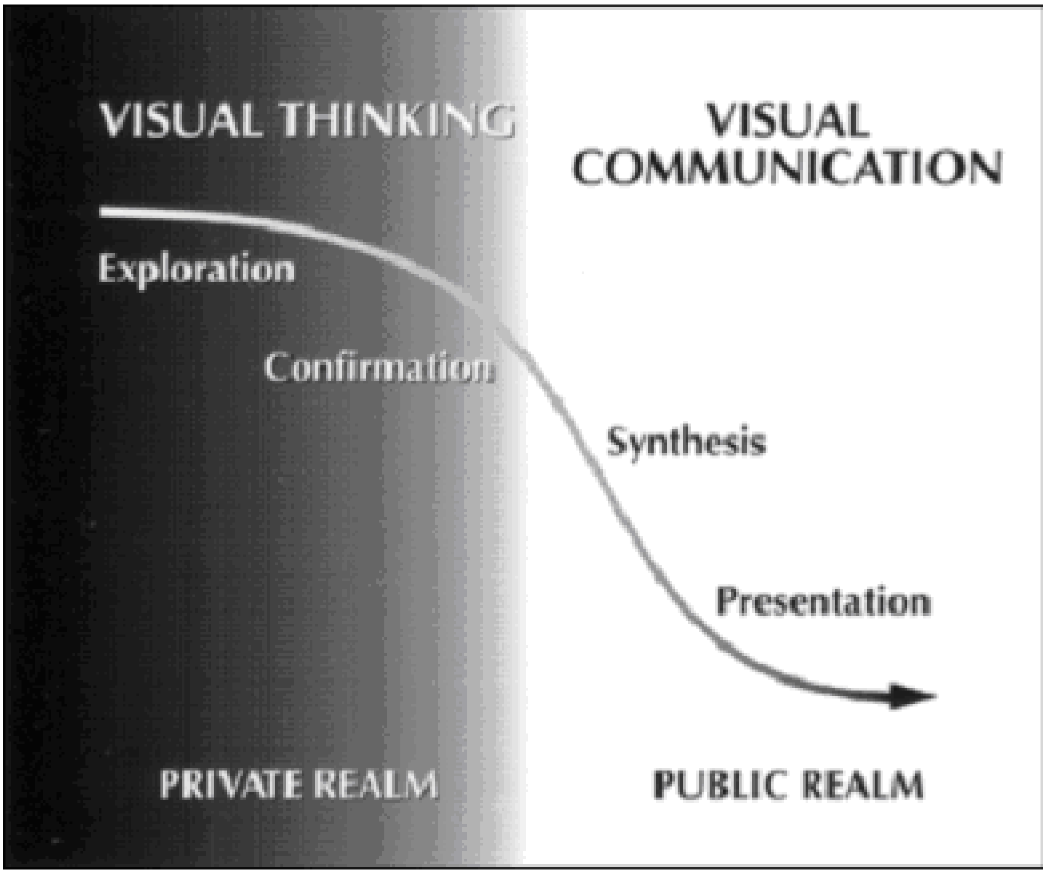


Figure 2.2: The range of functions in an idealized research sequence. (DiBiase, 1990; http://www.geovista.psu.edu/publications/others/dibiase90/swoopy.html)

MacEachren & Kraak (1997) extended this model to the usage of maps by focusing primarily on interactions. The model (map use cube, illustrated in Fig. 2.3) emphasizes visualization and communication within the map usage space.

Considering a continuum, map usage is subjective to the individual's needs and directed towards revealing unknowns or presenting the knowns and has higher levels of interactivity based on the purpose it is designed to serve. MacEachren & Taylor (1994, p.7) have outlined exemplars of the eight extremes of the map use space defined by the use of map use cube. Similar proposals have been made concerning the visual metaphors adopted by the designers, by Cartwright (1999), detailing nine possible metaphors that designers tend to adopt. However, it is essential that map designers in the due process of map making should bear in mind the importance of map usage.

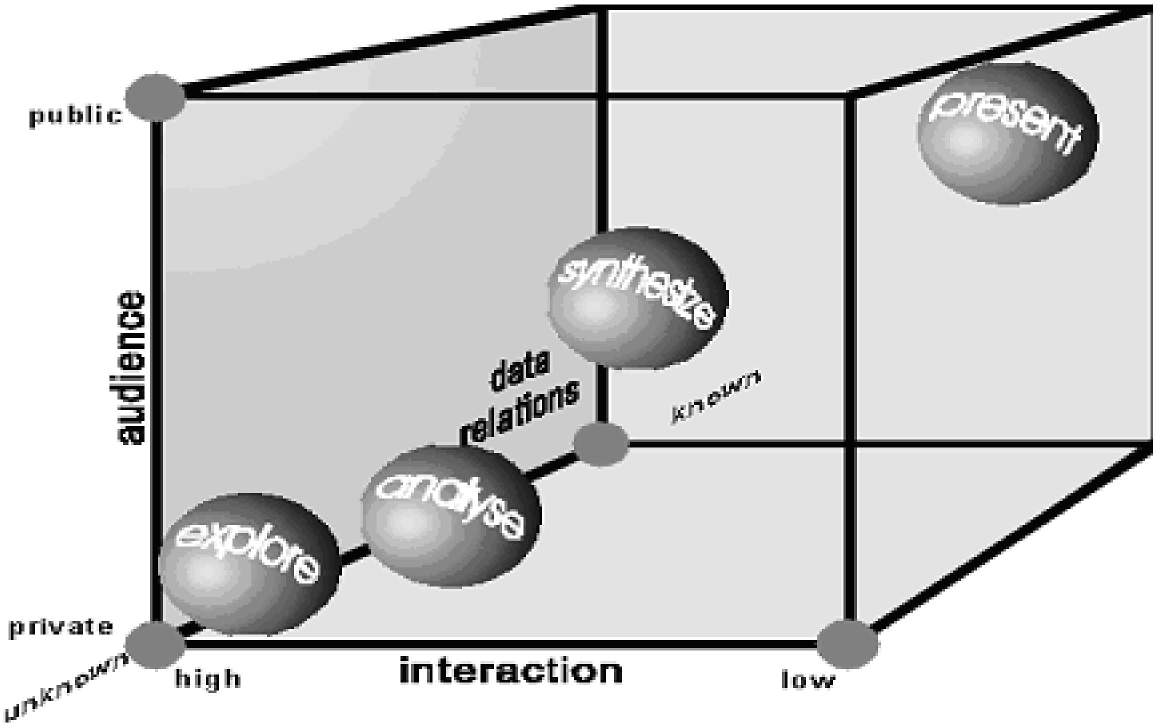


Figure 2.3 Goals of use arrayed in the map use cube (MacEachren & Kraak, 1997; p.338)

Visual exploration is intuitive and non-linear in nature involving higher degree of interactivity. Gahegan (1999) argues that users tend to build a hypothesis initially and later on discover information by processing, analysing, reducing or re-mapping from one space to another. Dykes *et al.*, (2005, p.25) term this the ‘*visual exploration paradigm*’ based on the ‘*information seeking mantra*’ (Shneiderman, 1996).

The degree of involvement of the end users and the purpose has unfurled an interest to develop novel techniques and models of information visualization (Pfitzner *et al.*, 2003; Hibino, 1999; Crampton, 2002) over the last decade. Chi (2000) presented a taxonomy of information visualization techniques using the data state reference model. Strongly criticising Chi’s work, Pfitzner *et al.* (2003) state that it has failed to present a taxonomy of visualization techniques, and that it rather focused on the taxonomy of processes involved. Furthermore, Pfitzner *et al.* (2003) created a unified taxonomic framework (as shown in Fig:2.4) for information retrieval visualization and interfaces which serves as guidelines during the design and evaluation phases.

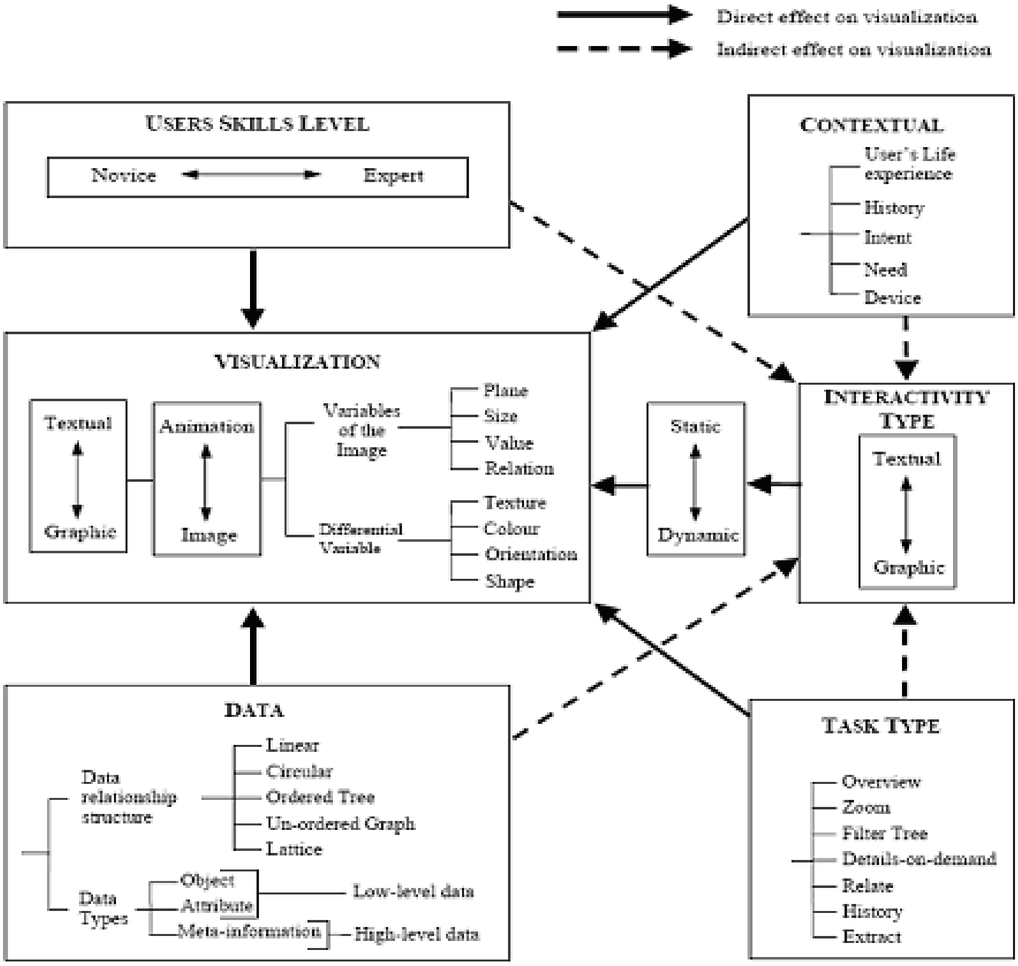


Figure 2.4: A unified taxonomic framework for information visualization (Pfitzner *et al.*, **2001; p. 64)**

2.3.3.1 Principles of visual information seeking systems

When confronted with the task of seeking specific information, visualization facilitates exploration. In the preceding section, we discussed the notion of the '*visual information seeking mantra*' (Ahlberg & Shneiderman, 1994) as being complex, non-linear and highly exploratory in nature. Systems that aspire to serve these characteristics need to follow a set of visual design guidelines. Drawing similarities with the context of exploratory data analysis and visual information seeking, they are primarily stated as (Andrienko & Andrienko, 2006):

- *Holistic view: demonstrate the holistic behaviour (the world of action) that can be perceived by the users.*
- *Simplify and abstract: separation of interesting data from excessive detail with emphasis on particularities, elimination of excessive details.*
- *Divide and group: progressive refinement of search parameters and subsequent classification.*
- *See in relation: identify the relationships and ability to select key features.*
- *Responsiveness: immediate and continuous feedback with respect to the interactions.*
- *Support for browsing: looking for specific recognizable pattern.*
- *Reformulation: ability to continuously reformulate goals.*
- *Formulization: establish structures in relation to the results from an interplay of several components.*
- *Zoom and Focus: thoroughly examine the details with respect to the degree of interest and identify key patterns.*

- *Establish linkages: integrate the key components to facilitate ease of understanding.*

2.3.3.2 Interaction/Interactivity

Interactivity is at the core of any dynamic visualization. The effectiveness of dynamic visualization is determined by ease of usage, understanding and predominantly the level of interactivity it facilitates. Modes of interactions signify the operations that can be applied to the data presented to the users, combine or relate specifics of information, compare and critique different views of data, and provide an effective mechanism to explore and understand the information.

Keim *et al.,* in Dykes *et al.* (2005, p.25-26) presented a classification of information visualization techniques based on three criteria:

- *the data to be visualized,*
- *the visualization technique,*
- *the interaction technique used.*

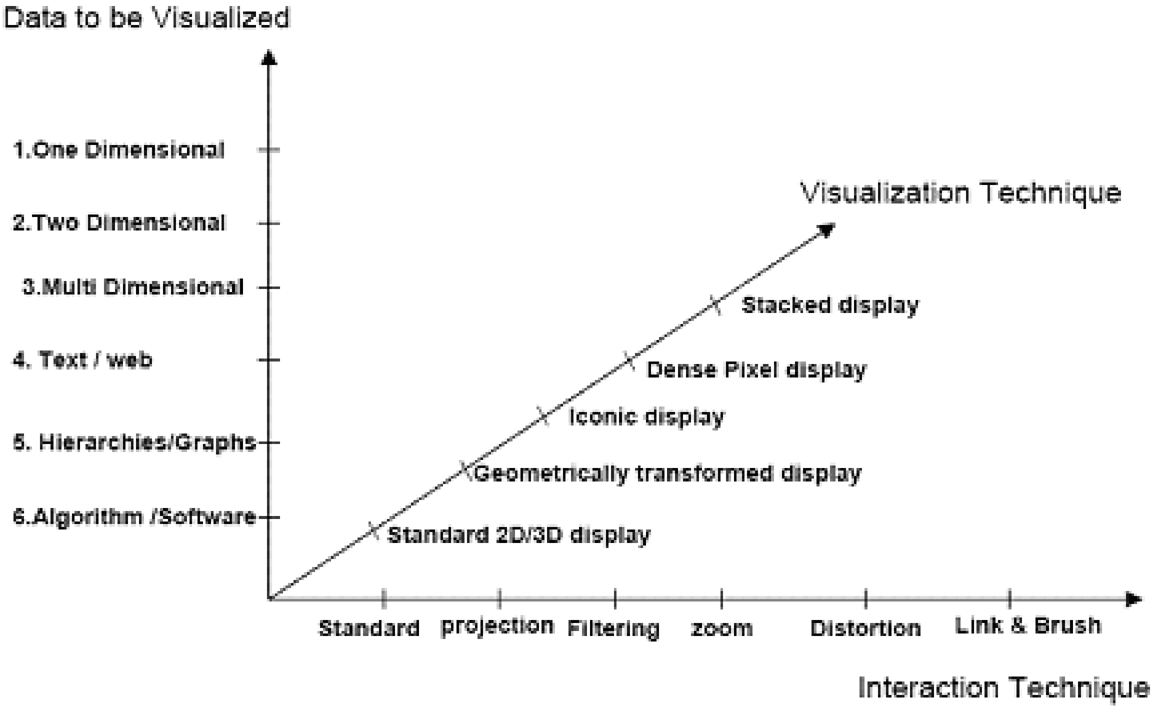


Figure 2.5: Classification of information visualization techniques (Keim ***et al.*****; p. 26 in Dykes** ***et al.*****, 2005)**

Figure 2. 6 illustrates the classification of these information visualization techniques based on the aforementioned criteria. Similarly, Goodell *et al.*, (2006) extended the notion onto developing a taxonomy to support analysis and modelling of the human data exploration process (as shown in Figure 2.6).

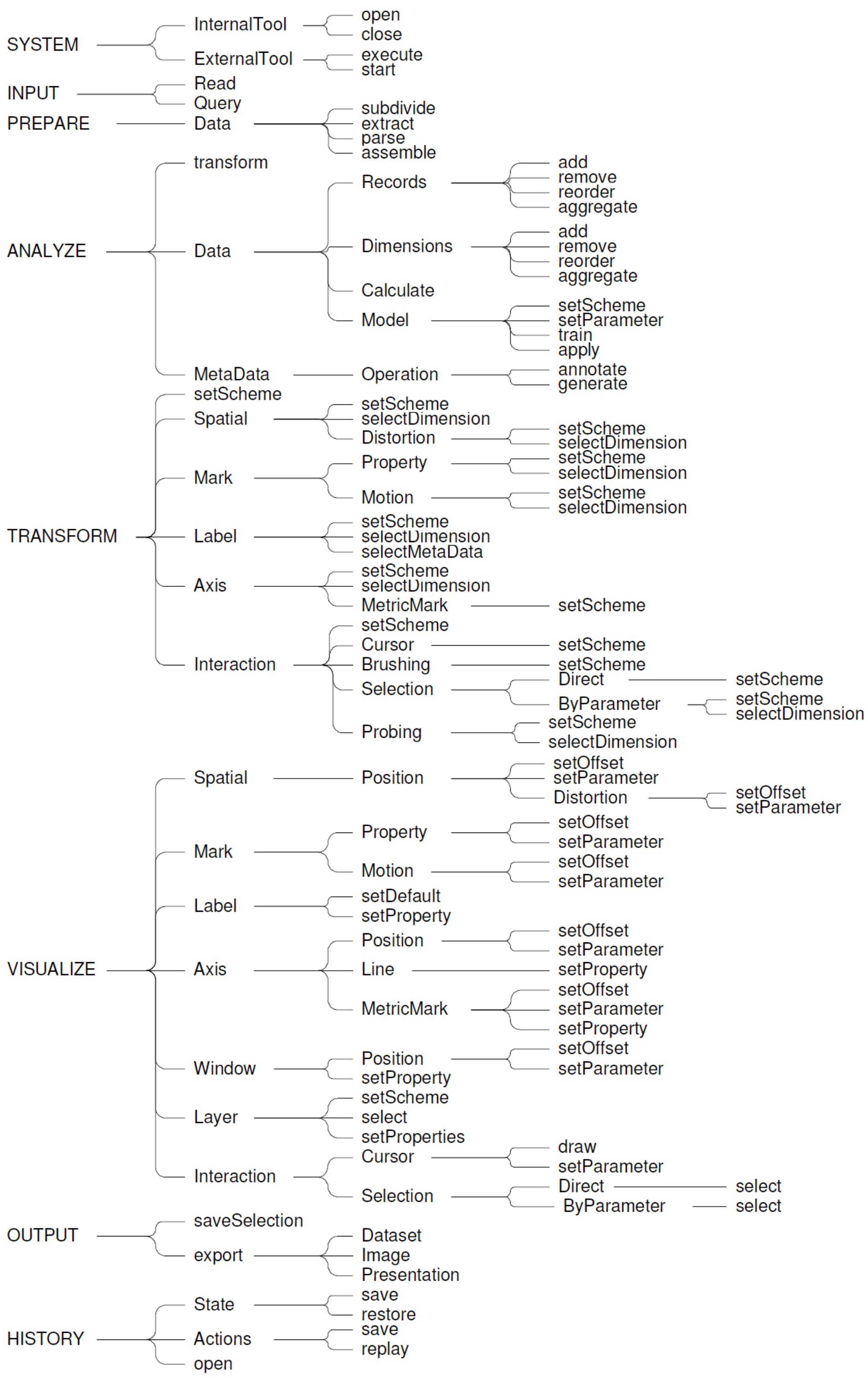


Figure 2.6: A taxonomy of visualization operations (Goodell *et al.*, 2006)

Crampton (2002) proposed a typology of interactivity based on increasing sophistication of interaction techniques (Table 2.4). In the remainder of this section 2.3.3.2, Crampton's typology will provide the structure of our discussion, as we believe that it covers the key interaction techniques that are relevant in geovisualization, and that the designers and developers of geovisualization systems should bear in mind these techniques to deliver a highly exploratory system.

Table 2.4: A preliminary typology of interactivity (Crampton, 2002; p. 90)

Interaction with data (high)	**Interaction with data representation (low)**
Database querying & Data mining	Lighting
Brushing	Viewpoint ("Camera")
-statistical	Orientation of data
-geographical	Zoom-in/out
-temporal	Re-scaling
Filtering (Excluding)	Remapping of symbols
Highlighting	
Interaction with temporal dimension (medium)	**Contextualizing interactions (high)**
Navigation	Multiple Views
Fly-through	Combining data layers
Toggling	Window Juxtaposition
Sorting or Re-expression	Linking

2.3.3.2.1 Interaction with data

'Interaction with data' is the first of Crampton's (2002) four broad classes of interaction techniques. 'Data' in this context is defined as the factual information available at the user's disposal, as raw material to analyse, reason, discover and select pertinent patterns in order to draw meaningful inferences. Often the data available are presented in different mediums and may be stored, communicated, interpreted or processed by automated means.

Database querying and data mining:

The automated process of discovery or extraction of factual information based on the user's request can be termed '*database querying*', whereas the automated discovery or extraction of hidden predictive information from databases can be termed '*Data mining*'. The latter process of late has been highly more interactive in nature than the former (Westphal and Blaxton, 1998, p.16). However, both processes are not directly dependent on visualization. Hence, we do not dwell further on the associated interaction techniques.

Brushing and Linking:

A process wherein users have the ability to interactively highlight, select or delete a subset of elements or conceptual entities with regard to the visualization performed, can be termed *brushing* (Monmonier, 1988, 1989). Often, brushing is associated with *linking*, which can be defined as the process of selective highlighting which directly affects or reflects the selection in other (linked) views.

Exploring possible usage of brushing techniques across the geographic domain, Monmonier (1988, 1989) presented a novel approach of *linking statistical plots* (Becker & Cleveland, 1987) with geography for analyzing spatially referenced data. This collection of techniques are termed '*geographical brushing*' and correspond to the events triggered by the users on interacting with the map display (i.e. with geographically referenced data). Note that this process is more than mere selection of geographic entities on the map. Besides geographical brushing, the notion of brushing can be extended to *attribute brushing* (clicking on items in a table or entities in a diagram), and *temporal brushing* (corresponding to timeline selection)(MacEachren *et al.*, 1998). The three techniques of geographic, statistical and temporal brushing are inter-linked and data can be explored to identify spatial patterns (Kraak, 2001; Crampton, 2002). Figure 2.7 illustrates an example of brushing and linking in the geographical domain.

Kraak (2001) states that "*these technological tools have embarked a new era of cartography which stimulates visual thinking, that would allow users to look at spatial and geo-referenced data in any combination, at any scale, with the aim of seeking or attribute or time or a combination of any of these three spatial components on to customized dimensions of interest expressed by the users*".

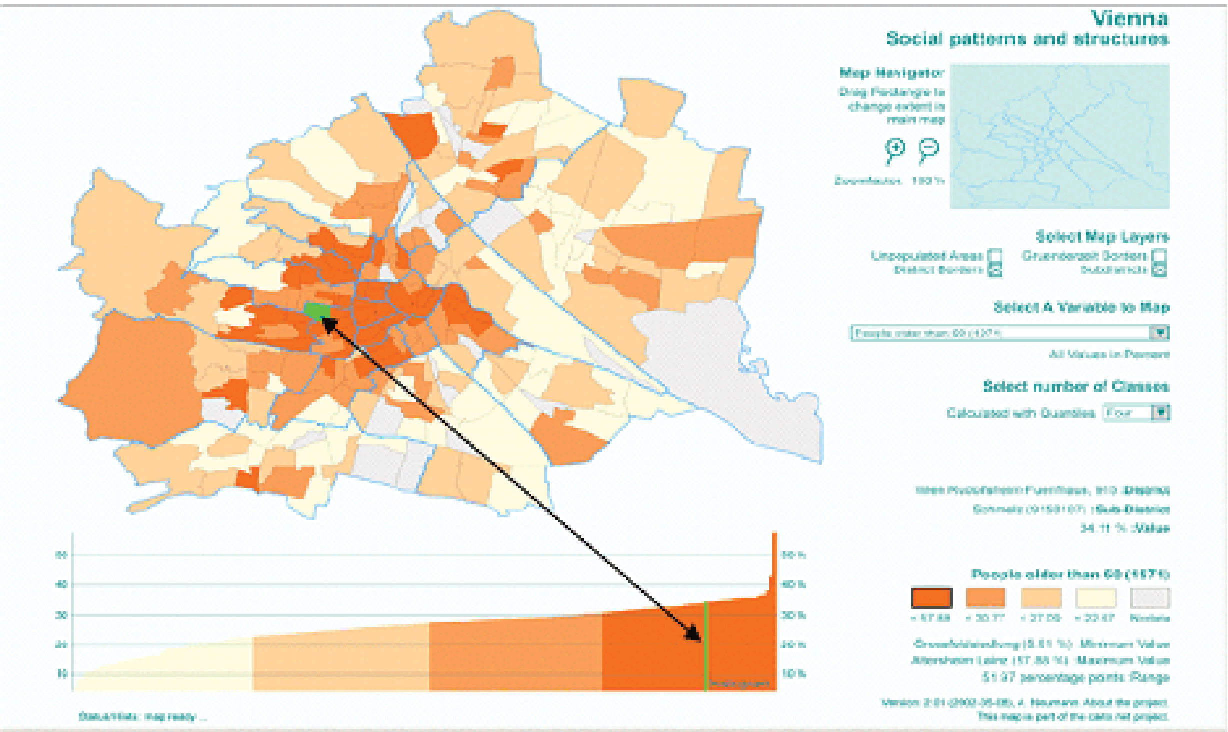


Figure 2.7: An example of brushing and linking. When the user moves over the map or histogram the corresponding element is highlighted (Neumann, 2003; http://www.karto.ethz.ch/neumann/cartography/vienna)

Many combinations can be perceived using brushing and linking, multiple views and focus, and context techniques to reduce the user's cognitive load. Ward (1994) extended the usage of brushing technique to dimensions greater than two, facilitating users to gain an understanding of spatial relationships in N-space by highlighting all data points which fall within a user defined, re-locatable sub-space. In addition to this, brushes are classified by identifying in which space the selection is being performed, screen or data space. Brushing in screen space is defined *as a continuous two-dimensional sub-space on the screen whereas in data space it is either by an enumeration of data elements contained within the brush or the N-dimensional boundary box that encapsulates the selection.* It is, however, argued that the objectives of brushing techniques can be to either specify containment with well-known boundaries by the user, or that brushing can act as a query tool to extract data by setting data ranges on different dimensions. Interactive manipulations using brushing techniques may be direct or indirect, if the user creates and modifies the brush directly by mouse on the data display, or indirectly, by usage of separate widgets like sliders or range sliders. Often direct interactive manipulations are carried out on screen space and subsequent responsive feedbacks are obtained as results, which make the process intuitive in nature. Data ranges for dimensions that are not shown must be specified with separate range sliders. Further, the brushes may be classified into range slider (Ahlberg and Shneiderman, 1994), beam brush, value brush (a data driven brush), and cluster brush (Sahling, 2002 & Kosara *et al.,* 2004).

Graphical formulation of queries is a process perceived as natural, where familiar brushing techniques can be of utmost usage. The notion of '*dynamic querying*' was promoted by Shneiderman

(1994) as an interactive approach to graphical database querying. However, numerous other conceived brushing techniques are discussed in the literature such as *conditioning* (Becker and Cleveland, 1987), *sectioning* (Furnas and Buja, 1994). Further work on enhancing the brushing concept was presented by Martin & Ward (1995). Their methods present novel concepts such as usage of non-discrete brush boundaries (ramped boundary brushes), simultaneous display of multiple (up to four) brushes, and creation of composite brushes via logical operators (data driven brushing). Wills (1996) demonstrated the extensive use of the concept of linking and brushing by extending it to multiple brushing types, and the flexibility in application of basic selection operations allow replacement, addition, subtraction and intersection of data on screen spaces.

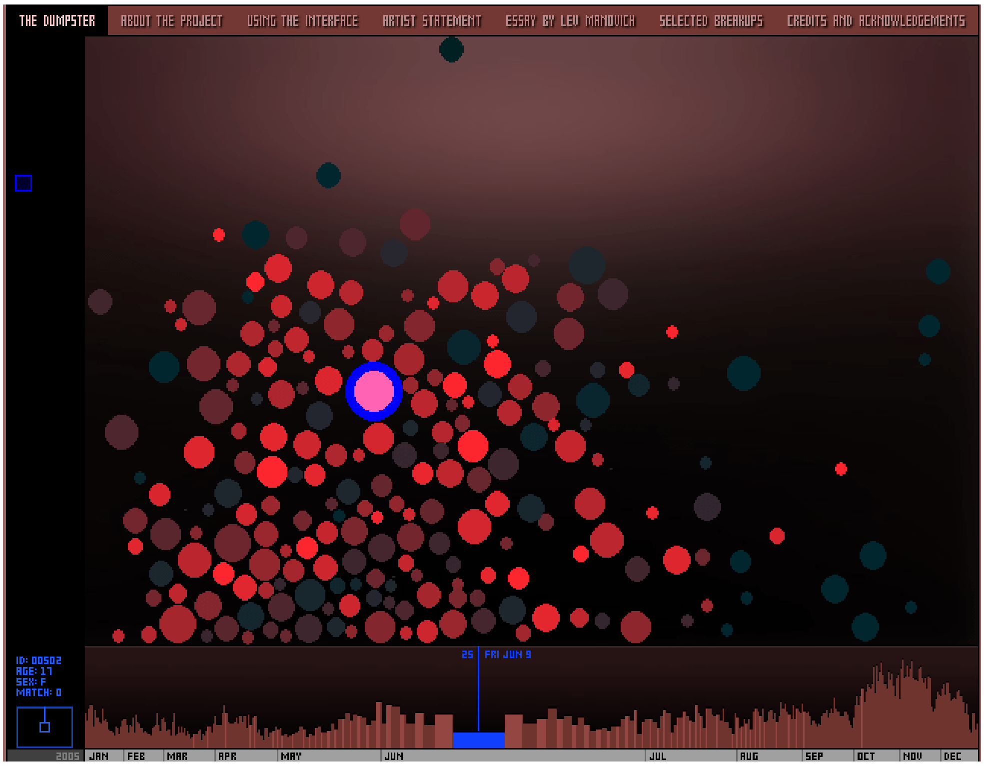


Figure 2.8: Dumpster – An interactive online visualization that depicts a slice through the romantic lives of American teenagers. (Levin *et al.*, 2006; http://artport.whitney.org/commissions/thedumpster/)

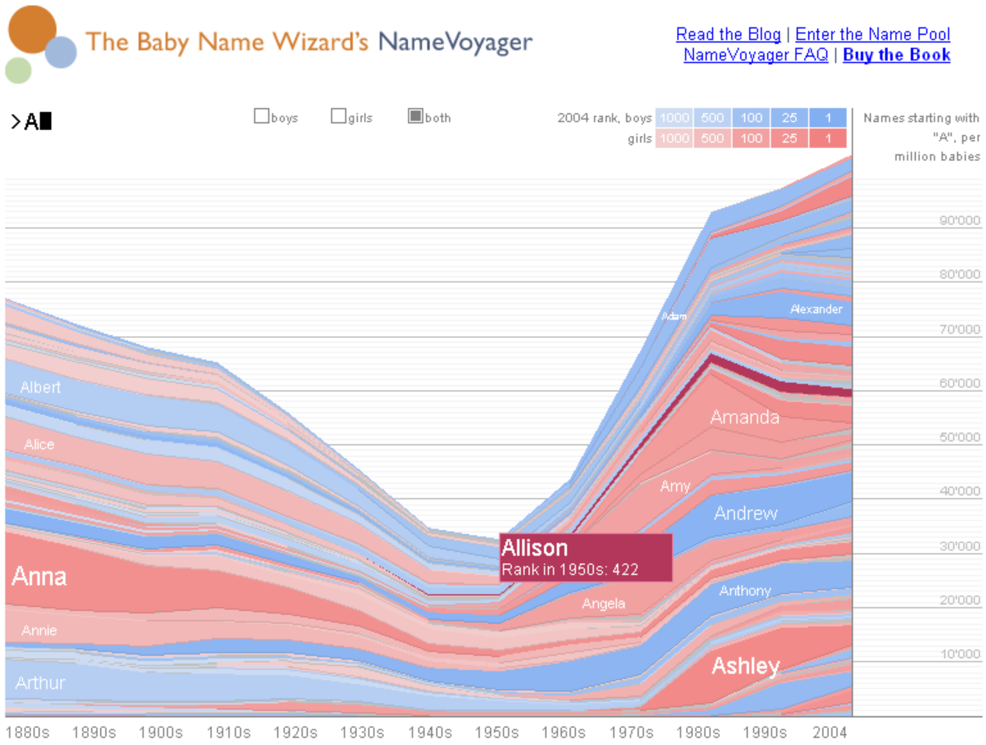


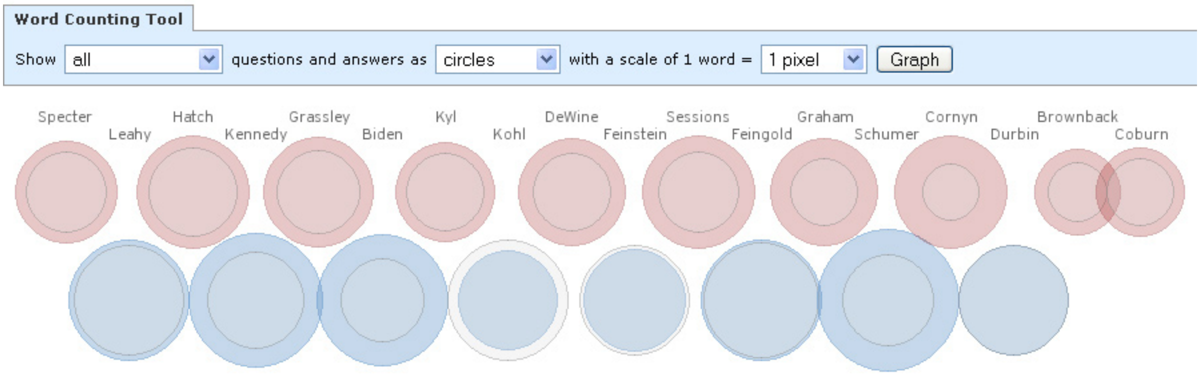
Figure 2.9: Namevoyager (Wattenberg, 2005; http://www.babynamewizard.com/namevoyager/Inv0105.html)

A few noted examples that present linking and brushing techniques are: XGobi -*Buja et al.,* (1996); *linked micromap plots* (Carr *et al.,* 1996, 1998, 2000, 2001); *Polaris* (Tang, 2001; Solte *et al.,* 2002); *Scalable Framework* (Lopez, 2001), *Compound brushing* (Chen, 2003); *Timesearcher* (Hochheiser *et al.*, 2001, 2003; Buono *et al.*, 2005), *Dumpster* (Levin *et al.*, 2006; Figure 2.8); *Namevoyager* (Wattenberg, 2005; Figure 2.9); *DynaMaps* (Li *et al.*, (2003).

Filtering and Highlighting:

Users seek the ability to sift through large amounts of data presented by *excluding, omitting or simplifying irrelevant or insignificant data from representation*. This ability is termed as filtering (Crampton, 2002). On the contrary, when *a specific set of data is emphasized or typified,* a technique often adopted to draw the user's attention to minute details of importance, this is termed as highlighting. An example of non-interactive nature of highlight- searching for word count tool (Corum, 2006; Figure 2.10). Figure 2.11 illustrates the example of filtering and highlighting by toggling on/off the layers that are not of interest to the users.

A few noted examples (snapshots provided in Appendix-E) of such usage are *Kartoo (www.kartoo.com)* where co-occurrent terms are filtered; *Grokker (www.grokker.com)* allows filtering based on the rank, domain and source; *Mapquest (www.mapquest.com)* where highlighting is accomplished by means of placing marker icons on places of interest such as schools, origin or destination of a route, etc.; and *Dynamic Queries* (Ahlberg and Shneiderman, 1994).



Show all questions and answers as circles with a scale of 1 word = 1 pixel Graph

Figure 2.10: An example of non-interactive highlighting (Corum, 2006; http://style.org/judgealito)

A combination of interactive filtering and highlighting techniques with focus-and-context, as well as a distortion technique to draw a different perspective is demonstrated in *Magic Lens* (Bier *et al*., 1993; Fishkin and Stone, 1995), which facilitates the user's ability to magnify and blur specific data.

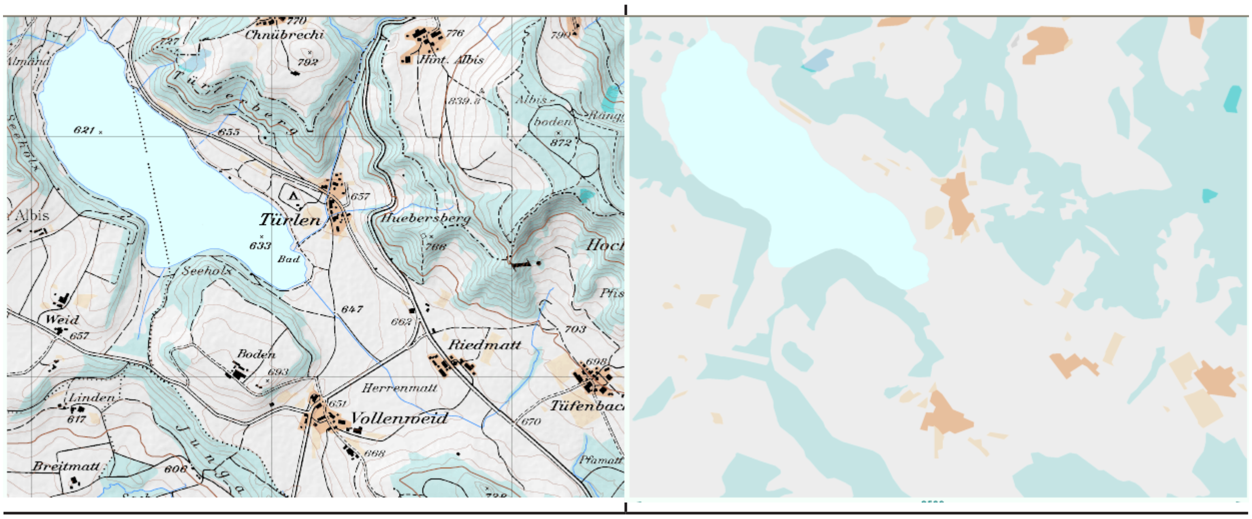


Figure 2.11 An illustation of simple filtering and highlighting by layer choice. Filtering information is achieved by by toggling on/off the layers that are not of interest to the user (source: http://www.carto.net/papers/svg/tuerlersee/)

2.3.3.2.2 Interaction with data representation

'Interaction with data representation' is the process of selectively modifying, by appropriate interaction techniques, the way in which the data (i.e. the factual information) is represented, such that the graphical representation better reflects the underlying information structure, allowing to stimulate a subjective cognition among the users. The interaction techniques of this class include the following:

Lighting:

Increasing or decreasing the illumination, ambience on a specific object in a representation has a significant impact on the user's perception. An example of lighting effect could be a shaded relief map with changing illumination angle to better bring out particular geomorphological features (Crampton, 2002).

Viewpoint ('Camera'):

A change in the viewpoint presents varied perspectives to the users of the same data. Gersmehl (1990) discusses the '*model and camera*' metaphor, which changes the view of the data scene based on the user-selected viewpoint of a virtual camera (Crampton, 2002, p.91).

Changing orientation of data:

It is an extension of the viewpoint techniques, where "*the model and camera metaphor consider how the model (data) can be re-oriented to allow a new perspective*" (Crampton, 2002; p.91). A noted example is Map24 (www.map24.com) which presents a change in orientation of the map perspective from 2-D to pseudo 3-D (2.5 D).

Zooming & focusing:

Zooming is a graphic interaction technique that facilitates scaled views of the representation on user's subjective selection to present the details. The process is either from abstract to detailed scaled viewing or vice versa, which is denoted as '*zoom-in*' or '*zoom-out*'. Navigation within these spaces is facilitated through panning either from left to right or up to down and vice versa. Furnas & Bederson (1995) present a broad range of techniques to display multiple scales of information in a two-dimensional representation, termed *multiscale viewing*. In the desktop metaphor, zooming, panning and scrolling address the constraints posed by the display to facilitate identification of essential information within large information spaces. Conversely, in the zoom metaphor, the scale is manipulated or varied over a single view using distortion techniques so that the outside information is not lost from view (Baik, Bala and Hadjaran, 2001) for instance in *fish eye views* (Furnas, 1986) or in the *bifocal metaphor.* These linear or non-linear visualization techniques impart a potentially powerful sense of control to the users presenting a macro/micro point of view (Tufte, 1990). Furnas & Bederson (1995, p.234) argued that "*in either case, the basic assumption is that by moving space and changing scale the users can get an integrated notion of very large structured and its contents, navigating through it is way more effective for their tasks*". Two conceptual issues need to be addressed concerning zooming, which are as follows (Bederson *et al.*, 2000): How to change the scale smoothly; and how to store data at different zoom levels.

Zooming can take two forms: *geometric* and *semantic zooming*. Geometric zooming simply provides a blow up of the graphical content whilst in semantic zooming the information content changes when approaching a particular area (von Wyss, 1996). Geometric zooming is linear in nature and depends on the physical properties of what is being viewed. Semantic zooming was first introduced by Perlin and Fox (1993) in the Pad System and later supported by Bederson and Hollan (1994). In semantic zooming the representation shown depends on the meaning to be imparted and is non-linear in nature. It usually requires an LOD (level of detail) concept and is closely related to automatic generalization (Cecconi and Galanda, 2002; Neumann, 2005). Semantic zooming not only affects the size of the objects but also their representations. Owing to its non-linear nature semantic zooming can be further classified into semantic lensing and semantic zooming.

Unlike traditional approaches to create multiple representations of an object and switching the view to the required degree of interest (DOI) Good *et al.* (2002) and Shipman (1999) present a novel approach for automatic generation of textual representations for changing space requirements. Another approach adopted is iconification and omission within the context of semantic zooming, proposed by Woodruff and Olston (1998).

Distortion:

Distortion is a view modification technique unlike traditional zooming, wherein concurrent presentation of local details together with the holistic view at a reduced magnification is provided interactively, without severely compromising spatial relationships. The foundation for these techniques were presented in 1973 in the form of a digital enhancing and continuity retaining lens called *Graphi-*

cal Fish Eye (Farrand, 1973), to address what he termed the DETAIL X SCOPE problem in information display. Extending the concept, Kadmon and Shlomi (1978) presented a polyfocal projection for the presentation of statistical data on non-interactive cartographic maps. They laid the mathematical foundation for a variety of techniques.

It is not our intention to present an in-depth study of various distortion techniques but only to outline significant contributions and bring about awareness with respect to the availability of a range of techniques available. Several noted examples are:

- Bifocal display (Spense and Apperley, 1982): A one dimensional form which involves a combination of a detailed view and two distorted side views, where items on either side of the detailed view are compressed uniformly in the horizontal direction.

- Fisheye view (Furnas, 1986): Suited for a hierarchical structure, where elements are assigned priority (relevance) and distance from the element under consideration and focus point in the structure. Subsequent selection of threshold value and the comparison of functions of these two numbers determine what information is to be presented or suppressed. Holland *et al.* (1989) and Mitta (1990) extended the work of Furnas.

- Perspective wall: A conceptual descendent of the bifocal display is presented by Mackinlay *et al.* (1991). The perspective wall is based on the notion of smoothly integrating detailed and contextual views to assist in the visualization of linear information.

- Hyperbolic visualization (Lamping *et al.,* 1995; Munzner, 1995,1997; Munzner *et al.*, 2000): Lays out the hierarchy on a hyperbolic plane and maps this plane onto a circular display region. The hyperbolic plane is a non-Euclidean geometry in which parallel lines diverge away from each other. This leads to the convenient property that the circumference of a circle on a hyperbolic plane grows exponential within its radius, which means that exponentially more space is available with increasing distance.

- Flip zooming (Holmquist and Ahlberg, 1997; Bjork, 2000): Data are broken into pages and flipped as in the pages of a book. The pages in focus are zoomed in, while those out of focus are displayed as thumbnail sketches, facilitating ease of access to the users to navigate through large data spaces.

Leung and Apperley (1994) outlined a taxonomy of distortion-oriented techniques, which classified them into continuous and non-continuous techniques, with bifocal and perspective wall assigned to the latter class, and fisheye view and polyfocal display to the former. Furthermore, Leung and Apperley broadly classified distortion and non-distortion techniques based on the characteristics of the data, as being inherently graphical in nature, with implicit spatial relationships or whether it is non-graphical in nature that facilitates representation in abstract graphical form. This taxonomy of presentation techniques for large graphical data spaces is illustrated in Figure 2.12.

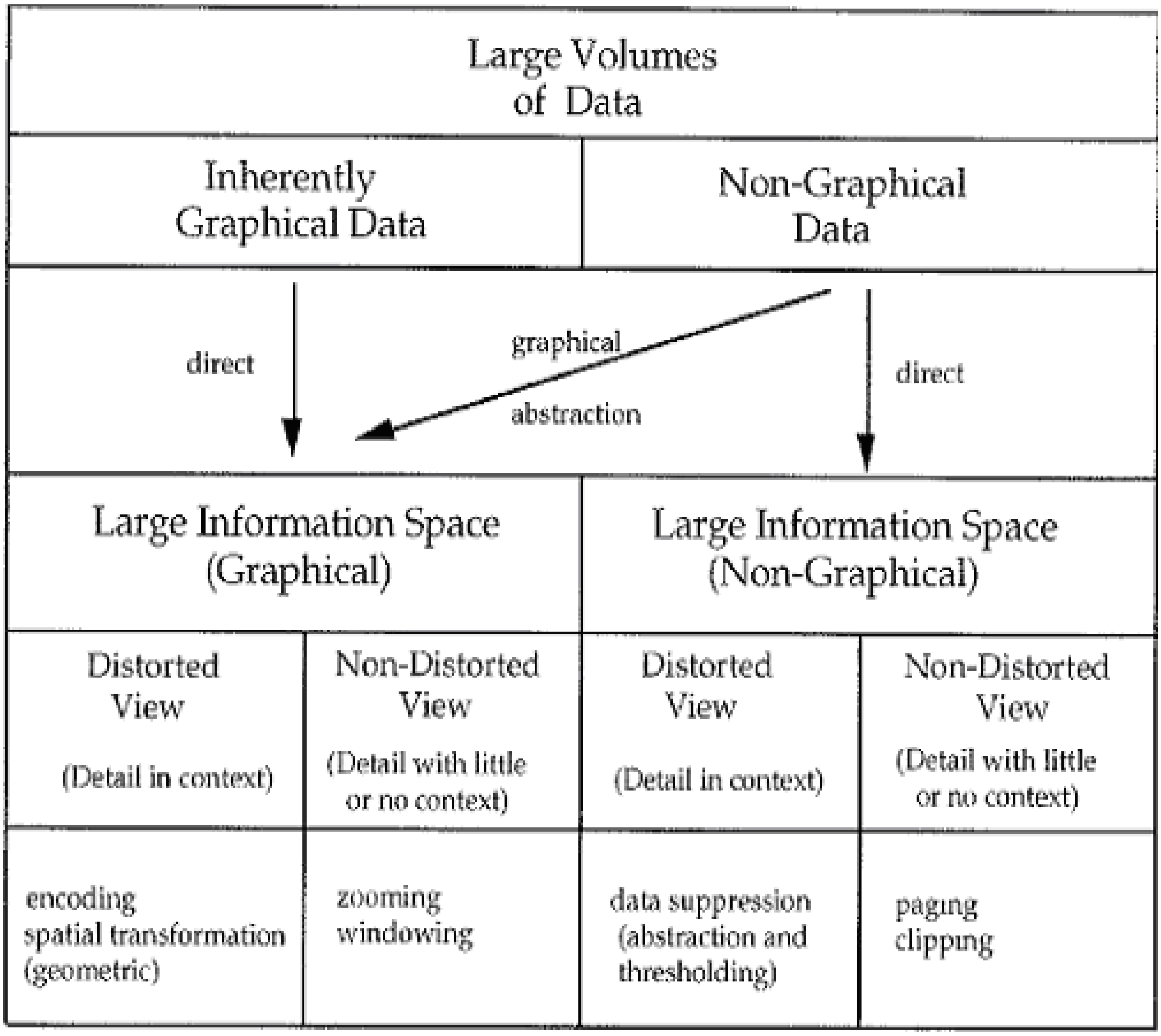


Figure 2.12: Taxonomy of presentation for large data spaces (Leung and Apperley, 1994; p. 127)

A combination of distortion techniques with map display, filtering, brushing and linking facilitate better interactivity subjective to the needs. A few examples are illustrated in Figures 2.13 to 2.17. Note that Figures 2.13 to 2.16 represent a particular form of distortion technique called 'cartogram', which we also intend to use in our experiments. Dorling (1995; p. xiv) defines cartograms as follows: "*A cartogram is a map or diagram showing geographical and statistical information. More specifically, the word cartogram has been defined as a 'combination of map and graph'* (Wilkie, 1976, p.1). *In fact, ordinary maps are equal land area cartograms, equal areas of space being allocated on the map sheet to equal areas of land*". Table 2.5 provides an overview of commonly known cartogram types. Note that Figures 2.13 to 2.16 all show examples of contiguous cartograms. We will see instances of Dorling cartograms in Chapter 4.

Table 2.5: Cartogram classification (Dent, 1998)

1.	Non-contiguous cartogram	The connectivity of geographic adjacent units is not preserved in non-contiguous cartograms. By freeing the objects from their adjacent object, they can grow or shrink in size and still maintain their shape. They are sometimes also referred to as shape preserving cartograms.
2.	Contiguous cartogram	In this case of cartograms, the adjacency is maintained (the objects remain connected with each other) which causes great distortion in shape.
3.	Dorling cartogram	Named after Danny Dorling. They maintain neither shape, topology nor object centroids. Objects are replaced with a uniform shape, usually a circle, of the appropriate size, for the reason that the shapes do not overlap. A similar approach is adopted with squares, thus leaving fewer gaps between the shapes.
4.	Pseudo cartogram	This is proposed by Tobler (1986,2004) and doesn't conform to any of the cartogram rules, but they resemble cartograms. The connectivity between objects is moved to a reference grid in order to give the same effect. Least square adjustment is used to find the "best fit" or cartogram that "is close enough"

Rescaling:

A graphic interaction technique which facilitates scaling back to an initial level of detail can be termed re-scaling or re-setting. It can be viewed analogous to an action of rollback, wherein users return to their initial state of interaction. This usually occurs when users wish to return the overview (initial state) after a zoom-in operation. Hence, this term refers to re-setting the scaling parameters of the data display.

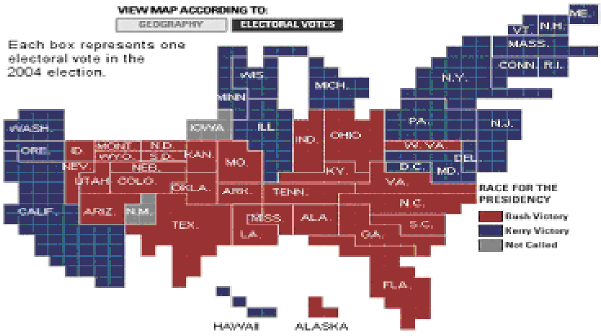


Figure 2.13: U.S Federal Election (http://www.nytimes.com/packages/html/politics/2004_ELECTIONRESULTS_GRAPHIC/)

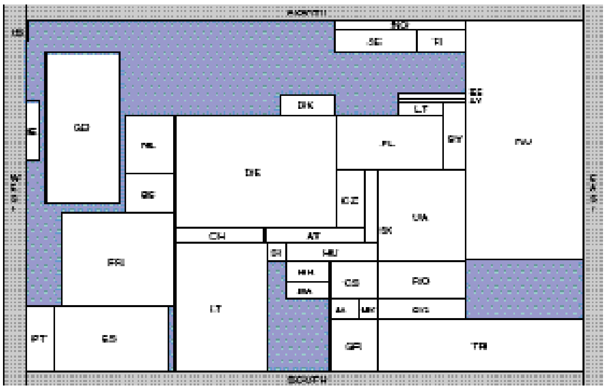


Figure 2.14: The population of Europe (country codes according to ISO 3611) (van Kreveld and Speckmann, 2004)

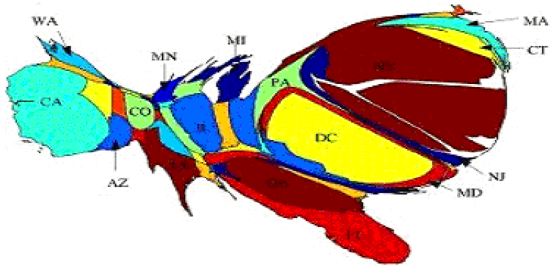


Figure 2.15: Geographical distribution of news stories. Newman and Gastner show (http:// www.sciencenews.org/articles/20040828/bob8.asp)

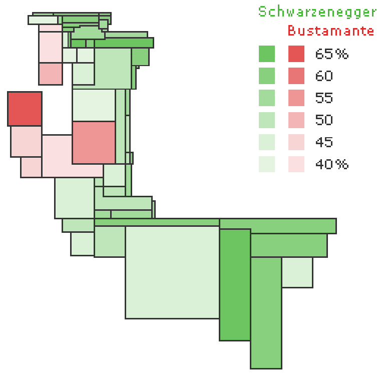
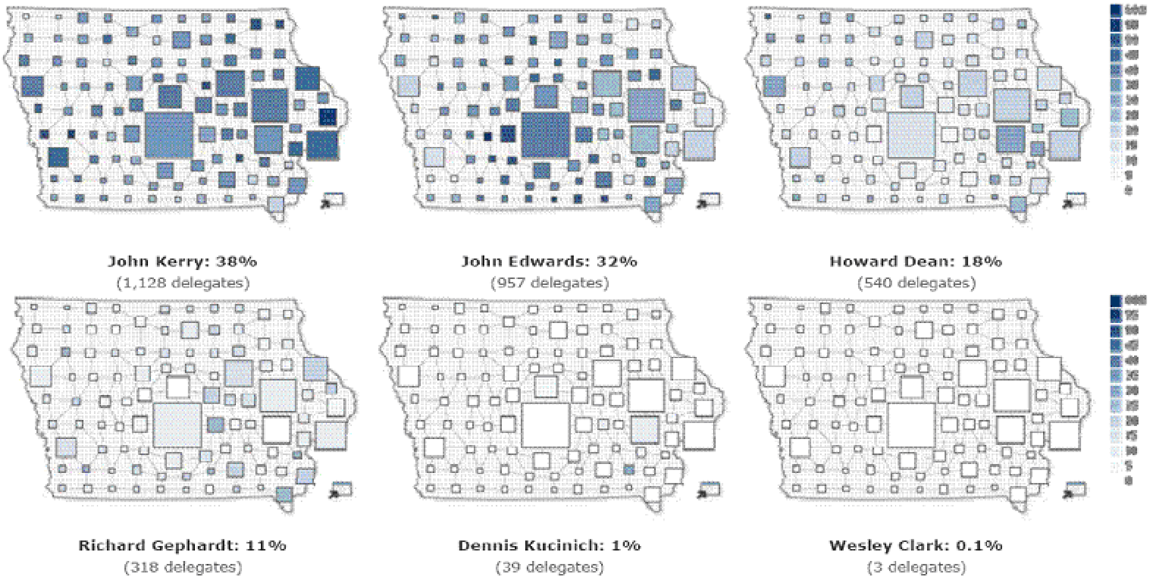


Figure 2.16: % of votes by county cast for the leading candidate (Corum, 2004. http:// www.style.org/mappingvotes/)



John Kerry: 38% (1,128 delegates)

John Edwards: 32% (957 delegates)

Howard Dean: 18% (540 delegates)

Richard Gephardt: 11% (318 delegates)

Dennis Kucinich: 1% (39 delegates)

Wesley Clark: 0.1% (3 delegates)

Figure 2.17: State delegate equivalents and the 2004 IOWA caucuses. Individual results per county are scaled in a checker board pattern. (Corum, 2004. http://www.style.org/iowacaucus/)

Remapping symbols:

In a user-centred environment, provision to customize the representation would help the users perceive the information presented in a better way. This ability to have customized symbolization and colouring schemes helps users to better perceive the information presented (Brewer, 1990). One such highly valued tool is ColorBrewer, designed to help people select good colour schemes for maps and other graphics (Harrower & Brewer, 2003).

2.3.3.2.3 Interaction with temporal dimension

Representing data and processes that involve both space and time poses a formidable challenge. In addressing this intriguing challenge, dynamic representations and interactions can help to create spatio-temporal representations that are cognitively informative to the user. These cognitively informed representations involve the depiction of change over a single time-space continuum that can be visually explored to trigger domain knowledge of dynamic phenomena (Kraak & Orneling, 2003). “*Representation variables in geodata context are interchangeable signs or signals that can be used to symbolize aspects of the data in perceptual form. Symbolization may vary from realistic to ab-*

stract; symbols may deliberately transform or exaggerate data characteristics or exhibit special effects (e.g. blinking spots or symbols that change their characteristics upon mouse-overs.)" (Blok, 2005, p. 61). DiBiase *et al.,* (1992) and MacEachren (1992) first introduced the notion of dynamic variables apart from those (static) visual variables presented by Bertin (1983).

Blok (2005) presents an exhaustive study of dynamic visualization variables evolved to date and outlines a set of variables of temporal dimension in animated representations expressed in terms of concepts introduced by DiBiase *et al.*, (1992) and MacEachren (1992), and relates them to concepts introduced by other authors, as shown in Table 2.6. In addition to Bertin's original set of visual variables, four dynamic visualization variables (*moment of display, order, duration, and frequency*) are defined and distinguished considering a design perspective and effects of the usage of these variables.

Table 2.6: Variables of the temporal dimension in animated representations expressed in concepts introduced by DiBiase *et al.*(1992) and MacEachren (1994); Blok (2005, p. 64, Table 4.1)

Concepts introduced by DiBiase et al. (1992) and MacEachren (1994b)	**Concepts used by other authors:**				
	Hayward (1984)	**Magnenat-Thalmann & Thalmann (1990)**	**Shepard (1994)**	**Green (1999)**	**Wilkinson (1999)**
Moment of display	none	none	timing	none	none
Order	scene (effects)	sequence	none	motion direction	motion direction
Duration	speed of movement	none	duration	motion velocity	motion speed
Frequency	none	none	motion/position change; blinking	frequency (flicker and phase)	none
Rate of change	none	none	none	none	motion acceleration
Synchronization	none	none	none	none	none

Navigation:

Navigation can be seen as the process of plotting a course of exploration involving all levels of interactivity to achieve specific objective within an information seeking environment, be it dynamic or static in nature. Navigation within the digital spatial realm has never been an easy task and often the user's perception of space differs from the one presented by the system. Visualizations in the recent past have become increasingly ubiquitous as a way to increase the engagement between users and the information to be explored. Figure 2.18 illustrates America's richest on a time scale, where in user's can navigate through this time scale revealing the statistics of each region.

Fly-through:

Fly-throughs are a sophisticated form of navigation where in users can choose to virtually walk or fly through selected consecutive sequences of views (called frames) of a terrain scene that give the appearance of moving through the real landscape (Crampton, 2002, p.93)

Toggling:

A technique of switching back and forth between time periods to highlight changes or to restore a state is termed 'toggling' (Crampton, 2002, p.93). Figure 2.19 illustrates a map depicting the frequency of conversations on Topix.net for the US based on time line and geo-categorization; the news topics can be toggled.

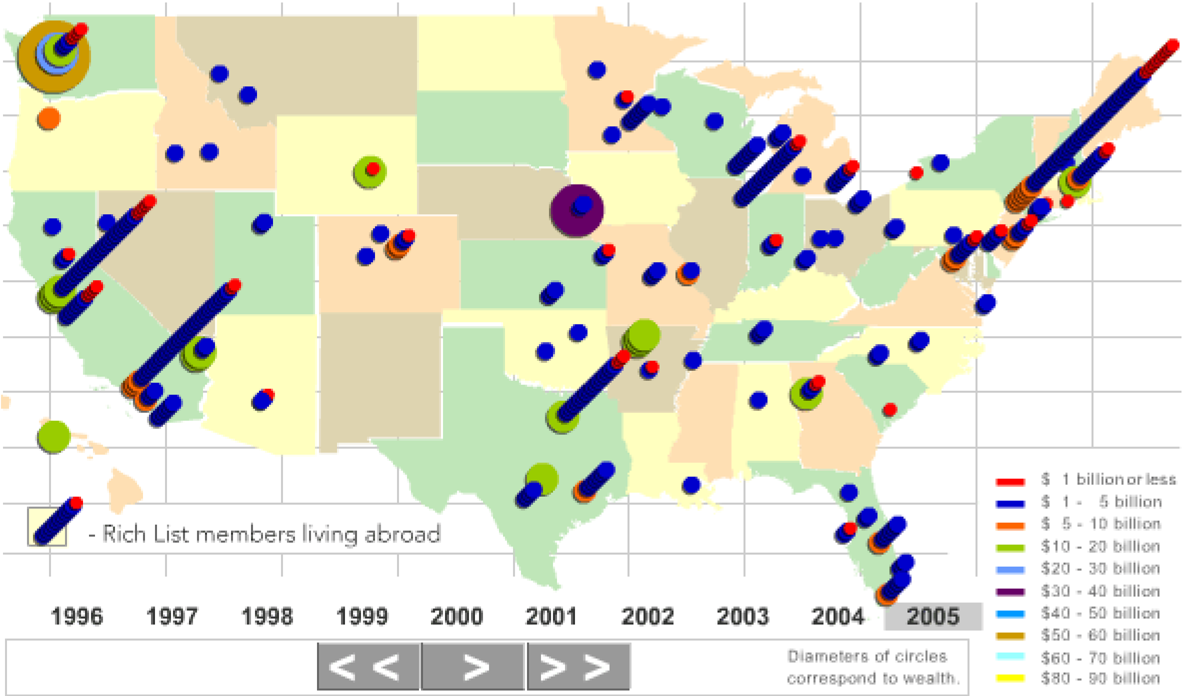
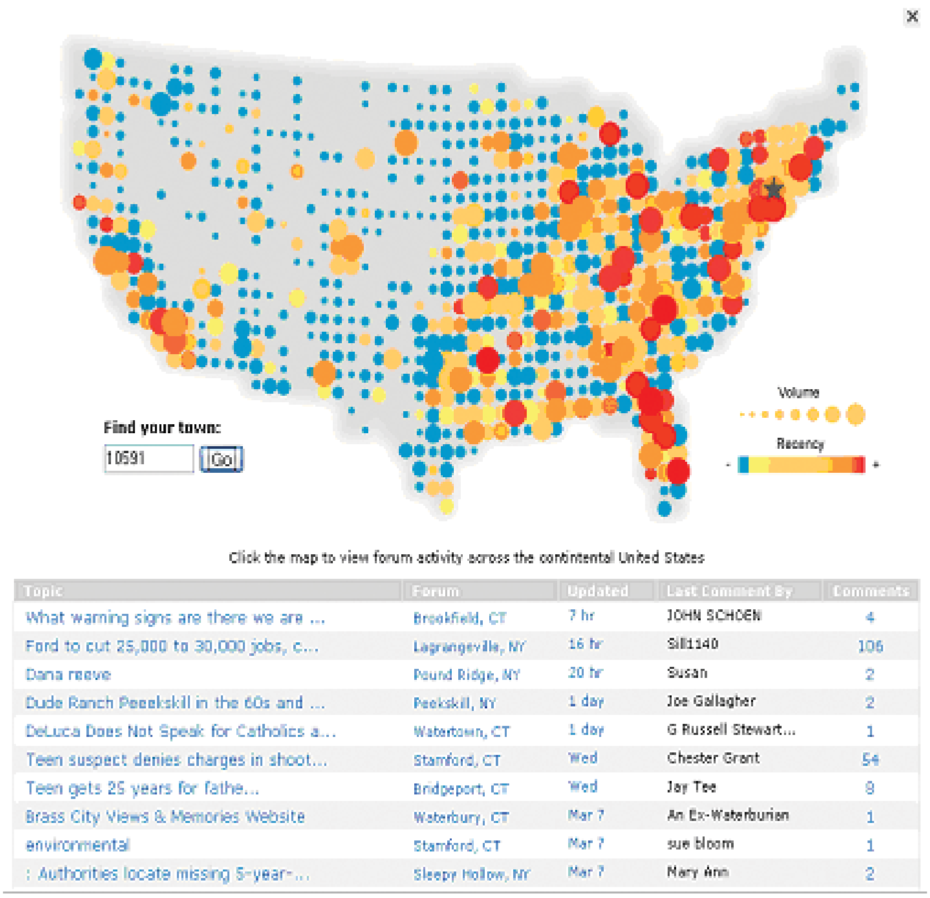


Figure 2.18: An illustration of America's richest – Hometowns of the richest 400 (http://www.forbes.com/2005/09/15/hometowns-networths-america-richest_05rich400_map.html)



Click the map to view forum activity across the contintental United States

Topic	Forum	Updated	Last Comment By	Comments
What warning signs are there we are ...	Brookfield, CT	7 hr	JOHN SCHOEN	4
Ford to cut 25,000 to 30,000 jobs, c...	Lagrangeville, NY	16 hr	Sill140	106
Dana reeve	Pound Ridge, NY	20 hr	Susan	2
Dude Ranch Peeekskill in the 60s and ...	Peekskill, NY	1 day	Joe Gallagher	2
DeLuca Does Not Speak for Catholics a...	Watertown, CT	1 day	G Russell Stewart...	1
Teen suspect denies charges in shoot...	Stamford, CT	Wed	Chester Grant	54
Teen gets 25 years for fathe...	Bridgeport, CT	Wed	Jay Tee	8
Brass City Views & Memories Website	Waterbury, CT	Mar 7	An Ex-Waterburian	1
environmental	Stamford, CT	Mar 7	sue bloom	1
: Authorities locate missing 5-year-...	Sleepy Hollow, NY	Mar 7	Mary Ann	2

Figure 2.19 Topix.net Conversation Map for the contiguous states of the USA (http://www.topix.net/blogs)

Sorting or re-expression:

Sorting or re-expression can be characterized as selective segregation or arranging items of the same kind, class or nature, etc. in some ordered sequence. By this process items with similar properties are grouped or labelled together, facilitating the ease of access for the end users. Figure 2.19 illustrates an example of sorting news articles based on location for a specific region in the USA.

2.3.3.2.4 Contextualizing interactions.

In geovisualization environments, where usually multi-faceted problems are explored, the understanding of context is crucial. Context can be referred to as how the user selects the relevant data or compares two different data sets of interest. If more than one theme or statistical variable has to be packed into a single map, the map will quickly become complex and possibly overburdened. Fig 2.20 illustrates one such example that attempts to depict, in a single map, two interrelated statistical variables, namely the percentages of votes by county cast for Schwarznegger vs. Bustamante in the 2003 election for Governor of California. While this map is highly sophisticated and differentiated, it is also quite complex to read. One way to deal with high levels of complexity is by interactions, which allow the user to explore the context of a map and to visualize the data from different perspectives. The interactions that will be discussed below include multiple views and combination of data layers.

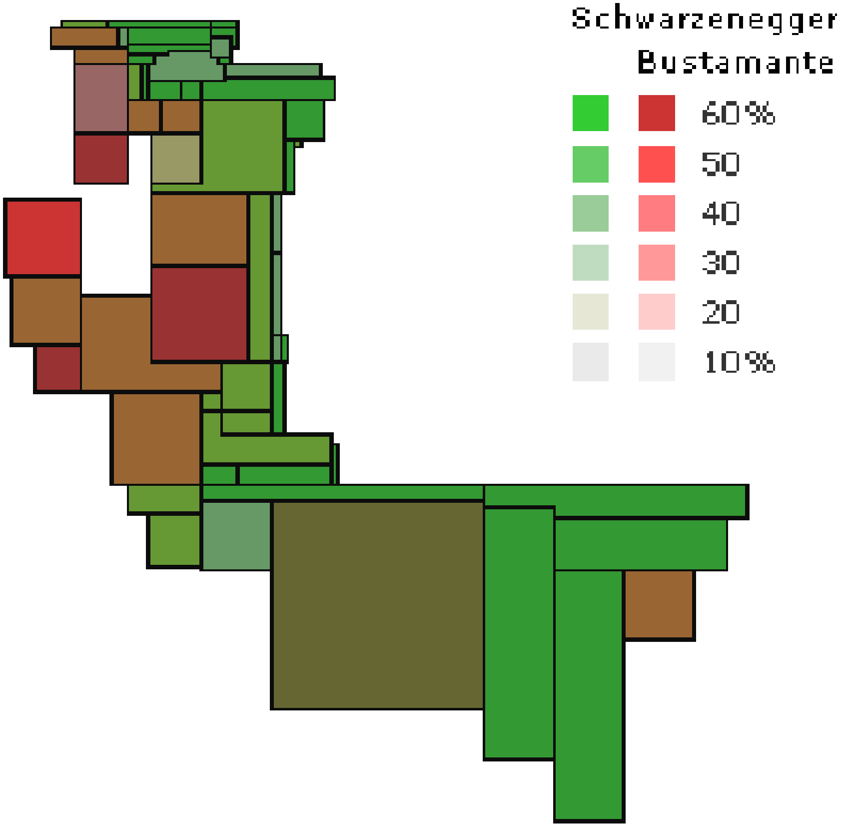


Figure 2.20: Percentages of votes cast by county for Schwarzenegger vs. Bustamante in the 2003 election for Governor of California [Source: http://www.style.org/mappingvotes/]

Multiple views:

Multiple views are often used when unforeseen relationships of a single conceptual entity need to be discovered (Shneiderman, 1987). The process is either sequential or simultaneous in nature. Buja *et al.* (1996, p.83) state that the sole "*purpose of arrangement of views is to facilitate meaningful comparisons*". Baldonado *et al.* (2000) identify the underlying complexities involved in coordination and subtle interactions among many dimensions of design space.

In the process of design, one should bear in mind the cognitive aspects involved in adopting multiple views within the system. Baldonado *et al.* (2000, p.11) point out that "*multiple views can provide utility in terms of minimizing some of the cognitive overload engendered by a single, complex view of the data. On the other hand, multiple views can decrease utility when added to the system, both in terms of higher cognitive overload (e.g. for context switching)*". By cognitive overhead is meant the effort in comparing a single conceptual entity in relation to others or to other dimensions. These interactions necessarily involve a higher degree of cognitive processing and computational requirement. The work by Baldanado *et al.* has identified the important dimensions of multiple view models: *selection of views, presentation of views*, and *interaction among views*. Furthermore, their work extends to outline four design rules that help to decide when multiple views may be adopted as an integral part of the system, including rules about diversity, complimentarity, parsimony, and decomposition (Baldanado *et al.,* 2000). Furthermore, they define four additional rules about how to use multiple views, including space/time resource optimization, consistency, self-evidence, and attention management. These eight rules are summarized in Table 2.7.

Table 2.7: Summary of rule and areas of major impact on utility of multiple views (Baldanado *et al.*, 2000; p. 112, Table 1)

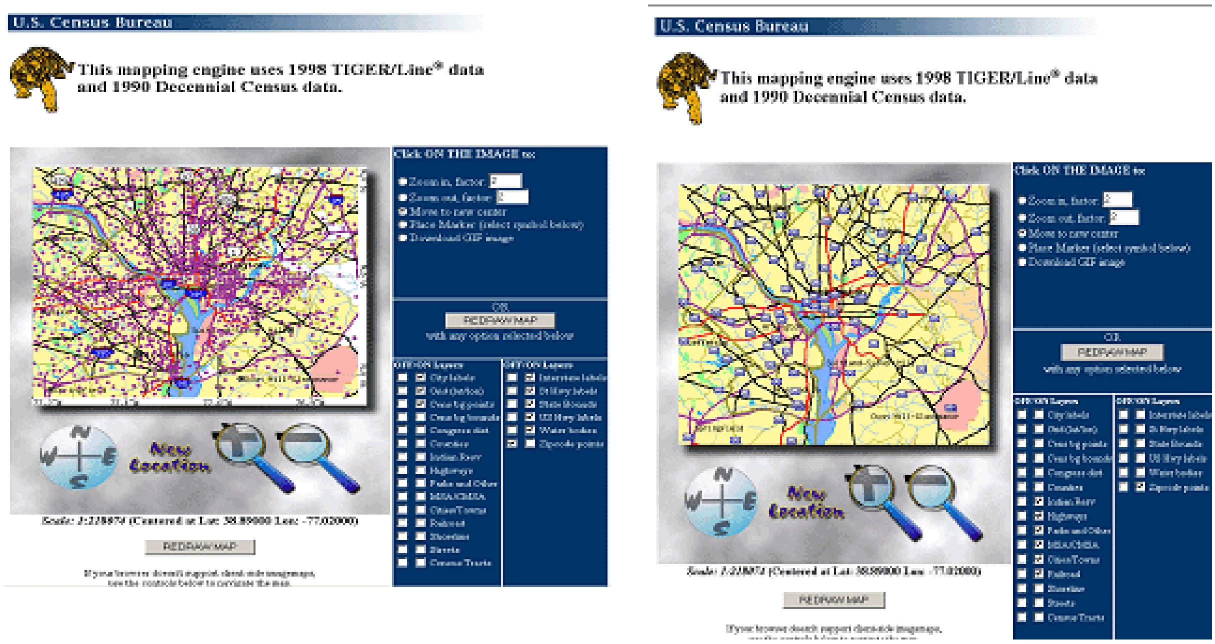
Rule	**Summary**	**Major positive impacts on utility**	**Major negative impacts on utility**
Diversity	To use multiple views when there is a diversity of attributes, models, user profiles, levels of abstraction or genres	Memory	Learning computational overhead display space overhead
Complimentarity	To use multiple views when different views bring out correlations and or disparities	Memory comparison context switching	Learning computational overhead display space overhead.
Decomposition	To partition complex data into multiple views to create manageable chunks and to provide insights into the interaction among different dimensions.	Memory comparison	learning computational overhead display space overhead
Parisomony	To use multiple views manually	Learning computational	Memory comparison context switching
Space/Time resource optimization	To balance the spatial and temporal costs of presenting multiple views with spatial and temporal benefits of using the views	Comparison computational overhead, display space overhead	
Self-evidence	To use perceptual cues to make relationships among multiple views more apparent to the user	Learning comparison	Computational overhead
Consistency	To make the interface for multiple views consistent, and make the state of multiple views consistent	Learning comparison	Computational overhead
Attention management	To use perceptual techniques to focus the user's attention on the right view at the right time	Memory switching	Computational overhead

Gleaning through the literature to identify prominent examples of the usage of multiple views sheds light on the works of North *et al.* (2000, 2002) who present a "*snap-together-visualization allowing users to construct customized, coordinated multiple views that fit their needs*"; Buja *et al.* (1996, 1998) with the XGobi and XGvis systems; Dai & Hardisty (2002), MacEachren (2003) who proposed a multiform bivariate matrix map showing the relationship between four variables for counties; and Bavoil *et al.* (2005) with the VisTrails system that enables interactive creation and maintenance of multiple view visualizations.

Combining data layers:

Users are often intrigued by questions about which phenomena are to be examined and how the analysis needs to be carried out when attempting to answer a geographic question. Some interactive geovisualization environments enable users to perform functions of spatial analysis such as projection and spatial transformation utilities; spatial retrieval classification and measurements; logical and visual overlaying; proximity and network functions; map algebra utilities; and output generation (Falbo *et al.*, 2002). Such analysis functions are performed by considering attribute and map (geographic) data, and are often supported by combining and overlaying (or merging) different visualizations of the data sets in order to deduce results of specific interest to the user (Crampton, 2002).

Notable exemplars of such interactive environment are the US Census Bureau's interactive mapping tool as shown in Figure 2.21; *USGS National Map Viewer (http://nmviewogc.cr.usgs.gov/viewer.htm), Geospatial One-stop (http://gos2.geodata.gov/wps/portal/gos), and the Hurricane Disaster Viewer (http://apps.arcwebservices.com/sc/hurricane_viewer/index.html).* The surge of interest in GIS and geography in general has been acknowledge by Nature's article on Geographic Mashups (Butler, 2006) which sheds light on merging two or more geoweb services and the integration of disparate data sets using geography as the key fundamental reference. Such integration also makes use of the combination of data layers.



U.S. Census Bureau

This mapping engine uses 1998 TIGER/Line® data and 1990 Decennial Census data.

Figure 2.21: US Bureau of the Census, Tiger Map Server: An exemplar illustrating the combination of data layers. The maps illustrate the overlay of different data layers superimposed on the map in order to infer the results of interest for the user. (Source: http://tiger.census.gov/cgi-bin/mapbrowse-tbl)

2.4 Taxonomy of user tasks for interaction

Mapping to the envisioned system requirements, a critical task analysis in essence can enable rapid construction of task descriptions with accompanying elegant task diagrams. Enabling designers to rapidly sketch task sequences that can be summed up to use-case scenarios. Shneiderman (1996) proposes a type by task taxonomy of information visualizations, where he assumes that users are viewing collections of items, where items have multiple attributes. In all seven data types listed below (1-, 2- , 3-dimensional data, temporal and multi-dimensional data, and tree and network data) the items have attributes and a basic search task is to select all items that satisfy values of a set of attributes. The seven tasks are at a high level of abstraction and they are:

Overview: Gain an overview of the entire collection.
Zoom: Zoom in on items of interest.

Filter: filter out uninteresting items.
Details-on-demand: Select an item or group and get details when needed.
Relate: View relationships among items.
History: Keep a history of actions to support undo, replay, and progressive refinement.
Extract: Allow extraction of sub-collections and of the query parameters.

Having learnt about the interaction techniques adopted in various visualizations, discussed above (Shneiderman, 1996), which serve as design guidelines for visualization. Pillat et al., (2005) criticize, however that they do not serve to represent a comprehensive framework for evaluation and usage guidelines.

A taxonomy based on use tasks for interaction (as shown in Figure 2.22) serves to assess the capabilities and efficiency of the techniques in visualization. Pillat *et al.* (2005) in their experimental studies on the evaluation of multidimensional visualizations adopted the taxonomy proposed by Valiati (2005). This taxonomy identifies seven tasks: *identify, determine, visualize, compare, infer, configure, and locate*. Of which i*dentify, determine, compare, infer* and *locate* are used to visually explore and analyse, while and the other two – *configure* and *visualize* – serve as supporting tasks.

Adopting a metaphor of a mathematical function, Andrienko & Andrienko (1996) identify the tasks involved in exploratory data analysis. Furthermore, they propose that a task consists of two parts: What information needs to be obtained? What conditions does this information need to fulfil? Andrienko & Andrienko (2006) then devote an entire chapter to the definition of a comprehensive list of tasks of exploratory spatiotemporal data analysis, further distinguishing between elementary tasks (referring to individual elements of the reference set) and synoptic tasks (involving the whole reference sets or subsets thereof).

2.5 Visualizing Text

2.5.1 Visualizing text document collections

In searching web documents the primary data retrieved is text. Many visualization proponents have developed techniques to browse different kinds of textual data. The Seesoft visualization proposed by Eick *et al.* (1992) enables mapping of text documents into an analytical visual form. A statistic associated with the text lines is mapped onto a colour ramp. Visualization of text to a 2-D display is accomplished by tiling or generating composite pages. A noted exemplar of this visualization is demonstrated in Figure 2.23.

For larger collections, attempts have been made to create a semantic 2-D map of the document space. A novel approach to these concerns has been presented by Lin (1991) using the idea of a self-organizing map originally developed by Kohonen (1990) to create a 2-D visualization of a large text collection. Several other visualization techniques have been proposed such as Web Book (Card *et al.*, 1996), Perspective Wall (Kohonen, 1995), Document Lens (Robertson & Mackinlay, 1991), Cat-a-Cone (Hearst *et al.*, 1997) and LightHouse (Leuski and Allan, 2000).

A novel approach to graphically represent the similarities amongst a corpus of documents based on word similarities and discovery of key topics in the text is presented by Miller *et al.* (1998) under the name TOPIC ISLANDS. Wise *et al.* (1995) developed SPIRE, designed with an objective to let users rapidly discover underlying relationships in a collection of documents by identifying pertinent documents, rather than wading through large volumes of text. The n-dimensional document signals are first clustered into groups of related documents; then projected into groups of related documents; and finally they are projected from n-space into 2-D space to create visual representations. The SPIRE system incorporates a number of visual representations, of which two are described and illustrated below:

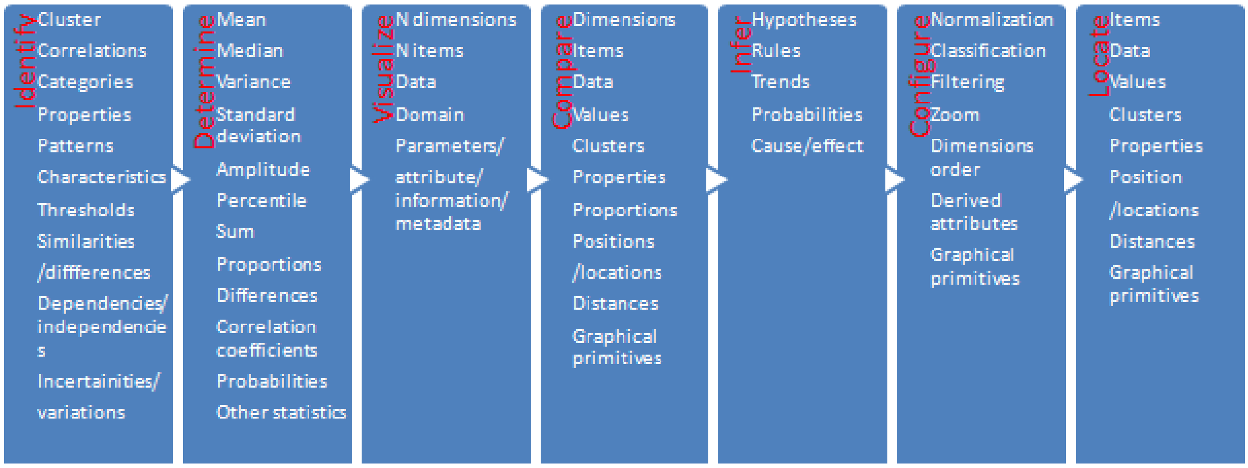
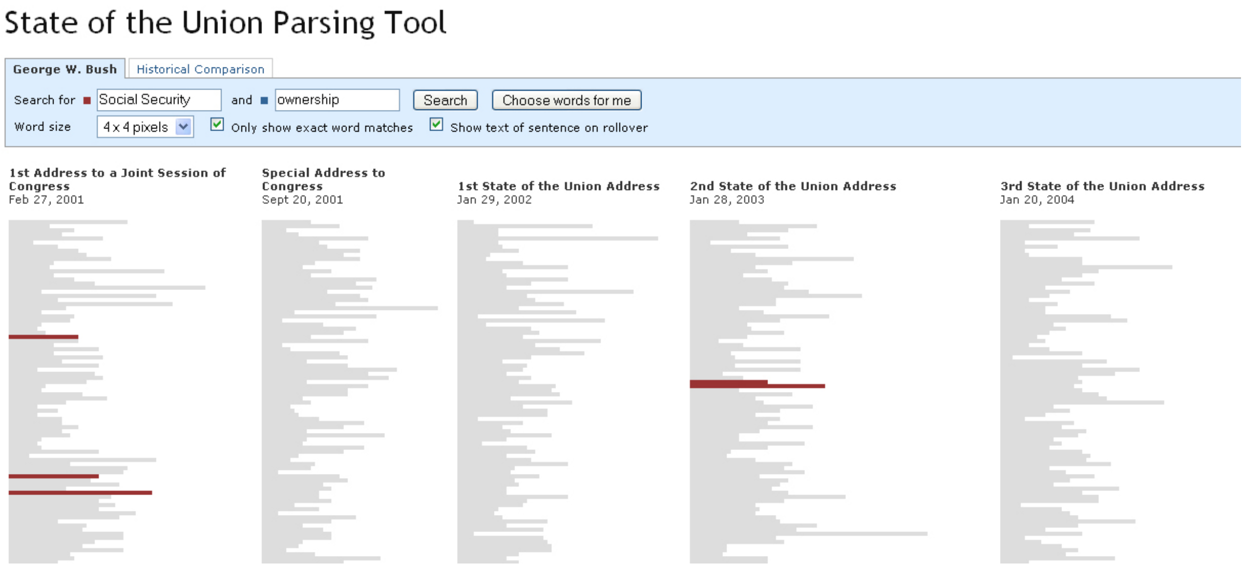


Figure 2.22: Taxonomy of user tasks for interaction with multidimensional visualization system (redrawn from Pillat *et al.*, 2005)


State of the Union Parsing Tool

George W. Bush | Historical Comparison

Search for Social Security and ownership

Word size 4 x 4 pixels ☑ Only show exact word matches ☑ Show text of sentence on rollover

Figure 2.23: State of the union parsing tool (http://www.style.org/stateoftheunion/parse/)

- Galaxies (Crow *et al.,* 1994) are based on a dimensional reduction technique called multidimensional scaling (MDS) with closely related documents clustered together in a 2-D projection. Galaxies display the documents as a universe of 'docustars'. Closely related documents cluster together in a tight group while unrelated documents are separated by large spaces (as shown in Figure 2.24).
- Themescapes (Wise *et al.,* 1995) produce a virtual terrain map of themes found in the corpus. Themes within the document spaces appear similar to a relief map of natural terrain. The mountains in the Themescape indicate dominant themes; valleys indicate weak themes. Themes that are close in content are visually close based on many relationships within the text spaces (as shown in Figure 2.25).

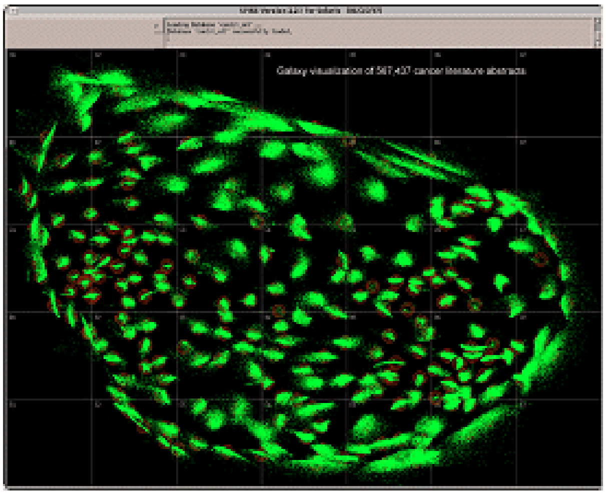


Figure 2.24: Galaxies visualization (Crow *et al.*, 1994)

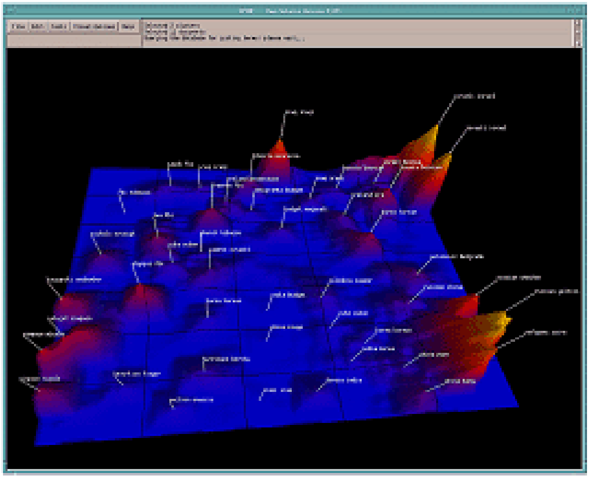
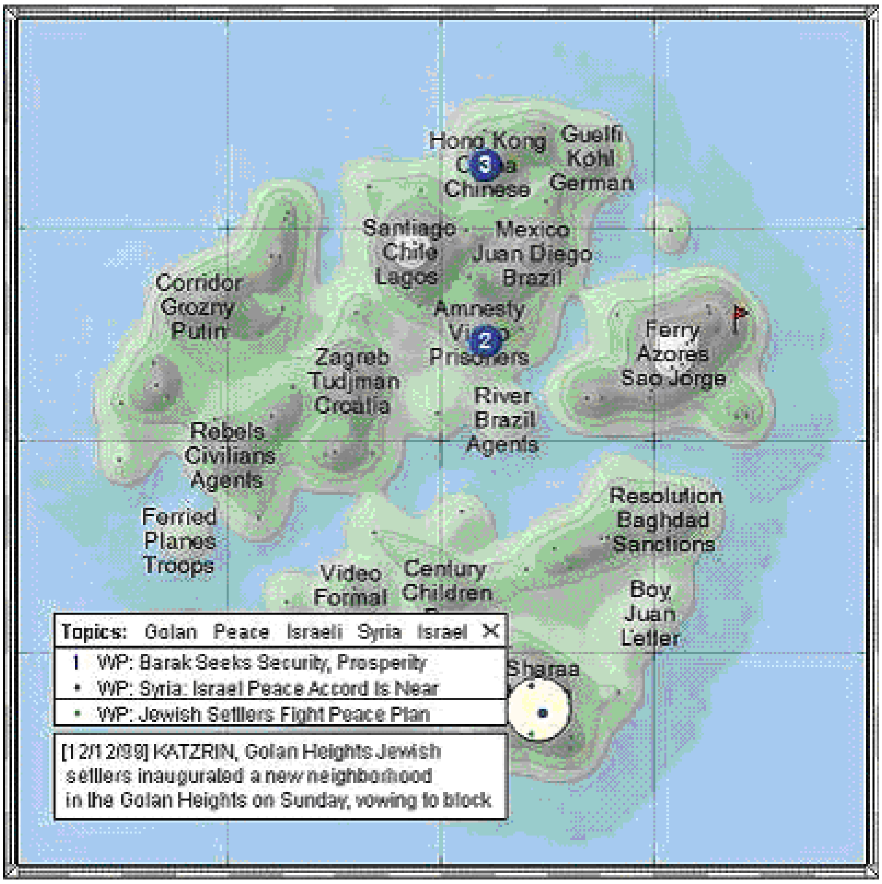


Figure 2.25: Themescape visualization (Wise *et al.*, 1995)

Cartia, Inc. demonstrated the use of the Themescape information analysis and mapping technology. These attractive and interactive NewsMaps provide a 'holistic' summary of large volumes of textual information represented as hills, valleys, and white, snow capped mountain peaks – a cartographic form common to topographic maps of the real world (Figure 2.26).



Topics: Golan Peace Israeli Syria Israel ✕

1 WP: Barak Seeks Security, Prosperity

• WP: Syria: Israel Peace Accord Is Near

• WP: Jewish Settlers Fight Peace Plan

[12/12/98] KATZRIN, Golan Heights Jewish settlers inaugurated a new neighborhood in the Golan Heights on Sunday, vowing to block

Figure 2.26: A typical newsmap using the Themescape technique (developed by Cartia, Inc. http://www.mundi.net/maps/maps_015/#ref_2)

2.5.2 Visualizing information hierarchies

Exploring information hierarchies can be achieved with tree views. Simple examples of such kinds of visualization are the Microsoft Windows Explorer, the Macintosh Finder, and the Java Jtree Component. This section discusses latest visual representations to view information hierarchies of relevance to this research.

A space filling approach to visualization of hierarchical information structures called Treemaps was proposed in the early 1990s by Johnson & Shneiderman (1991). Treemaps (as shown in Figure 2.27) are a screen-filling visualization of a hierarchy generated by alternating vertical and horizontal slicing of screen space. At any instance perhaps three to five levels of the tree are visualized. The interactive navigation allows the user to zoom-in and explore particular subtrees. A difficulty with treemaps is that overly broad trees (i.e. nodes having many subtrees) tend to generate overly narrow vertical or horizontal strips.

A similar approach of representation of hierarchy, using a three dimensional conical representation is presented by Robertson *et al.* (1991,1994) (as shown in Fig 2.28). In an effort to impart perception of depth, shadows are used for each node and for better reading the labels are placed on a horizontal layout. Animations are incorporated to facilitate transition and tracking of changes made by the user.

An extension of the Treemap was that proposed by Xerox PARC. The Hyperbolic Browser is a space-filling focus-and-context technique based on the principles of hyperbolic geometry, resulting in hyperbolic trees (as shown in Figure 2.29). A hyperbolic plane is generated, offering exponential space along each axis, and is then mapped to the unit disc in Euclidean space. With each child placing its children are compressed in a wedge of space. Functionalities available to manipulate the visualization in the Hyperbolic Browser include dragging the nodes around the display, bringing them into focus in the central area, and splaying the sub-tree like opening a fan (Lamping and Rao, 1994 & 1997; Lamping *et al.,* 1995).

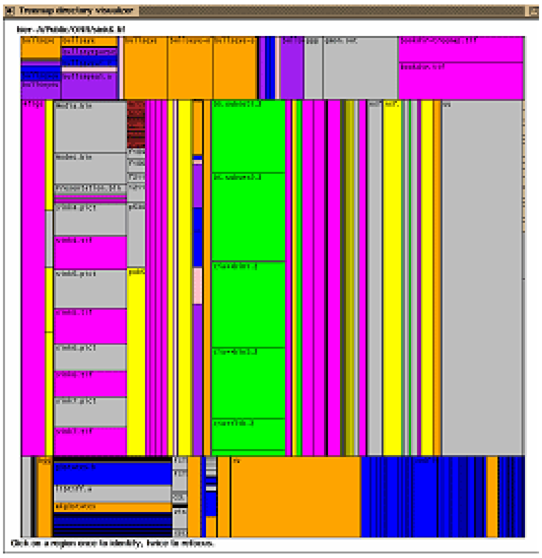


Figure 2.27: Treemap: a space filling approach to visualization of hierarchical information structures (Johnson & Shneiderman, 1991)

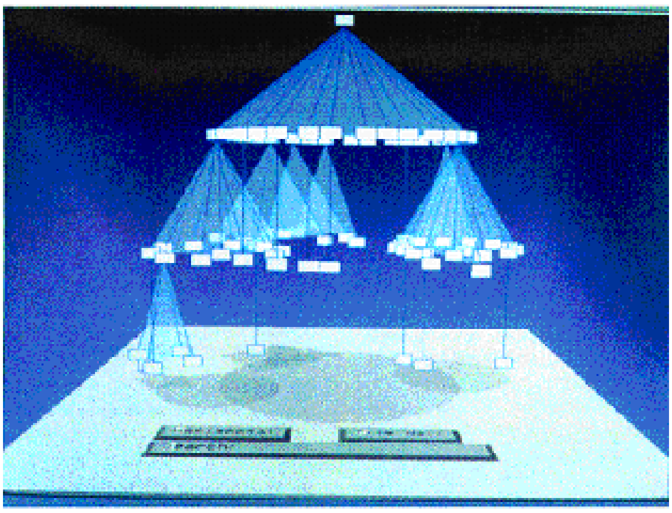


Figure 2.28: The Cone-Tree: a conical representation of hierarchy (Robertson *et al.*, 1991, 1994)

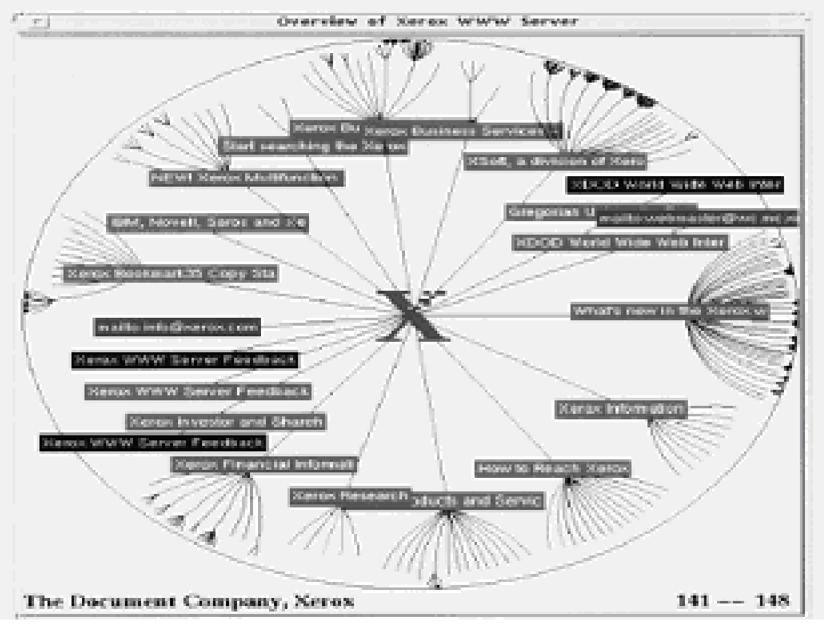


Figure 2.29: Hyperbolic Browser (Lamping & Rao, 1996)

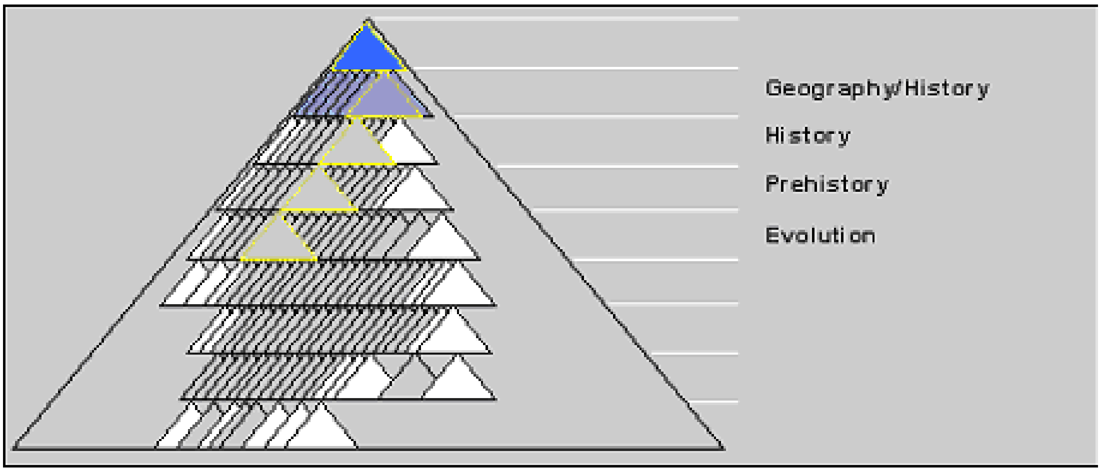


Figure 2.30: CHEOPS: A compact explorer for complex hierarchies (http://www.icg.tu-graz.ac.at/~hofer/mmis97/thh/TreeMaps/Cheops.htm)

Navigation systems are not only used for navigation through physical spaces, but also through cyberworlds of various kinds. Beaudoin *et al.* (1996) introduced CHEOPS (Compact Explorer for Complex Hierarchies; Figure 2.30). It uses overlapping triangles to provide a compact 2-D representation of hierarchy by overlaying (squashing together) children to save on screen space. Based on the selection of nodes (both parents and children) textual and visual information about the node is displayed in the far right panel. The location of a node is represented in a redundant fashion, that is, hierarchically, visually and also through textual labels.

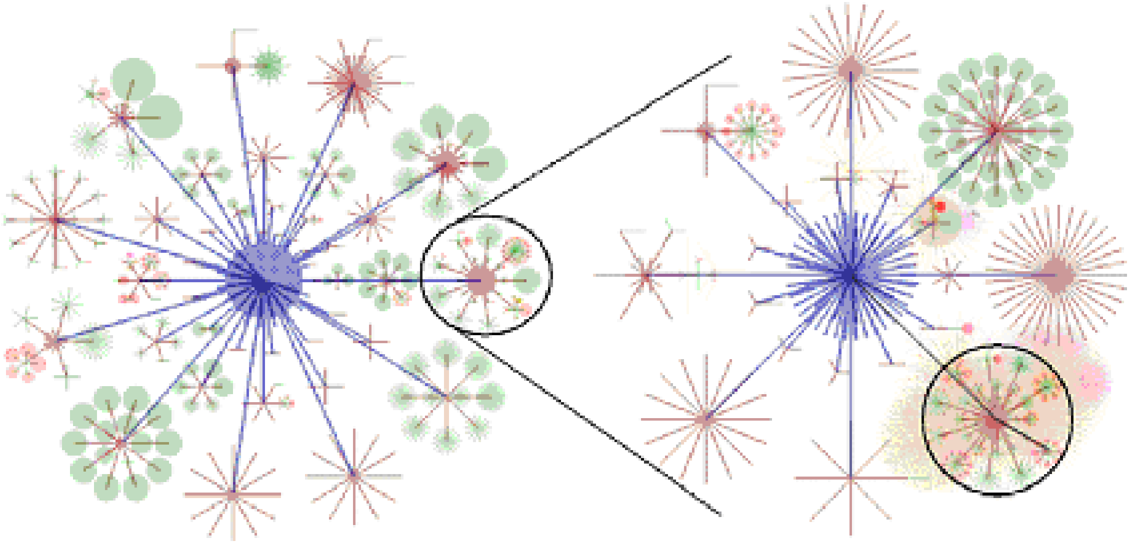


Figure 2.31: RINGS: Ringed interactive navigation graph system (http://www.cs.ucdavis.edu/~ma/papers/graph2.pdf)

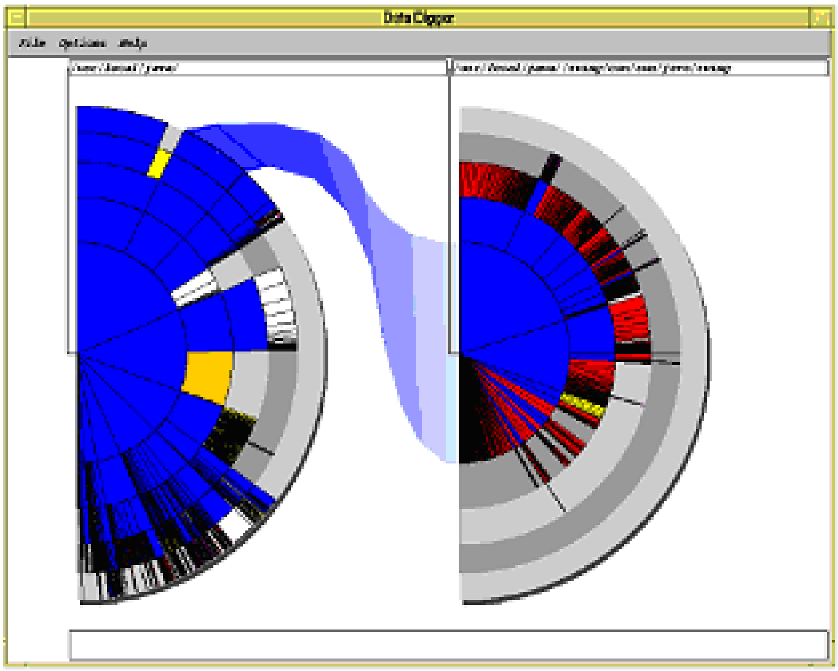


Figure 2.32: Information slices: expanding the swing/com/sun/java/swing subdirectory into the right-hand disc (http://www.iicm.edu/liberation/iicm_papers/ivis98.pdf)

RINGS (Figure 2.31) stands for Ringed Interactive Navigation Graph System and is a technique proposed by Soon TeeTeoh and Kwan-Liu Ma (2002) for visualizing large hierarchical data, as well as large graphs if a spanning tree of the graph is provided as input. The authors claim that their proposed visualization makes efficient use of limited display sizes by showing more distinguishable nodes at one time, without any of the other visualization methods' shortcomings. They also claim that their visualization brings in more contextual information without compromising clarity of the area in focus. RINGS also provides interactive navigation and focus-and-context solutions implemented in many existing systems.

Andrews and Heidegger (1998) propose another interesting method of visualizing hierarchies called *Information Slices* (Figure 2.32), using a series of semi-circular discs to compactly visualize large hierarchies in two-dimensional space. Each disc represents multiple levels of hierarchy; typically between five or more levels are visualized on each disc, the exact number being user configurable. Deeper hierarchies are represented using a cascading series of discs. At each level of hierarchy, the children are fanned out in the available space according to the total size of each child.

In an attempt to address the relationships between hierarchies as well as the relationships between multiple entities within a hierarchy, a new information structure called 'Polyarchies' was proposed by Robertson *et al.* (2002). Polyarchy visualization addresses the poly-hierarchy problem as well as more generic fundamental problems of hierarchical visualization, offering both the ability to scale and the ability to focus on multiple selections, without the loss of context. This has been termed 'visual pivoting', wherein two hierarchies are displayed simultaneously around a designated entity. Intersections of hierarchies share at least one node rather than sharing sub-trees. Robertson *et al.* (2002) claim to have addressed the issues relating to scale, flexibility and responsiveness. However, there are certain drawbacks to this approach, most importantly that of added complexity and potential confusion, as well as failure to address issues of ordering hierarchy and text rendering.

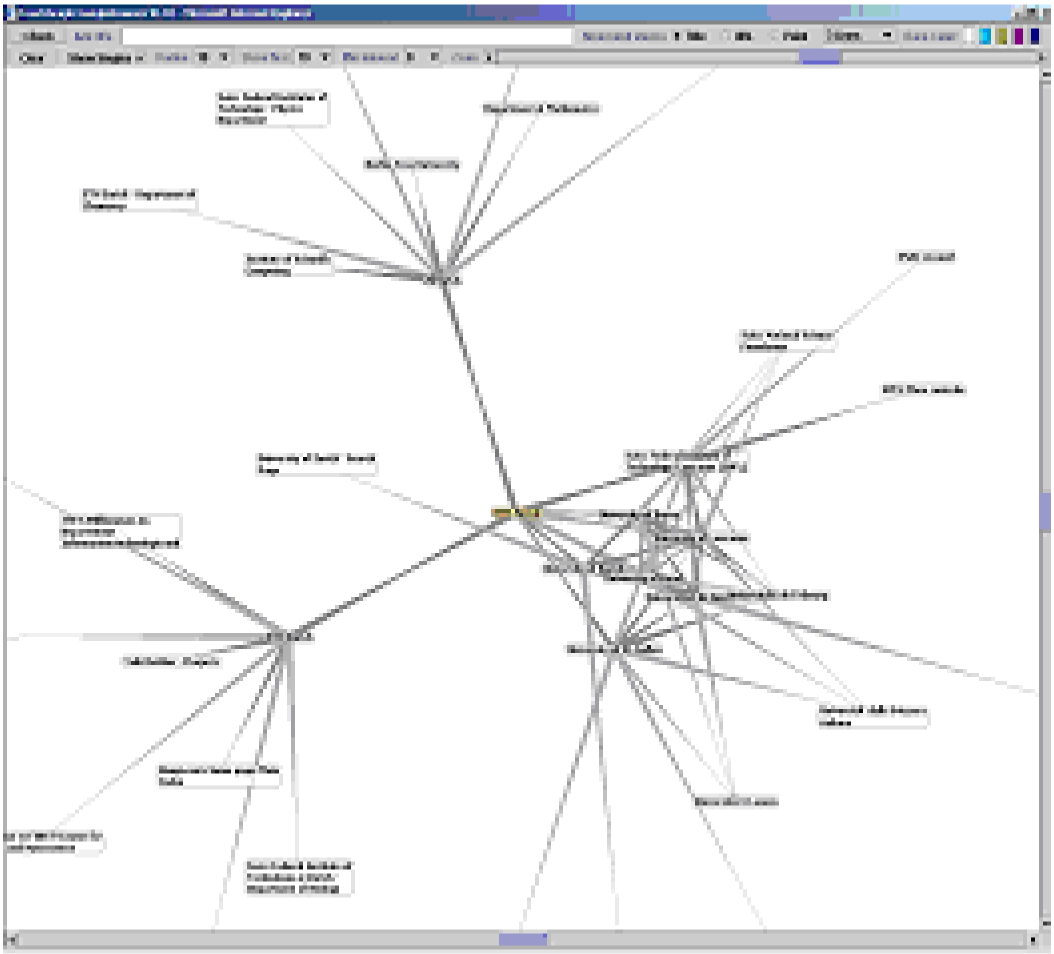
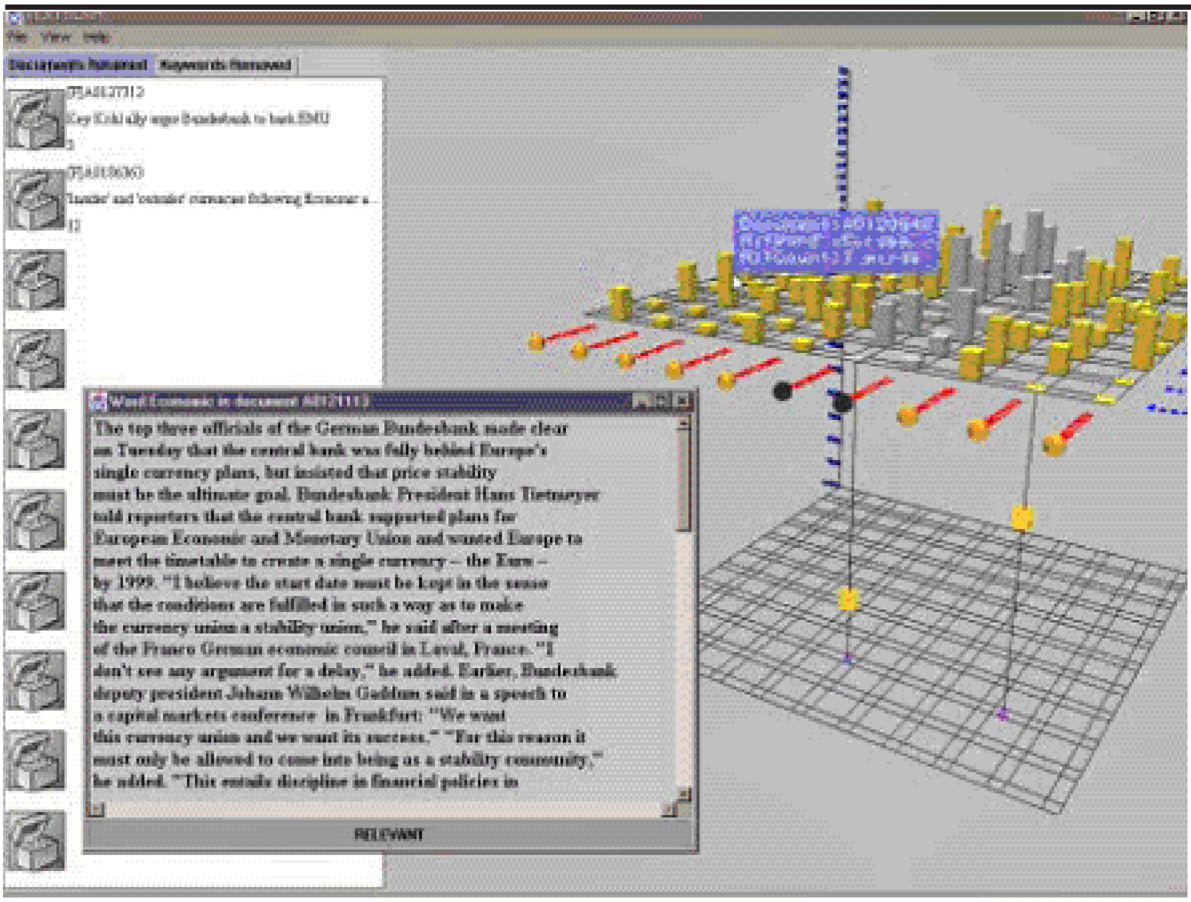


Figure 2.33: Touchgraph visualization (http://www.touchgraph.com/TGGoogleBrowser.html)



The top three officials of the German Bundesbank made clear on Tuesday that the central bank was fully behind Europe's single currency plans, but insisted that price stability must be the ultimate goal. Bundesbank President Hans Tietmeyer told reporters that the central bank supported plans for European Economic and Monetary Union and wanted Europe to meet the timetable to create a single currency – the Euro – by 1999. "I believe the start date must be kept in the sense that the conditions are fulfilled in such a way as to make the currency union a stability union," he said after a meeting of the Franco German economic council in Laval, France. "I don't see any argument for a delay," he added. Earlier, Bundesbank deputy president Johann Wilhelm Gaddum said in a speech to a capital markets conference in Frankfurt: "We want this currency union and we want its success." "For this reason it must only be allowed to come into being as a stability community," he added. "This entails discipline in financial policies in

Figure 2.34: Textscape (Rossi, 1999; Fig. 2)

Another web based visualization tool for ranking similarity is called Touchgraph (Figure 2.33). This tool has attracted increased attention within the visualization community for its capabilities to visualize the relationship between the query and resulting item set as a graph. It visually analogs items and similar pages. Koshman (2004) claims that this approach has the potential to enhance web search effectiveness based on similarity judgements. Rossi (1999) demonstrated Textscape (shown in Figure 2.32) as a 3-D prototype to identify similarities and differences within documents based on an extended cityscape design. Textscape is advocated to facilitate users with capabilities of visual data mining and cues (keyword clustering through threaded tiles), intuitive dataOIDs (unique data identifiers to interrogate for details), and dynamic scene changes linked to data feeds.

2.6 Role of Cognition

2.6.1 Cognition and spatial cognition

Cognition can be defined as

- *a comprehensive understanding built by recognition, discovery, awareness, rediscovery and exploratory activities carried out on visualizations.* (Guilford and Merrifield, 1960; Haynes, 1970, p.56)
- *as a barrier to understanding the basic features to which these functions refer to the clarification of the role of these underlying processes in total scheme of behaviour* (Haynes, 1970, p.55)
- *as the process that includes perception, learning, memory, thinking, reasoning, problem solving and communication* (Montello, 2002)

Cognition paves the way for the visual search processes adopted by the users in map reading. Card *et al.* (1999, p.16) identify six different ways in which information visualization can amplify cognition:

1. *by increasing the memory and processing resources available to users;*
2. *by reducing the search for information;*
3. *by using visual representations to enhance the detection of patterns;*
4. *by enabling perceptual inference operations;*
5. *by using perceptual mechanisms for monitoring;*
6. *by encoding information in a manipulable medium.*

Prior to examining the factors that affect the visualization of geographic information, it is important to understand the role of cognition. Haynes' (1970, p.58, Table:1) work sheds light on experimental studies detailed by Guilford and Merrifield (1960) to investigate the general cognitive ability of users. The studies detail that cognition is viewed as a complex process that is predicated on the interaction of an individual's sensory-motor and neurological systems.

Spatial Cognition is a vital element of general cognition, as it is the process by which an individual perceives, stores, recalls, creates and communicates about geographic spaces. McGee (1979, p. 893) relates this with spatial orientation and spatial visualization. Spatial cognition can be defined as

- *Inner space or spatial cognition, the spatial features, properties, categories and relations in terms of which we perceive, store and remember objects, persons, events, and on the basis of which we construct explicit, lexical, geometric, cartographic and artistic representations.* (Olson and Bialystock, 1983 p.2)

- *The underlying mental process that allows an individual to develop spatial abilities.* (Miller, 1996)

Numerous psychological studies have been carried out to identify factors affecting the spatial ability of an individual such as age, experience, gender, or individual differences (Miller, 1996; Liben, 1981; and McGee, 1979). The literature clearly established that an individual's cognition in general and spatial cognition in particular are governed by various factors. Hence, the role of cognition cannot be neglected, both in the design phase as well as during the evaluation phase, in order to establish how users perceive and respond to the visual representations that will be designed and tested during the course of our research.

2.6.2 Spatial awareness

Information seeking is the ultimate power of expression of an individual using a computer, looking at the world, and finding exactly the information that he/she wants. The semantic web is not only about web pages (documents) and links; it's about relationships underlying illustrated concepts, and their occurrence in space and time. It is not a mere challenge of recalling information after it's out on the internet. Search engines must help individuals to handle a body of information in a way that is conducive to action. In fact "*information is the relation that connects a sign with an intentionality*" (Couclelis, 1997). Software systems serve as mediums to cognitively construct information out of relationships between the outputs of the systems and the individuals' interest (Couclelis, 1998, p. 211).

"*Thus, in a space representing a corpus of documents (a 'document space'), the information objects will be the individual documents. Information objects are described by (a) intrinsic properties and (b) structural relations (Rennison and Strausfeld, 1995). The intrinsic properties of a document would be things such as author, title, date, subject matter, keywords, other important terms, and any other items appearing in a full bibliographic reference. They are the metadata explicitly coded into the system. The structural relations, on the other hand, emerge implicitly out of the affinities existing among information objects through their intrinsic properties*". Couclelis (1998, p. 212)

The review of search paradigms in information seeking earlier in this chapter (Section 2.2) suggests that ‘awareness’ (“Aware” in Table 2.3) is a form of information seeking that is undirected and passive in nature (Bates, 1986). The potential manifestations of structural relationships of information objects in space (and time) has been an area of interest in recent years. This pattern of seeking structural relationships on information objects in a geographical dimension, which is undirected and passive in nature can be perceived as an individual’s spatial awareness. Spatial awareness can be described as:

- *Spatial awareness as the ability to perceive (to see or recognize space)(*Poole, *et al.*, 2006)
- *Spatial awareness is the user’s implicit knowledge of his position and orientation within the environment during and after travel.* (Bowman *et al.*, 1997 p.46)
- *Spatial awareness is knowing where you are in relationship to other objects in your surroundings and how that relationship will change in the near future as you and other objects around you change positions.* (Poole *et al.*, 2006)
- *In architecture, spatial awareness refers to the feeling of open space a person gets from the interior or exterior of a building due to the way the architect has designed the building.*

Similar analogies are drawn with the way individuals perceive a map as a medium of communicating geographic information.

The internet (web) has been conceptualized as an information space that facilitates not only human readability, but also the participation of machines and users within the communication medium (Berners-Lee, 1999). Morville (2005, p.153) states “*as we build our internet of objects, the permutations of socio-semantic metadata will create new avenues of findability. Where has this object been? The era of ambient findability will overflow with metadata, as every object and location sprouts tags: social and semantic, embedded and unembedded, controlled and uncontrollable. Imagine the sensory overload of a walk in the park. Every path shimmers with the flow of humanity. Every person drips with the scent of information: experience, opinion, karma, contact. Every tree has a story: taxonomies and ontologies from bright lattices of logic. Desire lines flicker with unthinkable complexity in this consensual hallucination of space and non-space, a dynamic and diverse and yet overwhelming socio-semantic experience*”.

Drawing analogies and within the context of this research, spatial awareness can be described as the geospatial information space (information space projected onto a geographic dimension) which a user perceives to perform a certain activity of interest in relationship to other objects in the geographic space and how that relationship will change in the near future as the user and other objects around the user change their positions. Interest is usually expressed in the form of natural language with concepts, spatial relations and geographic locations as its key elements. Access of spatially aware information can be achieved based on the following criteria illustrated below. Specification of these spatial/geographic contexts frequently requires the use of a variety of spatial relationships concerning distance or containment or proximity or direction:

- *Identities of the entities*
- *Location of entities*
- *Spatial activities or events (an activity is an active expression of a set of actions, whereas an event is a passive expression of a set of actions, where a player performs or has performed at a specific instance of time)*
- *Time where the information is relevant*
- *The usage of geospatial information space, where entities exist (spatial affordance)*
- *Perceived attraction of a specific resource to a geospatial location*

2.7 Limitations of current research

In the previous chapter we have introduced the motivation and the research questions underlying this thesis and discussed the basic concepts of the internet and the world wide web, explained search engines and pointed out how web pages have a geographical scope (provider, content, and server location) that may be exploited to achieve spatial awareness in web searches.

In this chapter, then, the theoretical background of this thesis was introduced and the pertinent literature reviewed. We have started out by defining ‘information’ and by discussing the process and models of information seeking. Since the main focus of this thesis is on the visual representation of spatial information retrieved from the internet, the main part of this chapter has been devoted to the review of different visualization methods that might be potentially useful to for our purpose. Understanding that information seeking is an interactive process of iterative query formulation and query refinement, we have approached the discussion of visual representation methods from the perspective of interaction techniques that may be linked to the visualization process. Finally, we have briefly reviewed some of the basic definitions and concepts of cognition, with a focus on spatial cognition and spatial awareness.

We are now in a position to identify, in conclusion of this chapter, limitations of the current state of the art. We will distinguish between limitations of a more general nature, and those that apply specifically to visual representation in support of information seeking. The former triggered research work by other researchers of the SPIRIT consortium, while the latter gave rise to this researcher’s thesis.

2.7.1 General limitations in information seeking

The following limitations of recent research originally provided the motivation for the SPIRIT project, in an attempt to develop techniques for spatially aware information retrieval in web documents. Since the kick-off of this joint European project, geographic information retrieval has generated substantial research interest (Jones and Purves, 2006), including among the top players in the search engine industry. Thanks to projects such as SPIRIT the state of the art has been significantly advanced. However, the points mentioned below are still issues of current and future research.

- Current search engines largely ignore the geographical scope of web resources (Jones and Purves, 2006). They are ill-equipped to intelligently interpret geographic terminology or geographic prepositions such as “near”, “north of” or “within 5 km”.

- Moreover, geographic names, terms and prepositions are often fuzzy in nature and involve a number of further difficulties such as lack of uniqueness, spatial boundary changes, name changes, spatial and naming variations (e.g. modern vs. historial variants of a name), spelling variations, and neologisms.

- Existing Gazetteers are ill-equipped to interpret place names and phrases that contain place names for assigning footprints to documents that uniquely identify a location for a document.

- Text-based geo-referencing systems have to cope with these challenges and develop methods to delineate and represent imprecise geometries and regions (vernacular regions; Purves *et al.*, 2005; Arampatzis *et al.*, 2006) as well as techniques to deal with ambiguous semantics (e.g. London, UK vs. London, Ontario).

- Since search engines largely ignore the geographical scope of web resources, relevance ranking of retrieved documents is also restricted to text-based criteria. However, if web searching should become spatially aware, then relevance ranking must consider spatial criteria such as proximity as well (van Kreveld *et al.*, 2005).

2.7.2 Specific limitations in exploratory visual representation

This thesis focuses primarily on techniques and experiments in visual representation of search results in geographic information retrieval, as well as closely associated research issues. At the outset of this PhD project, the following limitations of the state of the art were identified:

- Existing user interfaces for query entry and reformulation to search engines are typically restricted to simple text interfaces since existing search engines are not spatially aware. Exceptions exist, such as kartoo.com (cf. Appendix-E), but these represent spatialized (Couclelis, 1998) versions of non-spatial information. However, in our case, we are dealing with geographically referenced information and hence visualization and map-like graphical user interfaces would seem a lot more meaningful.

- Existing user interfaces offer limited ability to refine, search and visualize the results, reflecting their lack of knowledge of the semantics of geographical terminology and the associated spatial structures.

- Due to the lack of spatial awareness of typical current search engines and the consequent lack of visualization in the user interface, it is also unclear which of the visualization methods and interaction techniques reviewed in this chapter offers the best potential to be used for the graphical representation and query refinement of web searches. Answering this question would require a comprehensive implementation and comparative empirical evaluation of alternative graphical representation options.

- Studies of the effectiveness of different modalities of user interfaces and result representation are lacking, both for general IR and specifically for spatially aware search engines.

- Relevance is a subjective term and is governed by user’s perception and expected interests. As existing, non-spatially aware search engines have not dealt with mixed semantic-spatial relevance, no empirical evidence exists how well the results of text-based ranking and mixed ranking algorithms (van Kreveld *et al.*, 2005) match user needs.

- Geographical information retrieval systems could be used in spatial activity planning, for instance, in tourism. However, that would require that the search engines are capable of interpreting subjective perceptions of distance in relation to effort over time expressed by users, and capable of constructing activity spaces based on the user’s cognitive model.

In the subsequent chapters, we work towards building on from the observations made above to construct a comprehensive framework for extracting and representing spatially aware information retrieved from the internet. In the next Chapter, we discuss the requirements gathered from experiments to understand and evaluate existing functions and tools available to address the research questions stated in Chapter 1.

Chapter 3

Spatially aware information seeking

(Requirements gathering)

"what we find changes who we become"

- Peter Morville, The Age of Findability, Boxes and Arrows.

3.1 Introduction

The course of this chapter builds on from research statement and the theoretical background (Chapters 1 & 2) related to the extraction of geographic information from unstructured textual sources such as the internet. The chapter sets out to describe two key ways in which the requirements for the SPIRIT project were defined – through the elicitation of use case scenarios from the development team and by the use of these scenarios in gathering qualitative and quantitative data from prospective users of the system. Based on this requirements gathering process, a set of requirements are formulated and used, together with theory, to develop an information seeking model.

Users are at the core of the spatially aware information seeking process. Thus, user centered design demands understanding of the behavioural modes and motivations behind the acquisition of spatially aware information on the web. Adopting a range of methodologies from the information retrieval and interaction perspective, user-centered design methods and behavioural studies, requirement studies focused on:

- extracting usage and tasks involved in the spatially aware information seeking process;
- qualitative (experimental) studies of a few chosen search engines and geoweb services;
- collection of cognitive and statistical data from users performing spatially aware information seeking tasks using qualitative (questionnaire) data collection methods; and
- analysing how users interact with the existing search engines and geo-web services.

Table 3.1 illustrates a model for user-centered information retrieval interaction and interface design detailing methods and tasks adopted for early stage requirements gathering.

Table 3.1: Model for user-centered IR interaction and interface design (based on the model by Allen, 1996, p.24) (Hansen, 1997, p.8, Table 2)

Component	Method	Task
Resource Analysis	Description of information system functionality	Describe resource(s) that are used to complete the task
User Need Analysis	1. Questionnaire with 5-point scale ratings and open-ended questions (qualitative and quantitative data). 2. Log statistics (quantitative data)	1. Users goals, purpose, objective, actions, individual preferences. 2. Logging user transactions. Measures like time no of actions and type of actions.
Task Analysis	Hierarchical task analysis (HTA)	User's tasks, goals and activities that they accomplish when meeting their needs.
User modelling		Merging needs, user tasks and goals and system tasks
Designing for usage	Requirement lists (qualitative data)	Requirements elicitation for redesign of the user interface

Unfurling interest for sophisticated forms of geographic information can be found in many different contexts, ranging from city development planning and monitoring to general travel and tourism information gathering. The context of usage and the requirements, depends on the kind of end users which the system intends to address. The system was envisaged to address the requirements of:

- those who wish to find information about topics of interest or phenomenon that occurs in, or is associated with a place. For example,they might be interested in services such as shops, garages, museums or sporting facilities. Alternatively, they might be interested in documents or images that refer to some aspects of history, culture or environment of a place of interest.

In the early stages of the life cycle of this projects, it has been decided to focus on tourism domain, which later on has been extended to serve various interests across domains. To identify detailed requirements for the design of the SPIRIT system a three pronged approach was adopted:

1. construction of use-case scenarios;
2. illustrations of these scenarios using story boards; and
3. simulated work tasks based around these scenarios using existing search engines and geo-web services.

3.2 Scenario development

In a user-centered design, scenarios (Rosson & Carroll, 2002) facilitate a narrative description of user interaction and experience with a system. Their strong narrative nature helps to obtain insights to potential problems with the planned system and envision a comprehensive solution that can be extended to specific use cases. Furthermore, scenarios can be a powerful tool in actively engaging designers and the developers with user-centred design.

A total of 11 scenarios were generated by a group of experts working in the Geographic Information Sciences domain. Each of these scenarios had an associated need for services or activities and reflected varied understanding and perspectives of the functionality of the envisaged SPIRIT system. In addition, these scenarios facilitated in providing information for contextual analysis based on end users, tasks involved, complexities of interactions, context of usage. A summary of key elements of scenarios extracted is illustrated in Table 3.2

Table 3.2: Summary of the scenarios generated by focus groups (Balley *et al.*, 2002, p.8)

User Context and Task	Global Requirements (Expression of Need and SPIRIT Requirements)
User: Ordinary user Task: to find a place where to build a beach volley ground	**Expression of need:** expression of constraints (wind, accessibility, price) regarding the place. **SPIRIT answer:** SPIRIT integrates pieces of information from varied resources that can be in the SPIRIT (ontology of things and geographic places to express the need) or outside SPIRIT (routing site, weather forecast, real estate broker.) SPIRIT integrates the resources, at least spatially.
Context: An event (fire) happens somewhere. User: Ordinary web user Task: to evaluate the population safety actions	**SPIRIT answer:** SPIRIT locates a place and addresses things relatively to this place (communities living near it, nearest highway and communication facilities, nearest water body)
User: Ordinary web user Task: to book a hotel for business travel	**Expression of need:** expression of spatial constraints on the hotel looked for (close to railways station, close (walk) to a Conference center). **SPIRIT answer:** SPIRIT located a Conference center and addresses hotels relatively to expressed spatial constraints.
User: Ordinary web user Task: to prepare holidays (to find a city and some specific places associated to it)	**Expression of need:** expression of constraints on the city (must have spatial relationships with other things like rivers, bicycle ride facilities, historic villages, monuments of Bacsteingotik style). **SPIRIT answer:** SPIRIT locates things and evaluates spatial relationships.
User: Ordinary web user Task: to find a place for wind surfing during holidays	**Expression of need:** expression of constraints to find a place (near a specific place, supporting windsurfing activity) **SPIRIT answer:** SPIRIT locates things and evaluates spatial relationships.
User: ordinary web user Context: user has just moved Task: to find facilities in the new neighbourhood	**Expression of need:** designation of a place through a map. **SPIRIT answer:** SPIRIT is able to address things (facilities) in a given place (neighbour or user place) and to detail their location to the user on a map. SPIRIT ranks the results.

Table 3.2: Summary of the scenarios generated by focus groups (Balley *et al.*, 2002, p.8)

User Context and Task	**Global Requirements (Expression of Need and SPIRIT Requirements)**
User: ordinary web user Context: have just moved and already explored their new area Task: to find a house to buy in their new place possibly a house they have noticed during exploration.	**Expression of need:** designation of place with many criteria. **SPIRIT answer:** SPIRIT is able to determine service sites selling houses in their area and more specifically the services that sells a house located at a specific address.
User: Ordinary web user wants to visit Japan but not gifted in geography of japan. Task: to choose a place for holiday	**Expression of need:** expression of constraints in places (in Japan, big city) and searched relationships between these places (intercommunications). **SPIRIT answer:** SPIRIT interprets the constraints and locate places and test their relationships. SPIRIT provides two parallel search methods (SPIRIT and GOOGLE)
User: ordinary web user (hobby: archaeology) context: goes to a conference abroad Task: to plan holidays after the conference (to find places supporting specific activities)	**Expression of need:** designation of a zone of interest. **SPIRIT answer:** SPIRIT gives the possibility to express vague queries and to browse all types of answers associated to it.
User: biologist Context: visit friends in Australia Task: to find a zoo to visit during this visit	**Expression of need:** definition of a zone of interest where to look for things. **SPIRIT answer:** SPIRIT addresses things inside a zone, locates them on a map.
User: ordinary web user (hobby: bicycle) Task: to learn things about varied topics	**Expression of need:** the user might change information retrieval strategy during the query process (if he thinks of something new he wants to learn). the user may define a zone on a map. **SPIRIT answer:** SPIRIT displays browsable semantic nets associated to the user query terms. SPIRIT proposes a synthesis of varied information associated with a place

An extended usage example of the envisioned SPIRIT system that demonstrates one of the scenarios developed is illustrated below. It consists of three elements, a narrative, describing the perceived background of a pair of hypothetical users and the way they might interact with the SPIRIT system, a set of assumptions regarding the SPIRIT system and its abilities and some open questions with respect to the envisaged design.

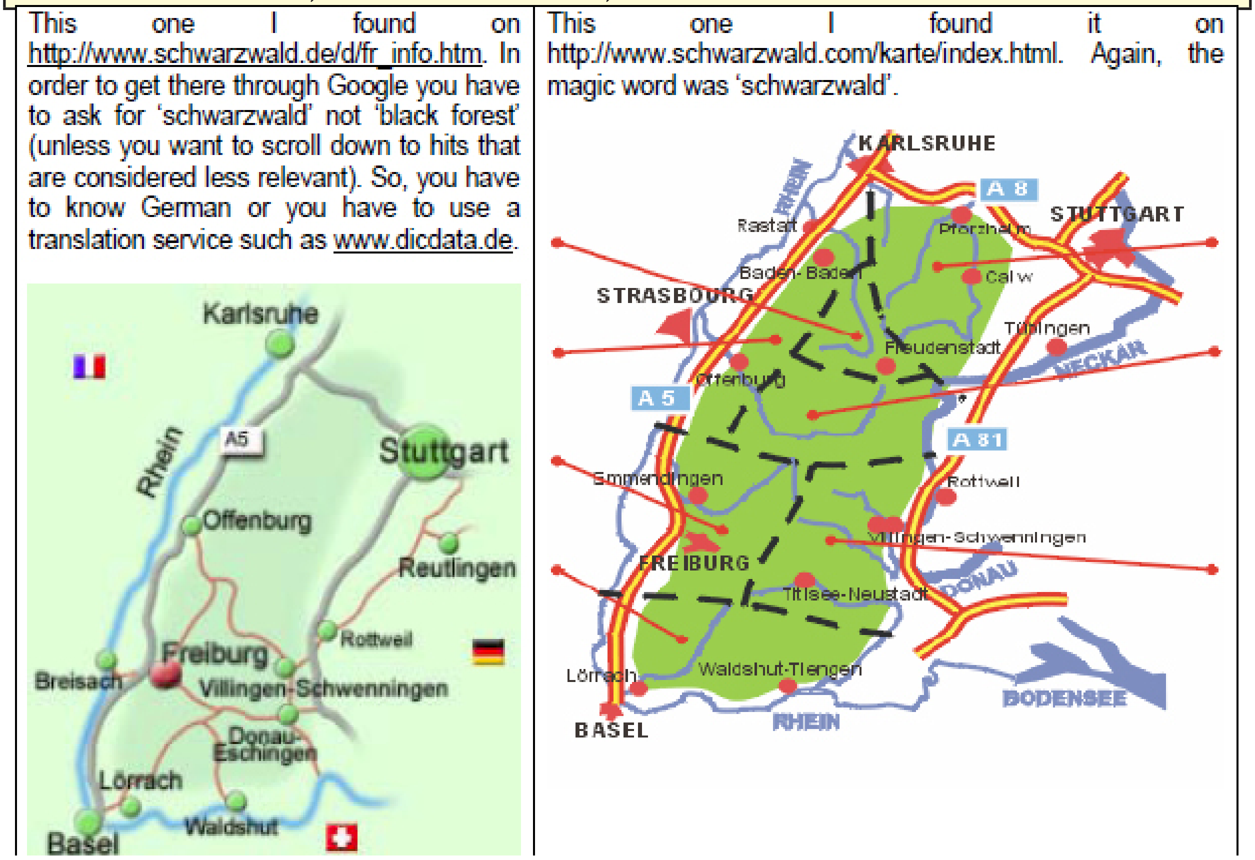
Example Scenario (Source: SPIRIT deliverable D3 7101)		
Narrative	Assumptions	Open issues/ Questions
Juliette and Marc-Antoine are a happy young couple from Paris, and fond bicyclists. Both are high school teachers, and since they believe in the use of ICT in teaching, they have a computer with ASDL connection at home which they regularly use to retrieve information from the web that they can use in their teaching.	Users are computer and internet literate. They are typically people with a solid education and a keen interest in other places, sports and culture. Though not themselves IT specialists, they see IT as an opportunity and use broadband internet connections.	Is bandwidth an issue? Are mobile devices allowed?
They like to spend their vacations doing bicycle trips. As a rule, they are visiting the home countries of former Tour de France winners.		
They have already been to Texas, the home of Lance Armstrong. So, this year, it's 1997 winner Jan Ullrich's Germany. They search Google for 'Jan Ullrich', and, sure enough, they find his homepage www.janullrich.de. Jan Ullrich even lists his place of residence on his homepage: a village called Merdingen in the Black Forest.	SPIRIT must be more than Google as far as geographic search is concerned.	
Juliette's aunt is married to a German and Juliette remembers eating Black Forest Cake when she was visiting her aunt as a child. She didn't know, however, that this was also a place name, which the information given on Jan Ullrich's homepage does seem to suggest. Juliette would like to know more about that region, where it is and how large it is. She tries her favourite map services, www.mapquest.com and www.map24.com.	SPIRIT must be more than conventional map services as far as geographic search is concerned.	How do we deal with cross-language issues in ontologies and gazetteers?

Example Scenario (Source: SPIRIT deliverable D3 7101)		
Narrative	Assumptions	Open issues/ Questions
However, both of them don't provide this kind of information; they require an address query. A Google search for 'merdingen' brings her to the www.merdingen.de, the homepage of the town of Merdingen (not maps, but pictures of Jan Ullrich at some local fun race…). Again through Google, looking for 'black forest' she finds www.schwarzwald -tourist-info.de, a portal with informations for tourists but no maps.	Language is an issue also (and perhaps particularly) for geographic names.	How do we ensure that the SPIRIT service is easily found by those not knowing it?
(Did Juliette speak German, she would have searched for 'schwarzwald' and would have found www.schwarzwald.de, a German language portal for tourism services, but with a nice little map of the Black Forest region; see below.)	It must be possible to find SPIRIT through other search engines. SPIRIT should provide links to information related to geographic regions and places, as well as footprint maps and pictures of geographic entities.	Are these links provided by (tbd) industrial partners, and might some of them even charge a premium? Do our ontologies provide abstraction hierarchies?
Hence, Juliette asks Google for 'geographic location search' and finds a service called SPIRIT. Using the 'region footprint' feature of SPIRIT, the delineation of the Black Forest is found. Juliette and Marc-Antoine can now see that the Black Forest is called 'Schwarzwald' in German, that it is a geographic region in the South of Germany, in the State of Baden-Württemberg, and that it is not an administrative unit (but made up of several administrative districts, called Kreise). SPIRIT also provides loads of links to other information about the region, along with pictures which allow them to get a first impression of the area. It does indeed look pleasant there, with some exciting topography for bicyclists. Now, they are definitely convinced that this should indeed become a worthwhile trip.	Semantic net display of related terms for geographic entities. SPIRIT handles visualizations of variable granularity in the spatial/geographical as well as semantic sense. SPIRIT must definitely be better than text-based search engines, by enabling to make spatial inferences. SPIRIT allows to make spatial queries that involve topological as well as (geo)metric relations.	What are the things that SPIRIT does different and better than Google & Co.? We need to clearly spec this out.

Example Scenario (Source: SPIRIT deliverable D3 7101)		
Narrative	Assumptions	Open issues/ Questions
And SPIRIT does more than just produce maps of an area. Because it is driven by a spatial ontology, it can display a semantic net of geographic entities that are semantically (and spatially) related to 'Black Forest'.		
Both the geographic map visualization as well as the semantic net visualization can be navigated horizontally (on the same level of granularity or scale) as well as vertically (coarser / finer levels of granularity).	SPIRIT should allow to make inferences about spatial relations as well as visualize them, so the user can gain a better understanding of the portion of geographic space that he/she is interested in.	How extensive should the mapping cabablities be? Can we not make use of existing mapping services through APIs, like we intend to do for the search engine functions through the Google API ?
Juliette and Marc-Antoine are now definitely thrilled by the capabilities of SPIRIT. They start looking for route information on bicycle routes. Initially, they had intended to do this through Google, but since they are convinced by the apparent features of SPIRIT for geographic search, they decide to stick to it and try there. And SPIRIT doesn't let them down.		Again, is it safe to assume that we link up to services such as map24.com through their API (see question above)? – If this was our aim, we would have to come into an agreement with the service provider of map24, a company called Netsolut (www.netsolut.com).
The town of Merdingen is already a given on their trip; so, they decide to first ask for routes connecting to this place. This seems to be a small place, as it is only connected by small county roads (K4929, K4930, K4979). (All the better: You don't cycle on motorways if you can avoid it.) To find out more about the surroundings, Juliette and Marc-Antoine ask SPIRIT for 'populated places > 5,000 inhabitants up to 50 km from Merdingen'. They enter this query through a query interface that allows them to compose spatial queries in an intuitive way (menu-driven, but not via natural language). As a result, they find Freiburg (im Breisgau) and Breisach, along with a few other smaller cities.	It would be beneficial if we tried to team up with a mapping service. Map24 (driven by technology developed by www.netsolut.com) provides probably the best mapping service (in terms of cartography) as well as APIs and hence might be a natural partner. Just like Google for the search engine part.	

Example Scenario (Source: SPIRIT deliverable D3 7101)		
Narrative	Assumptions	Open issues/ Questions
Using the maps of SPIRIT and the built-in zooming/panning capabilities, they soon are able to gain a picture of the region:		
Merdingen is not really located in the hilly and mountainous part of the Black Forest; it is located at the edge of the Upper Rhine Valley plain. The city of Freiburg lies some 8 km to the east, at the foothills of the Black Forest (which extends further to the east). With this enhanced understanding of the area, it becomes easy for Juliette and Marc-Antoine to decide how they should plan their trip. They decide to spend a couple of days in Freiburg (which seems to be a pretty city, guessing from the pictures provides on the official city tourism website), make a 'detour' to visit Merdingen as well as a few day trips to the hills of the Black Forest. After that, they plan to spend another few days in the higher parts of the Black Forest, in Titisee (Lake Titi), and perhaps even take short trips to Switzerland (e.g. to a fabulous town called Hallau, east of Schaffhausen).		
For detailed planning of the trip, they find that SPIRIT is not really the best place to look for such information. It is not really geared towards locator and route maps. However, in tandem with www.map24.com, route planning becomes easy, with map24 delivering the detailed route maps (with excellent zooming facilities) and SPIRIT taking care of the spatial and semantic linkages.		
Perhaps if Jan Ullrich had SPIRIT, he could win the Tour de France again?		

<table>
<tr><th colspan="3">Example Scenario (Source: SPIRIT deliverable D3 7101)</th></tr>
<tr><th>Narrative</th><th>Assumptions</th><th>Open issues/ Questions</th></tr>
<tr><td colspan="3">Is it possible already today (without SPIRIT) to find out about the whereabouts of the Black Forest?
Yes. it is. However, it is really painful and you might get a completely wrong idea, depending on which website(s) you end up on. It is also close to impossible to find a delineation (footprint) of the region. Most often, what you find is road maps that somewhere have the name ‘Black Forest’ or ‘Schwarzwald’ written on it – amidst zillions of city names, road numbers etc. Given that the Black Forest is a rather important tourist region, it is quite surprising that it would be so hard to get good visualizations today. Apparently, most providers are mainly interested in providing links to tourist sites, which themselves mainly push hotel reservation services, city tours, etc.
Also of note: The Black Forest seems to be a nice example of competing portal providers (frequent in the tourism domain) that use similar names and offer services and content of very variable quality – and of course never link up to their competitors (who may actually be different administrative units, such as different cities, counties, provinces, etc.). Some examples: www.schwarzwald.de, www.schwarzwald.com, www.schwarzwald.net, www.schwarzwald-tourism-info.de etc.</td></tr>
</table>



This one I found on http://www.schwarzwald.de/d/fr_info.htm. In order to get there through Google you have to ask for ‘schwarzwald’ not ‘black forest’ (unless you want to scroll down to hits that are considered less relevant). So, you have to know German or you have to use a translation service such as www.dicdata.de.

This one I found it on http://www.schwarzwald.com/karte/index.html. Again, the magic word was ‘schwarzwald’.

Figure 3.1: Example Scenario obtained in terms of region delinations of the Black Forest (Source: SPIRIT deliverable D3 7101)

3.2.1 Scenario analysis

Analyzing the fourteen scenarios generated revealed that the key elements of an expressed need are frequently an aggregation of several simpler needs. From a holistic perspective of this need, two situations may occur:

- the need is too complex to be expressed in a single query; or
- users need intermediate/supportive information to express their needs.

In either of these cases, the process is of an iterative nature, enabling users and systems to progressively understand each other by, for example, disambiguating user's search terms and validating systems responses, to build "*common ground*". Based on these characteristics, a user's state can be expressed as "*definitive*" or "*inquisitive*" in nature (referring to Chapter 2, information seeking behaviours). Through contextual analysis of the scenarios two key components are identified in the queries generated: a "*where*" and a "*what*" together with a spatial relationship between the geographical element (where) and the non-geographical objects (what). In some cases, there may be more than one such relation for example, "*hotels in Munich close to the main station*", where there are 2 geographical entities (Munich and the main station) and the non-geographical objects are related to the locations by "*in*" and "*close to*". Specific geographic locations and regions are among the key elements expressed by individuals. Table 3.3 presents contextual analysis of the scenarios generated by the focus group detailing the geographic and non-geographic components as well as the relations used to express them in their information request.

Table 3.3: Key elements of scenarios (Balley *et al.*, 2002, p.14)

Scenario	**Service**	**Location**	**Relation**	**Other**
1.	Accommodation, Hotels	Paris, cities near by	In, Near, No more than 15 miles away	
2.	Airport, cinemas, school, museums	North-East, Cardiff, Somerset, Ile de France, Ruhrgebeit	Near by, in, in the near by area, within 10 miles/30 minutes drive	
3.	Hotels	West of London	West of	
4.	Parks, Zoo	South Australia, Sydney Melbourne	Between, near	Active maps, free typing
5.	Cities	Garonne South, France	In (graphically expressed)	
6.	Hotels	Munich, Main station	In, Close to, Short walk	Distance with respect to 2 points
7.	Accommodation, Bicycle	River, bicycle paths	Along	Distance calculation
8.	Generic services, geo-info	Upper Petaca, New Mexico	Surrounds, close	Structured interface
9.	Hotels, Windsurf renting	Mediterranean coast of France, Nice	In, within 20 min drive	The coast, free typing
10.	Big cities, hotels, museums, transport	Japan	In	Structured interface
11.	Parks, Archaeological sites	Rome	Neighbourhood	
12.	Accommodations, mountaineering, weather forecast	Switzerland, les diablerets (mountain)		Mixed Languages

Table 3.3: Key elements of scenarios (Balley *et al.*, 2002, p.14)

Scenario	Service	Location	Relation	Other
13.	General Info,Bicycle routes	Black forest, Merdingen	Up to 50 Km	Mixed Languages, Structured Interaction
14.	House nursery	University campus, hospital	Distance from	Distance respect to 2 points

Similar findings have been cited in the literature by Couclelis (1998, p.214) who states that "*place, way, region are among the most fundamental geographic concepts. They are deeply meaningful to people, full of connotations derived from their multiple practical and symbolic roles in everyday life*". In addition, to these key elements "*distribution, spatial interaction, scale and change are other prominent elements recognised by the geographers (National Research Council, 1997)*". Furthermore, a typology has been presented (Table 3.4) based on "*Mathematical, Socio-economic, Behavioural and Experiential space where an increasingly complex conceptualization of space is used in human geography... Experiential Space is defined to be the most intuitive and least theorized in the sequence, and some of its underlying concepts are chosen*" which is concerned with the perceptions and experiences users carry and their understanding of relations in space. Table 3.4 details about conceptual correspondences that exist in individual's minds which aid to draw meaningful associations, but in reality spatial relations do not exist in the world in any meaningful sense. These conceptual correspondences tend to differ when users/individuals interact with them. Couclelis (1998) (ref. Table 3.4) details how conceptual correspondences of individuals based on experience, perception of phenomena and relations in space can be transformed onto a mathematical space in order to represent in mathematical space and digital computers.

Table 3.4: Conceptual correspondences between mathematical space and experiential space (Couclelis, 1998, Table 2, p.215)

Mathematical Space	Experiential Space
Point	Place
Line/Boundary	Way/limit/Boundary
Area/Polygon	Region/territory
Plane/field	Domain/Expanse/Continuum
Object	(other) Geographic Entity

These elements of information landscape and association functions shown in Table 3.5 illustrate individual's spatial awareness needs. The use-case scenarios described by the focus group have illustrated typical contextual usage of the envisioned SPIRIT system. Much may be hoped from an initial review of capabilities of various search engines and geo-web services that could meet the requirements envisioned by the scenarios illustrated by the focus group (see Table 3.2).

Table 3.5: Elements of the information landscape and associated informational functions. (Couclelis, 1998)

Query and retrieval operations	Higher level functions
Place: input, store, retrieve, save, find, in, sort, classify, delete and replace	contain, enclose, include, exclude, gather, separate, remove substitute, relate, associate, fit in, belong to, be part of, be similar to, keep together, keep whole, preserve, pull-out of, destroy, break up.

Table 3.5: Elements of the information landscape and associated informational functions. (Couclelis, 1998)

Query and retrieval operations	**Higher level functions**
Way Barrier: Locate, search, find, start from, end at, sequence, rank, sort, order.	link, connect, associate, interact, influence, disconnect, separate, reach, lead to, flow, come from, go leave, traverse, cross, return, meet before, meet after, end, stop.
Region:(Same as Place)	{Same as place, plus:} be prominent, be most important, be at the center, dominate, be peripheral, depend on, surround, embed, converge, diverge, extent, spread, cover, distribute, support, be on, be on top, link, connect, associate, interact, influence, disconnect, separate.

3.3 Storyboards and simulated work tasks

The scenarios developed and described above provided a first set of tools to consider the requirements which the proposed SPIRIT system must meet. However, they were generated by members of the SPIRIT team, and reflected their preconceptions as to the likely direction of the system development. In a second stage of requirements capture, story boards and simulated work tasks were prepared using existing web-based tools, to explore in more depth the likely interactions of users with the planned SPIRIT system, and to identify which needs could or could not be met by existing solutions.

3.3.1 Storyboards

Story board interfaces are used illustrate possible uses of a system in a sequential way. They are intended to illustrate the desired functionality to be provided in a system and are particularly useful in the conceptualization of a process, Thus helping designers to better communicate the perceived reality of the system with potential users.

A set of simple interactive mock-ups were created using MS PowerPoint presentations, which presented a limited set of possible interactions such as selection, drawing a window on a map and responses to basic queries. (refer to Annexure DVD/CD accompanied with this thesis- Folder on mock-ups and storyboard containing storyboards created in MSPowerPoint).

3.3.2 Simulated work tasks

Simulated work tasks are used to present a model of a realistic system to the subjects involved in the study with short descriptions of a situation “ *that may lead to the information retrieval seeking tasks*” (Borlund, 2000, p.80). These simulated work tasks are used to explore user requirements in a controlled situation. Borlund (2000, p.140.) suggests that it is possible to “*substitute real information needs with simulated needs through the application of simulated work task situations*” (Borlund, 2000, p.140) and thus to draw meaningful conclusions about a planned system. Several stimulated work task situations that reflected possible interactions between users and the systems and that of the graphical components were delivered in a so-called “*Wizard of Oz*” experiment – that is to say a set of storyboards with differing results were prepared, where according to the user’s interaction an appropriate result was presented.

The simulated work tasks were repeated with, but in a web environment using existing search engines by selecting three use-case scenarios generated during the requirements gathered in the early stage experiments.

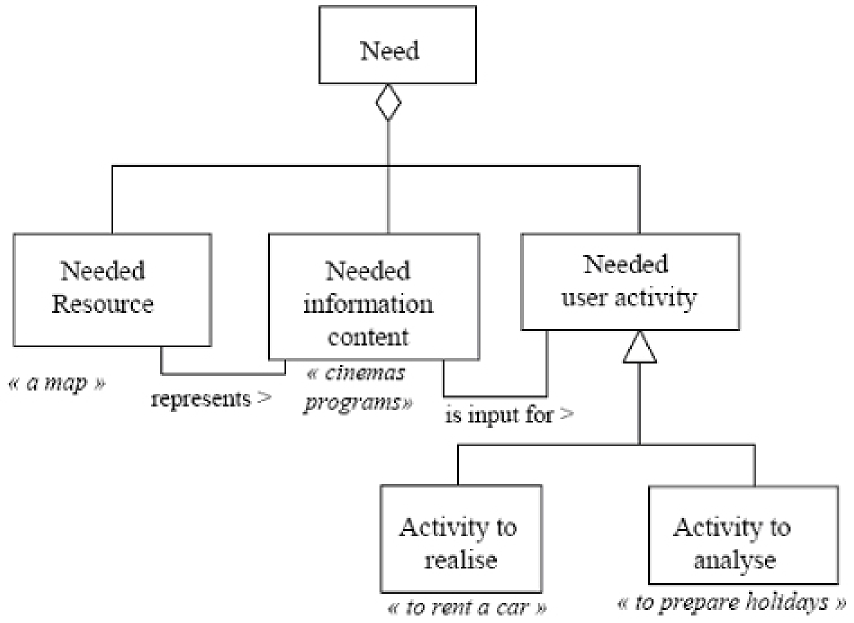


Figure 3.2: Dimension of need of SPIRIT user (Balley, 2002 p.11 Fig 1)

These carefully chosen scenarios (Table 3.6) can help us identify gaps in the current search engines and geo-web services available at the time of the study. The selected search engines and geo-web interfaces were: *Google(www.google.com), Kartoo(www.kartoo.com), Multimap(www.multimap.com), Map24(www.map24.com), Mapquest(www.mapquest.co.uk), Upmystreet(www.upmysteet.com).* These search engines and geo-web services were chosen because of their popularity, ease of use and functionality. Kartoo was specifically chosen because it provides a cartographic approach to the representation of the results instead of simply listing them in a traditional manner.

Table 3.6: Chosen use-case scenarios for the assessment of early stage requirements of existing search engines and geowebservices (Balley *et al.*, 2002)

Scenarios	**Description**
1.	You are planning your two week vacation to Black Forest (Schwarzwald) in Germany. You would like to go cycling. You have to identify the best route cycling around. You would also include visits to small cities as well as big cities in your travel plan.
2.	You are planning to spend a vacation on the Mediterranean coast of France and wish to stay in a reasonably large city with good restaurants, commercial shopping centers. You would prefer accommodation in a city on the coast since you wish to practice windsurfing in a near by place. You wish to hire the equipment for windsurfing thus you need a fully equipped beach service
3.	You are planning a trip to the Swiss alps with your family. You have heard from a friend that there is some nice mountaineering near a place called Agile. and also some pleasant walks that you could do with your children (who are aged 7 and 9 respectively). You wish to book a chalet to stay in with your family and also get information about mountaineering routes, rock climbing and hiking in the area. You have heard of one mountain called Les Diablerets with nice landscapes

3.3.3 Environment and subjects

"Wizard of Oz" Powerpoint storyboards were created based on the preconceptions of the

SPIRIT development team reflecting the use case scenarios, using existing web-based tools such as google, which provided the results for these use case scenarios and map based representations have been obtained from Map24 (www.map24.ch) and some other sources based on the availability of a specific map that addresses the needs of the scenario. The Wizard of Oz storyboards were developed after a series of iterations by actually carrying out the tasks of the use case scenario with existing systems such as google and map24. Based on our understanding of the interactions and the system responses, the Wizard of Oz storyboard were developed. The Wizard of Oz experiments were developed jointly by Zurich Team & IGN (France). The Wizard of Oz experiments were roughly scheduled to last for one hour with each user carrying out 2 scenarios.

The focus groups for these experiments were drawn from domain experts working in GIS/ geography or in domains requiring regular use of geographic data. Representative users for various domains were also pooled into evaluate story boards and simulated work tasks from different nationalities. Table 3.7 illustrates the representative users from various domains recruited to participate in the Wizard of Oz experimental studies with considerable geographic domain experience.

Table 3.7: Representative users from various domains **recruited for Wizard of Oz studies (Bucher** *et al.*, **2004 p.12 Table 2)**

Domain	SPIRIT site
Real estate	Sheffield(U.K), Cardiff(U.K)
Travel	IGN, Paris(France), Sheffield(U.K), Hannover(Germany)
Event Organizers	Zurich(Switzerland)
Others (e.g. employment agencies)	Cardiff(U.K.)

With the web as delivery platform, personal computers with pentium IV configuration (Windows O.S) and Java plug-in enabled browsers (IE & Mozilla) have been used to perform the simulated work tasks using existing search engines and geo-web services. The experiments were roughly scheduled to last for one hour, acquiring user's background, self learning process of some selected geo-search engine and later on to carry out the simulated work task scenarios illustrated in Table 3.6. In total 14 participants carried out these experiments using existing search engines and geo-web services. All the participants were doctoral, post doctoral or graduate students of Department of Geography, University of Zurich; Thus having considerable geographic domain experience. The subsequent sections detail the insights drawn from these experimental studies.

3.3.4 Analysis methods

For these early stage experimental studies, a combination of data analysis methods were adopted (Table 3.8). Individual records for each user were created reflecting the envisioned SPIRIT system, various tasks involved to achieve the user's intermediate as well as final goals and potential difficulties in accessing information from varied sources.

Table 3.8: Types of data collected, data collection methods and analysis methods.

Data collection methods	Types of data collected	Data analysis method
Evaluation questionnaires before and after spatially aware information seeking tasks	1.Qualitative data: 5-point likert scale for questionnaire 2. Qualitative data: Written (Open-ended) data to the 5-point Likert scale 3. Interview: discussing key features and functionalities.	1.Quantitative data analysis. 2. Qualitative analysis of written data 3. Comparison of statistical data 4. Task Analysis of qualitative data

3.3.5 Findings

Functional analysis of interface modalities

Before reporting on the user studies, a brief description of the differences between various search and geo-web services is given. This analysis reveals the different ways in which it was both possible to formulate, reformulate and retrieve query results with existing systems.

Among the many geo-web services selected, each exhibited a different functional mapping interface. The interface is the medium that presents a functional overview of the underlying system and the success and failure of user-centered systems is often primarily based on the user interface design. Figures 3.3-3.7 illustrate various interfaces adopted by geo-web services as starting points and for result presentations. It is to be noted that a few tend to opt for simple structured query interfaces along with a graphical interface ("*map*"). This is however governed by the service it tends to offer. Route planning services provides a standard structured query interface with start and end points, where as for identification of a specific address, structured query input is supported with maps aid exploration.

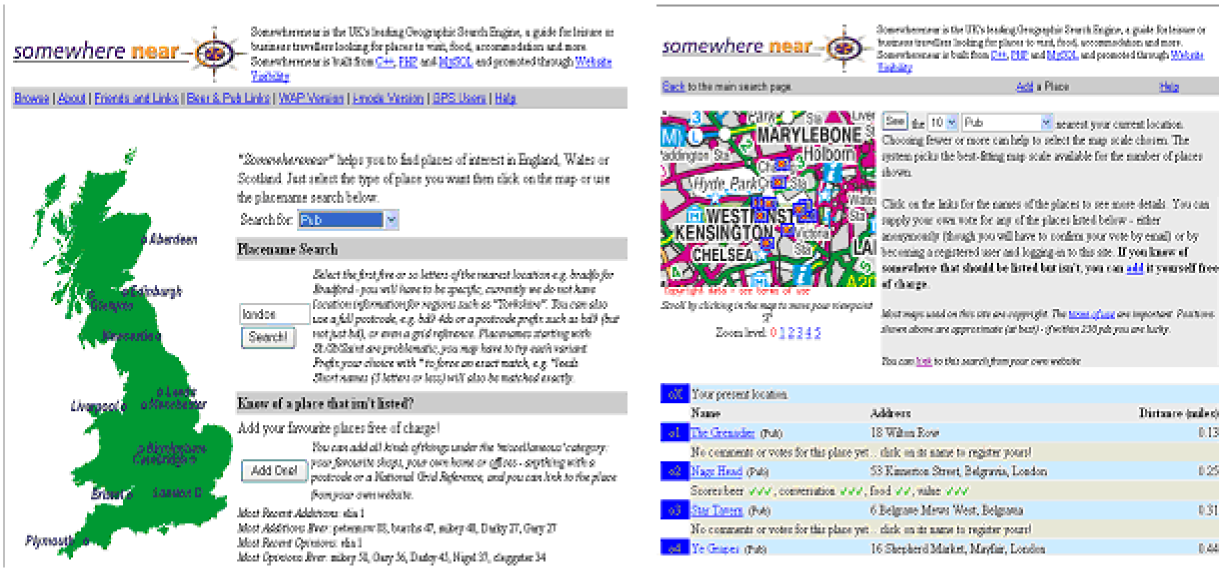
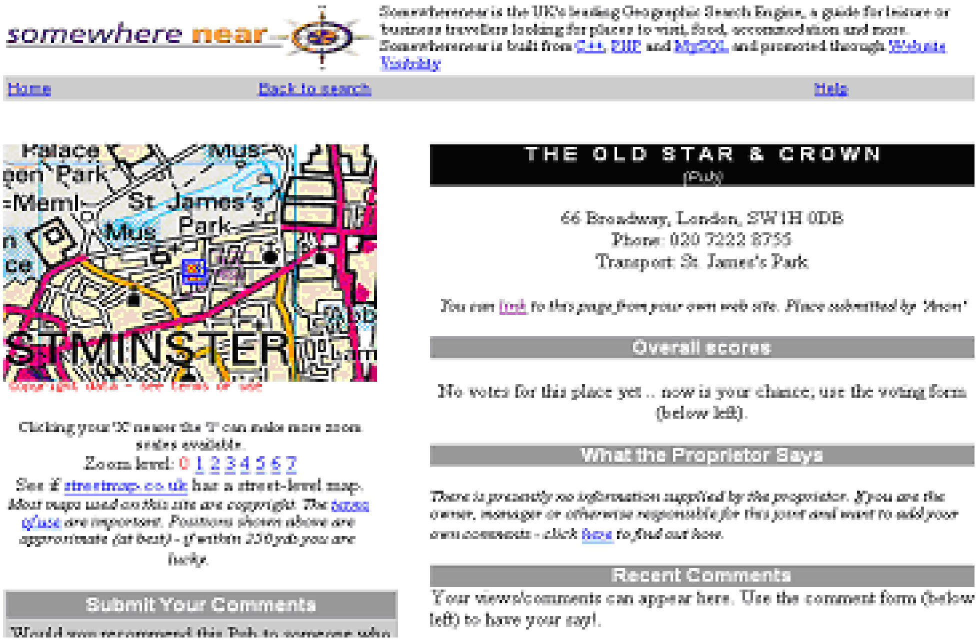


Figure 3.3a: A snap shot of the Starting Point , results presentations of geo-web service: SomewhereNear (http://www.somewherenear.com)



Somewherenear is the UK's leading Geographic Search Engine, a guide for leisure or business travellers looking for places to visit, food, accommodation and more. Somewherenear is built from C++, PHP and MySQL, and promoted through Website Visibility

THE OLD STAR & CROWN
(Pub)

66 Broadway, London, SW1H 0DB
Phone: 020 7222 8755
Transport: St. James's Park

You can link to this page from your own web site. Place submitted by 'Anon'

No votes for this place yet .. now is your chance; use the voting form (below left).

Clicking your 'X' nearer the 'I' can make more zoom scales available.
Zoom level: 0 1 2 3 4 5 6 7
See if streetmap.co.uk has a street-level map.
Most maps used on this site are copyright. The terms of use are important. Positions shown above are approximate (at best) - if within 250yds you are lucky.

There is presently no information supplied by the proprietor. If you are the owner, manager or otherwise responsible for this joint and want to add your own comments - click here to find out how.

Your views/comments can appear here. Use the comment form (below left) to have your say!

Would you recommend this Pub to someone who

Figure 3.3b: A snap shot of the Starting Point , results presentations of geo-web service: SomewhereNear (http://www.somewherenear.com)

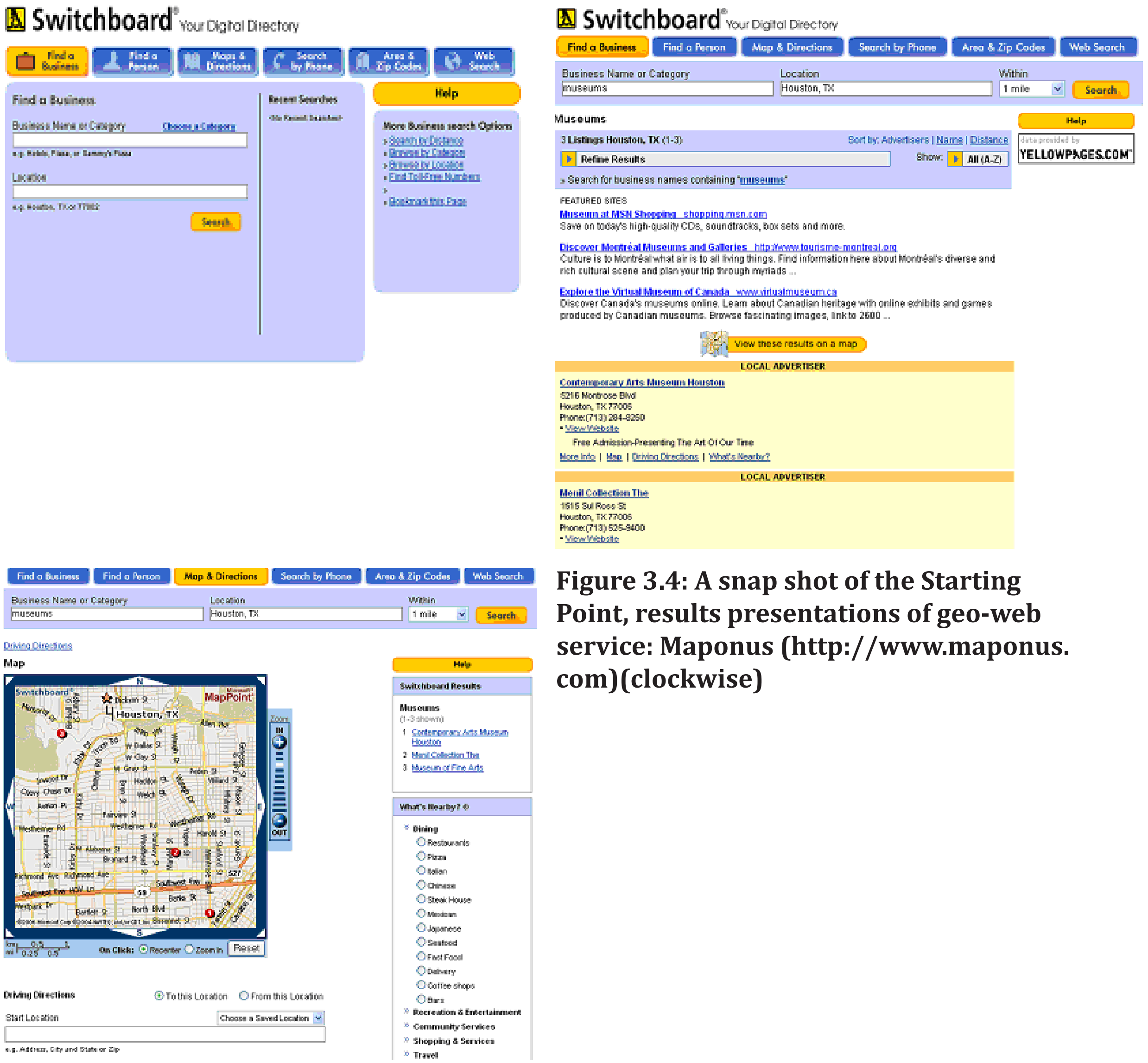
The criterion of selection of geo-web services and search engines was based on popularity and functionality and their characteristics illustrated in Table 3.9. Unfortunately, at the time of this study important spatial search engines such as *Google Local* and *Mirago* were not available or only at an experimental stage. In late 2005 the launch of map based search services was limited to the datasets collected from yellow pages. *Google*, *Yahoo* and *MSN* adopt a semi-structured graphical interface for their map based services (See Appendix- E)

Table 3.9: Geoweb services feature comparison

Feature	**Somewhere-near**	**Maponus**	**Mapquest**	**Multimap**	**Map24**
Starting Point (Structured)		Concept + Place Name	Place Name + Address+ City+ State + Zip Code, Directional input		Address Input + Route Input
Map +Structured	Concept + Place Name + Less Interactive Map			Street+ Town+ Postal Code + Interactive Map	
Geographic preposition	No	No	No	No	No
Exploration(Interacitivty	Map based (little) + Zoom level(3)	Text listing (Map on selection - little) + zoom level (12)	Text + Map (little)	Map (medium) + zoom (12)	Map (high) + Route + 3D perspective+ Rocket Zoom

Table 3.9: Geoweb services feature comparison

Feature	Somewhere-near	Maponus	Mapquest	Multimap	Map24
Presentation	Textual Listing + Distance from referred location	Text listing or on map selection	Textual listing + Map (Little)	Textual listing + Map + Near by (distance from referred location)	Textual Listing + Map + Points of Interest.



Find a Business

Business Name or Category

e.g. Hotels, Pizza, or Sammy's Pizza

Location

e.g. Houston, TX or 77002

Business Name or Category: museums
Location: Houston, TX
Within: 1 mile

Museums

3 Listings Houston, TX (1-3) Sort by: Advertisers | Name | Distance

Show: All (A-Z)

» Search for business names containing "museums"

FEATURED SITES

Museum at MSN Shopping shopping.msn.com
Save on today's high-quality CDs, soundtracks, box sets and more.

Discover Montréal Museums and Galleries http://www.tourisme-montreal.org
Culture is to Montréal what air is to all living things. Find information here about Montréal's diverse and rich cultural scene and plan your trip through myriads ...

Explore the Virtual Museum of Canada www.virtualmuseum.ca
Discover Canada's museums online. Learn about Canadian heritage with online exhibits and games produced by Canadian museums. Browse fascinating images, link to 2600 ...

LOCAL ADVERTISER

Contemporary Arts Museum Houston
5216 Montrose Blvd
Houston, TX 77006
Phone:(713) 284-8250
• View Website
Free Admission-Presenting The Art Of Our Time
More Info | Map | Driving Directions | What's Nearby?

LOCAL ADVERTISER

Menil Collection The
1515 Sul Ross St
Houston, TX 77006
Phone:(713) 525-9400
• View Website

Driving Directions

Map

Switchboard Results

Museums
(1-3 shown)
1 Contemporary Arts Museum Houston
2 Menil Collection The
3 Museum of Fine Arts

What's Nearby?

- Dining
  - ○ Restaurants
  - ○ Pizza
  - ○ Italian
  - ○ Chinese
  - ○ Steak House
  - ○ Mexican
  - ○ Japanese
  - ○ Seafood
  - ○ Fast Food
  - ○ Delivery
  - ○ Coffee shops
  - ○ Bars
- Recreation & Entertainment
- Community Services
- Shopping & Services
- Travel

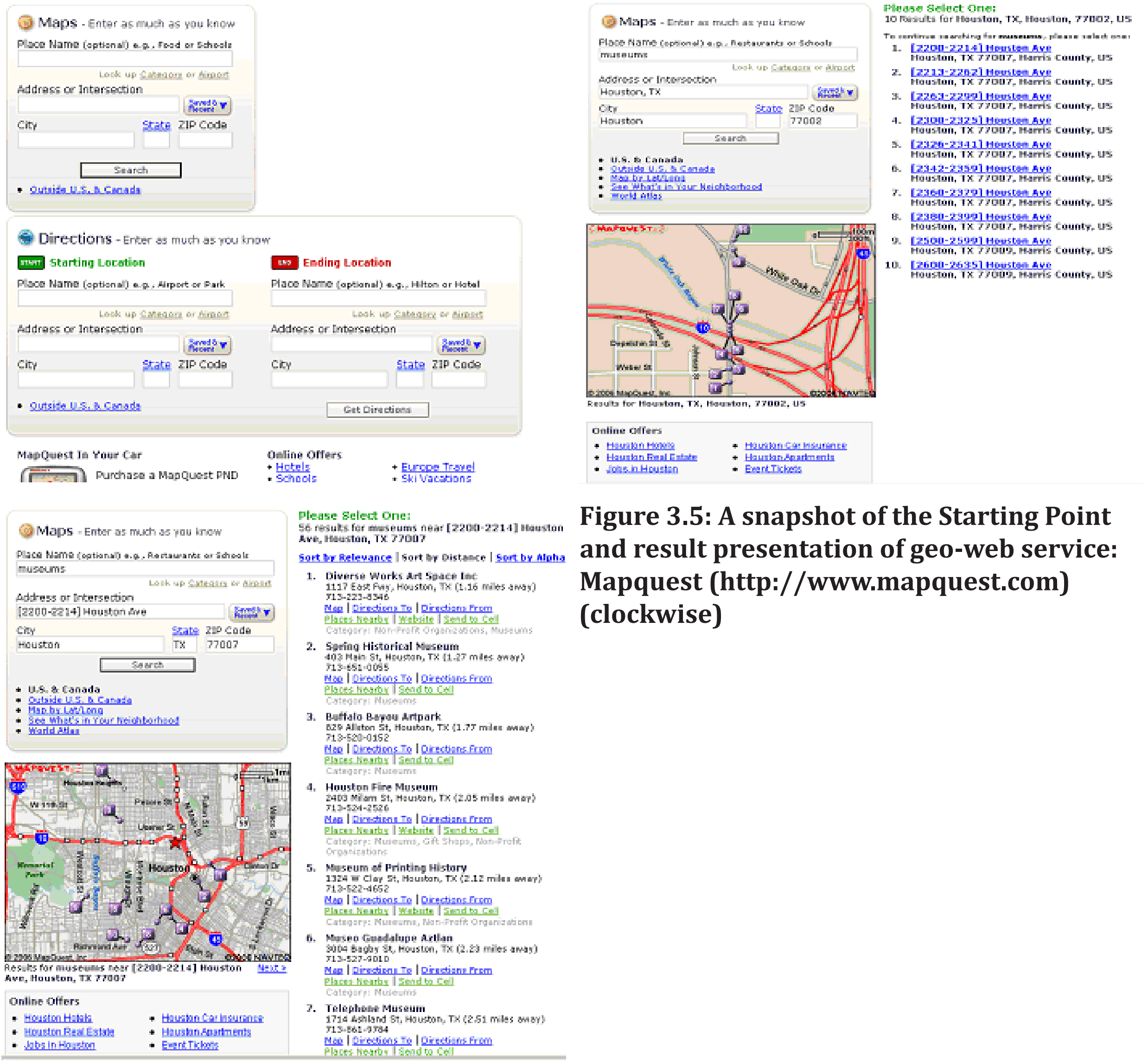
Driving Directions ● To this Location ○ From this Location

Start Location

e.g. Address, City and State or Zip

Figure 3.4: A snap shot of the Starting Point, results presentations of geo-web service: Maponus (http://www.maponus.com)(clockwise)

Appendix-E illustrates a few selected simple text based interfaces of Google, Kartoo, Grokker and Webbrain all of which are similar, whereas their approaches to presenting results vary considerably with Google adopting a simple textual listing of the results based on pagerank, while grokker presents features such as interactive filtering based on recency and classification based on keywords in addition to textual listing.



Place Name: museums
Address or Intersection: Houston, TX
City: Houston
ZIP Code: 77002

Results for Houston, TX, Houston, 77002, US

Please Select One:
10 Results for Houston, TX, Houston, 77002, US

To continue searching for museums, please select one:

1. [2200-2214] Houston Ave
Houston, TX 77007, Harris County, US
2. [2213-2262] Houston Ave
Houston, TX 77007, Harris County, US
3. [2263-2299] Houston Ave
Houston, TX 77007, Harris County, US
4. [2300-2325] Houston Ave
Houston, TX 77007, Harris County, US
5. [2326-2341] Houston Ave
Houston, TX 77007, Harris County, US
6. [2342-2359] Houston Ave
Houston, TX 77007, Harris County, US
7. [2360-2379] Houston Ave
Houston, TX 77007, Harris County, US
8. [2380-2399] Houston Ave
Houston, TX 77007, Harris County, US
9. [2500-2599] Houston Ave
Houston, TX 77009, Harris County, US
10. [2600-2635] Houston Ave
Houston, TX 77009, Harris County, US

Place Name: museums
Address or Intersection: [2200-2214] Houston Ave
City: Houston
State: TX
ZIP Code: 77007

Results for museums near [2200-2214] Houston Ave, Houston, TX 77007

Please Select One:
56 results for museums near [2200-2214] Houston Ave, Houston, TX 77007

1. Diverse Works Art Space Inc
1117 East Fwy, Houston, TX (1.16 miles away)
713-223-8346
Category: Non-Profit Organizations, Museums
2. Spring Historical Museum
403 Main St, Houston, TX (1.27 miles away)
713-651-0055
Category: Museums
3. Buffalo Bayou Artpark
829 Allston St, Houston, TX (1.77 miles away)
713-520-0152
Category: Museums
4. Houston Fire Museum
2403 Milam St, Houston, TX (2.05 miles away)
713-524-2526
Category: Museums, Gift Shops, Non-Profit Organizations
5. Museum of Printing History
1324 W Clay St, Houston, TX (2.12 miles away)
713-522-4652
Category: Museums, Non-Profit Organizations
6. Museo Guadalupe Aztlan
3004 Bagby St, Houston, TX (2.23 miles away)
713-527-9010
Category: Museums
7. Telephone Museum
1714 Ashland St, Houston, TX (2.51 miles away)
713-861-9784

Figure 3.5: A snapshot of the Starting Point and result presentation of geo-web service: Mapquest (http://www.mapquest.com) (clockwise)

Kartoo provides a rather more interactive approach presenting the results as a spatialisation, while Webbrain adopts theTouchgraph visualization technique for interactively navigating through the hierarchy. Table 3.10 illustrates a detailed comparison of result presentations of these selected search engines.

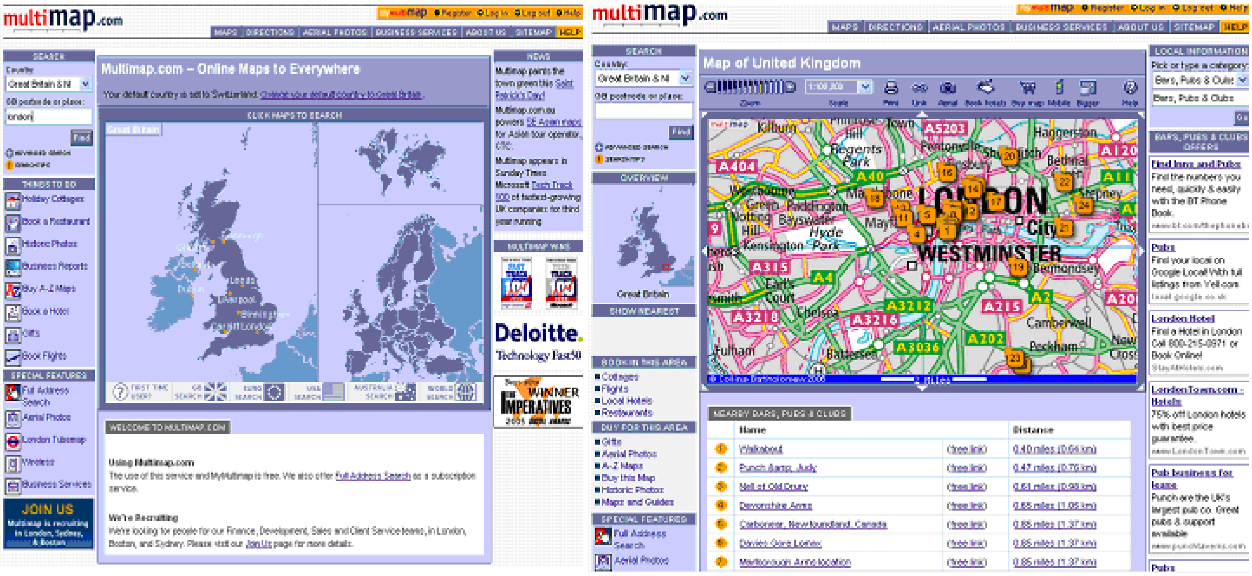


Figure 3.6: A snap shot of Starting point and result presentation of geo-web service: Multimap (http://www.multimap.com)

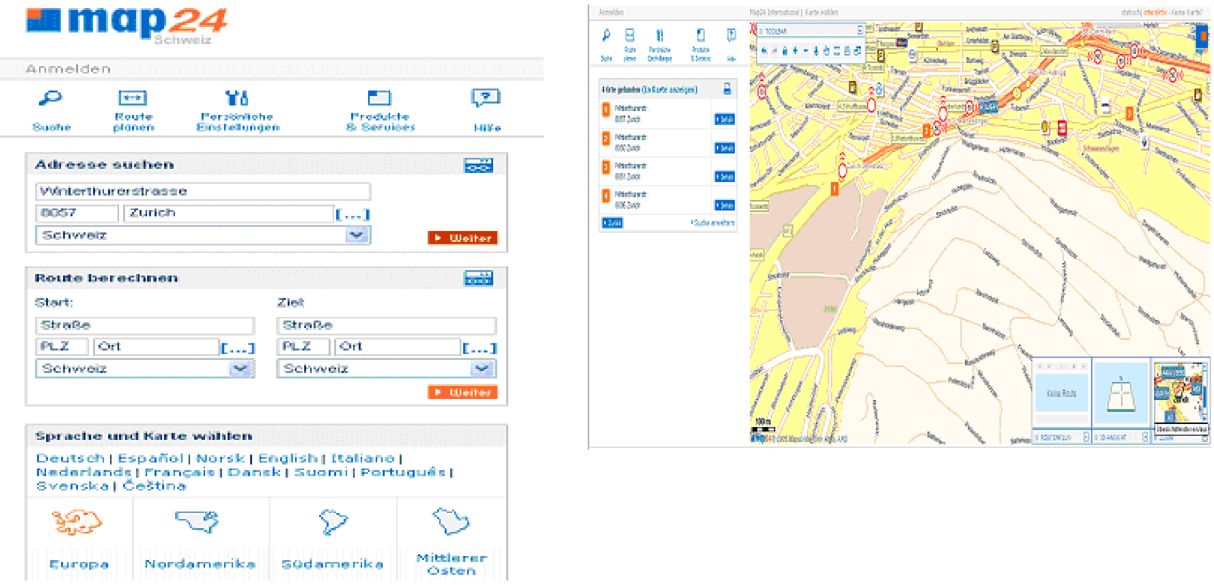


Figure 3.7: A snap shot of Starting point and result presentation of geo-web service: Map24 (http://www.map24.com)

Table 3.10: Comparison of results presentation of selected search engines.

Graphical Visualization	Grokker	Kartoo	The Brain (TouchGraph)	Google
Hierarchical clustering	Yes		Yes	
Clustering visualization	Yes	Yes	Yes	-
Map based visualization	No	Yes	No	No
Graphical interaction	Yes	Yes	Yes	No

Table 3.10: Comparison of results presentation of selected search engines.

Graphical Visualization	**Grokker**	**Kartoo**	**The Brain (TouchGraph)**	**Google**
Visualization manipulation	Yes	Yes	Yes	-
Interaction and visualization				
Highlighting	Yes	Yes	Yes	No
Coloured query result	No	Yes	No	No
Filtering results presentation	Yes	Yes	Yes	no
Cooccuring terms interaction/ visualization	Yes	Yes	Yes	Yes
Textual listing	Yes	Yes	Yes	Yes
Classification	Yes	Yes	No	

User characteristics and basic background knowledge

6 of 14 subjects considered themselves as experts with respect to their knowledge of computers in their daily life (Table 3.11). 7 of the 14 users search for information for work purposes as shown in Table 3.12. 11 of the 14 subject have widely used English as the preferred language of choice, while a few state that they often prefer to search in a language other than their native language as illustrated in Table 3. 13.

Table 3.11: User's expertise and knowledge with respect to usage of computers [Q: As a computer user, do you consider yourself as?]

Total	Novice	Basic	Moderate	Advanced	Expert
14	0	2	3	3	6

Table 3.12: Search at a work place [Q: How often do you search for information for work purposes?]

Daily	Several times a week	Once a week	Occassionally	Never
7	6	1	0	0

Table 3.13: Language preference and search [Q: How often do you use another language than your native one?]

Proficiency	English	Italian	French	German	Spanish	Portuguese
Native	1	1	1	10	0	0
Fluent	5	0	3	1	0	0
Fair	2	0	1	0	2	0
Basic	3	5	6	1	0	1

50% of the subjects, i.e 7 of 14 considered themselves having moderate search expertise, while only 2 of 14 consider themselves as expert as shown in Table 3.14.

Table 3.14: Search expertise [Q: As a searcher do you consider yourself as?]

Total	Novice	Basic	Moderate	Advanced	Expert
14	0	1	7	4	2

Among the different search engines listed in Table 3.15, Google is the most widely used, followed by Yahoo and Altavista. The familiarity of a specific search engine or the selection of a specific information source can influence how users choose their goal posts towards achieving their final objective.

Table 3.15: Search Engine usage [Q: Which web search engines do you use?]

Altavista	Google	Yahoo	Northen-light	Excite	About	Ask Jeeves	Go	Kartoo
7	14	8	0	0	0	0	0	0

Table 3.16 presents various other information sources that users are aware of and would choose preferably based on the need and the knowledge they possess. The amount of knowledge an individual possesses and their perceived understanding clearly influences search strategies and styles.

Table 3.16: Information sources [Q: What are different information sources the users are aware of , for the scenarios presented and would choose?]

Scenarios	Paper based Maps	Travel Agencies	Search Engines
1	1	2	1
2	2	2	2
3	13	13	14

Table 3.17 details the characteristics of the individuals about the search practices and styles adopted based on the information they encounter.

Table 3.17: Search strategies and styles [Q: If you do not get the result you expect, you normally:?]

Slightly change the query and search again	Completely change the query and search again	Switch to an "advanced mode"	Change the search engine	Ask somebody for help	Giveup	Others, please specify
14	6	9	6	1	3	0

As it is envisaged that the end users of the SPIRIT system are users who aim to exploit geographic information and constantly use it in their daily life. It is important for us to identify the awareness and usage of various geo-web services of such users. Table 3.18 illustrates user's experience and usage of various geo-web services.

Table 3.18: Geo-web services [Q: Have you ever used geographical information services on the web?]

	Daily	Several times a week	Once a week	Monthly	Occasionally	Never
Never	0	0	0	0	1	2
Multimap	1	1	0	0	0	0
Map24	1	2	0	6	0	0
Some-wherenear	0	0	0	0	1	0
Upmystreet	0	1	0	0	0	0
Mapquest	1	0	0	1	5	0
Mapblast	0	0	0	1	1	0

Furthermore, Table 3.19 illustrates the context of usage of these selected geo-web services in their daily life, which are planning a business travel, planning a leisure travel, looking for accommodation, to find a place, to find a service on a map.

Table 3.19: The context of usage of the geo-web services illustrated in Table 3.16 [if you have used one of the services above, in which context?]

Planning a business travel	Planning a leisure travel	Looking for an accommoda-tion	To find a map of a place	To find a serv-ice on the map	Others, Please specify
9	6	4	12	2	0

Query formulation preferences

Among the many factors that influence the outcome of the search process, the functionalities and characteristics of each of the search engines influence the individual's understanding. One important factor that has considerable influence is the ease of use in expressing the information request. Most users stated that a relatively simple text based interface is the preferred choice, enabling them to express their information request in simple natural language, which explains why Google has faired better than several other search engines for simple text based starting point. Table 3.20 supports such an argument with empirical evidence showing that of the 3 scenarios presented a majority preferred a simple text based starting point.

Table 3.20: Preferred starting points of interface [Q: What are preferred starting points in the interface?]

Scenario	Text based Simple	Map based	Text + Map
1	10	4	0
2	9	5	0
3	12	2	0

Task analysis for scenarios

During these experiments, individuals were asked to devise and breakdown the scenarios into the keywords and search goals to identify their expectations, both before and after using the system. This was accomplished by provision of two open-ended sets of written questionnaires:

- The first set to describe the search objectives in terms of keywords and search goals.
- The second set was to document individuals' interactions with existing search interfaces during the spatially aware information seeking process.

Analysing the goals and tasks for each of the scenarios given to the users, provided a greater insight into how individuals perceived each of the scenarios and also the keywords specified to express their needs to meet their final objectives. Most of the keywords chosen reflect concepts and the geographic location. Table 3.21 illustrates the keywords used for each of the scenarios presented to the users. An analysis of these keywords provides the following insights:

- Keywords are often expressed in an individual's preferred language for example: "*Nizza in German* "refers to " *Nice in french*", similarly "*radfahrar*" for cyclist.
- Keywords often are in combination illustrating "*concept*"and "*geographic location*". By geographic preposition, we mean the relationships envisioned by the users based on the scenarios presented to them.

Table 3.21: Keywords identified by the users during the early stage requirements assessment of existing search engines and geoweb services (Simulated Work Tasks) (ref. Table 3.6)

Scenario No.	Keywords identified
1.	Schwarzwald, Fahrard, Jan Ulrich, Bike tour, Heimatstadt, Merdingen, Hometown, Cycling,Velofahren, Bike Holidays, Camping, Black Forest, "Heimstadt von Jan Ulrich", Germany, Radfahren/Rattouren, Towns and Cities, Leisure, Timesports.
2.	Hotel, France, Windsurfing, Marseille, restaurants, Hiring equipment, Mediterranean coast, Shopping center, Cote D'Azur, Cannes, Nizza (Nice, France), Camoing de Gleins, French Riviera, Monaca, Pension Hotels.
3.	Aigle, Chalet, Les Diableret, Swiss Alps, Leysin, Hiking, Bergsteigen, Familienferein, Wandern, Chalet mieter, Mountaineering, Climbing.

The strategies adopted by the users during the process, when interacting with a simulated work task draw the following insights:

- Prior knowledge about a geographic location influenced the choice of the starting interface. Individuals with no knowledge about a specific geographic location, tend to learn about the geographic location by choosing a text based search interface. This, however, does not hold true to a very great extent for individuals with adequate geographic knowledge. They prefer to browse through a map based interface, by zooming to the required level of detail.
- Expression of the needs and the results obtained has been preferred in native languages.
- Goals and search keywords defined prior to the search process are abandoned and their peregrine nature shifts to a new set of goals based on the information encountered. Keywords are modified to address their newly defined objectives.
- To accomplish their final objectives users randomly choose a range of existing search engines and geo-web services provided to them. The order of usage of these search engines

and geo-web services differed based on the individuals geographic knowledge, expertise, results obtained from the search engines and required level of detail obtained from geo-web services.

- "Phrase" searching and boolean logic has been widely adopted to identify services such as *hotels, restaurants, windsurfing.*

Existing geographic knowledge

The knowledge of geography relevant to a query is a prime driving factor in the scenarios presented to the interested subjects. The studies reveal varying knowledge of a specific geographic location ("Place"). The choice of information source is however governed by the amount of geographic knowledge each of the subjects are familiar with.

The three approaches to assessing requirements, through the use of scenario development (ref. Section 3.2 Scenario Development), storyboarding (ref. Section 3.3 Storyboards and Simulated Work task) and task analysis using existing geo-web services (ref. Section 3.3 Assessment of Existing geo-web services) described above facilitated the drawing of holistic as well as serialistic views of users in terms of usage and interactivity. The insights drawn from these approaches appear to be useful to categorise user needs into three distinct dimensions as illustrated in Figure 3.2. The independent nature of these dimensions means they need to be individually addressed.

The dimensions themselves are as follows (Balley, 2002):

- User activity- a specific action by the user to illustrate a need which has been further classified into an activity to be realized and an activity to be analysed. The former relates to the user's intention to realize a specific activity immediately with the resource obtained, while the latter involves the user's need for supportive information.

- Required information content- one of the core functionalities of the SPIRIT system is to deliver the information that has a spatial component in it, which can be depicted as " something that is related to someplace".

- Required resource- the user may express specific needs regarding the very resource, he/she is looking for e.g. aerial picture, web documents, etc... One of the primary objectives of the SPIRIT system is to deliver on demand maps by rendering the geographical footprints of the web documents.

3.4 Hierarchical task analysis

Having completed the experiments described above to elicit user requirements, a hierarchical task analysis was performed to identify the functionalities required of the final SPIRIT system. This analysis took the form of a comprehensive task breakdown based on the functionalities and features that would facilitate accomplishment of a set of action or operations and aimed to map corresponding interactivities with the system. Figure 3.8 details these tasks in different phases of spatially aware information seeking process.

The spatially aware information seeking process comprises of a set of episodes, each of which aims to address a certain intermediate goals to achieve their final objectives. Each episode is carried out in three phases:

- **Definition or pre-retrieval:** In this phase, information request is expressed in one of these 3 modes: *Querying mode, Learning mode or Browsing mode.* Users adopt the querying mode when there is definite or precise indication of what their needs are. This doesn't however mean that users have adequate knowledge about a geographic location. By contrast, in the Browsing mode, users have little knowledge of what they need to know, but have an adequate geographic knowledge to kickoff their quest. Finally, when the learning mode is adopted, the users have relatively little knowledge about a geographic location. The number of iterations of information acquisition may vary based on the information obtained and users understanding.

The key elemental tasks in this phase are *locate, identify, differentiate, diverge, discriminate* and *scope* which are performed prior to the execution of their query.

- **Retrieval:** During the retrieval phase, there is relatively little interaction between the user and the system. However, when the system finds ambiguous place names user's are prompted to identify a single location through a process termed as *disambiguation.*

- **Post-retrieval:** During the post-retrieval phase, there is a relatively high amount of interaction between users and the system, whereby subjects tend to *refine, manipulate, organize, differentiate, correlate, classify, discriminate, detail , overview, simplify* and *draw multiple perspectives* of the retrieved information. The order of occurrence of each of these tasks is however undefined and subject to individual users' relevance assessment often termed as "*perceived relevance*".

The number of episodes within a spatially aware information seeking process can thus be seen to vary based on the complexity of the information needed, perceived relevance and of course the precision and recall of the system used for retrieval.

3.5 Defined requirements

Table 3.22 & 3.23 illustrate the user requirements identified during the experimental studies. However, the feasibility of certain points is questionable for example, automatic translation of geographical names and searching in multiple languages were unlikely to be realised as a feature of the first release.

Table 3.22: User Requirements List (Balley *et al.*, 2002)

1.	Support geographical concept expansion e.g. from "The Midlands" into "Nottinghamshire Leicestershire Derbyshire Northamptonshire Warwickshire Staffordshire " or from "Schwarzwald" into "Murgtal-Schwarzwaldhochstraze Baderregion-Nordschwarzwald-Weinland Ortenau-vom-Kinzigtal zum Neckar-Breisgau, Feldberg-Sudlicher-Schwarzwald, Titisee-Villingen-Schwenningen". This feature is considered essential for supporting an exhaustive retrieving of the information referred to a specific area.
2.	Support spatial concepts that relate different geographical entities as well as objects. e.g. "main station in Munich", "beaches out of Nice", "Cities in South England" as well as "Cinemas in Cardiff", "Hotels north of London"
3.	Support different terminologies and notations of spatial relations. Different terms are used in natural language to mean similar concepts. Hence the system must be able to recognise synonymous notions for spatial relations. For e.g. the term "near" is synonymous with the terms "close to", "not too far from" and the term "touch" is synonymous with the terms "adjacent" and "nearby".
4.	Support any kind of spatial relationship like "in Zurich" and " not in Zurich".

Table 3.22: User Requirements List (Balley *et al.*, 2002)	
5.	Support Fuzzy measures of distance e.g. "15 minutes drive", "20 minutes walk", "walking distance", possibly combined. This is equally true for qualitative distance as in "in walking distance". The system must also be able to interpret an approximate measure of definitive distance. e.g. within 10 miles could be matched with 8 or 12 miles etc.
6.	Automatic identification and interactive disambiguation of place name when the same name is shared with other place in the same country or in others.
8.	Support description of geographical features e.g. "main rivers in France", "abbeys in South England", "Adriatic coast of Italy". SPIRIT should be able to detect the areas or the points under discussion.
9.	Support Fuzzy geographical areas, e.g. "Les Diablerets mountain area", " South Wales", "North East London".
10.	Maps are preferred display medium • Maps have to be enriched with other types of information (mainly links and images) e.g. hotels, travel information, routes and paths. • Direct manipulation of the displayed map has to be supported; zoom-in and zoom-out are considered the basic, measuring the distances would be an interesting feature to have.
11.	Clustering of retrieved pages with respect to particular semantic concepts, e.g. information about accomodation classified in hotels, resorts, B&B, houses, huts, camping etc.
12.	Offer document ranking respect to distance from a specific place, e.g. "hotels close to the Munich Station" than the closer will be presented first.
13.	Support ranking by spatial non-spatial concepts. e.g. hotels ranked respected to the distance from the station vs. respect to the number of stars.
14.	Keep the history of search sessions for future queries. e.g. different queries, area specification and relevant results
15.	Support query by sketching e.g. allow users drawing a square on the map illustrating his current location and automatically identifying the place and retrieve related information
16.	Maps and listed information have to be visually related e.g. map position blinking when the corresponding term in the list is brushed, automatic scroll of the info list when a certain place on the map is clicked.
17.	Ubiquitous retrieval of geographical information, i.e while travelling by means of a PDA:
18.	Support multi-language naming. e.g. Black Forest has to be automatically translated into Schwarzwald, would be highly desirable. Often foreign geographic names have the original name translated, mapping multiple names on the same concept is thus fundamental for the retrieving.
19.	Some sort of automatic translation feature can be explored when essential information is retrieved only in a different language than the once in the query

Table 3.23 Broadening of SPIRIT Requirements (Bucher *et al.*, 2004 p.18-20)	
1.	Provision of advanced interface to allow users to specify additional or multiple geographic criteria, precise address location or temporal annexation of web sites.
2.	Spell check functionality or "did you mean" should be provided.
3.	Automatic place name disambiguation functionality to be provided.
4.	Proactive alternative search terms provision.
5.	Bookmarks should be categorised by search (i.e. the performed query should be recalled for every bookmark link) and linked to the maps.
6.	Provision to save book marked maps within user's profile and easily exportable.

Table 3.23 Broadening of SPIRIT Requirements (Bucher *et al.*, 2004 p.18-20)

7.	Provision of statistical information relating to a specific geographical location or tourist information and a set of predefined links to other sources of information
8.	Provision of route planning services on interactive highlighting between source and destination.
9.	Provision of statistics associated to number of documents associated to a particular area or region.
10.	Results should be ranked with respect to the criteria relevant to the particular query. e.g.: a search for accomodation in sheffield would yield results ranked by cost.
11.	Investigate the feasibility of showing topographic information on the map as a layer and if feasible, provide it. More precisely, user should be able to toggle these layered topographic data presented on the map.
12.	Provision of map tool box to explore the map, with a descriptive note on its usage.
13.	Improve the map readibility by appropriate selection of colours and symbology to represent the features.
14.	A comprehensive suite of spatial relationships based on proximity, containment, associations and distance function are presented. However precise spatial relationships " with in x minutes drive" are preferred choice.
15.	Clustering of the results obtained and subsequent highlighting of the clusters should be illustrated in the maps.
16.	Discrete representation of document footprints on the map by exploiting colour and symbols based on relevance rank.
17.	Provision of a help option explaining the usability of the interface and its functionalities.

3.6 Development of an information seeking model

The process of defining requirements for the SPIRIT system revealed a web of contributing factors that influence the spatially aware information seeking process as a whole. Based on these factors, and research discussed in terms of information seeking models in Chapter 2, an information seeking model specific to the spatially aware information seeking process was developed. The purpose of this model is to investigate the potential significance of fundamental geographic concepts, such as place, way, and region in an information space that is only metaphorically geographical in nature; which are then associated with specific functions and user actions that make sense in the context of exploring and searching an information space consisting of unstructured textual document objects. The nature of the queries posed by an individual illustrate geographic concepts expressed in natural language and require practical interpretation in terms of interacting with unstructured textual sources such as internet to make sense of vast array of possible geographically-structured representation and analytical tools for the analysis of spatial information. The purpose of this model was also to investigate the potential metaphors in the design of visualizations for spatially aware information retrieved from unstructured textual sources such as the internet.

An informavore (an individual who consumes information) chooses a specific course or strategy based on the visual and linguistic cues encountered. Observations drawn from the studies, reveal that the spatially aware information seeking process comprises of the six key components illustrated in Figure 3.8.

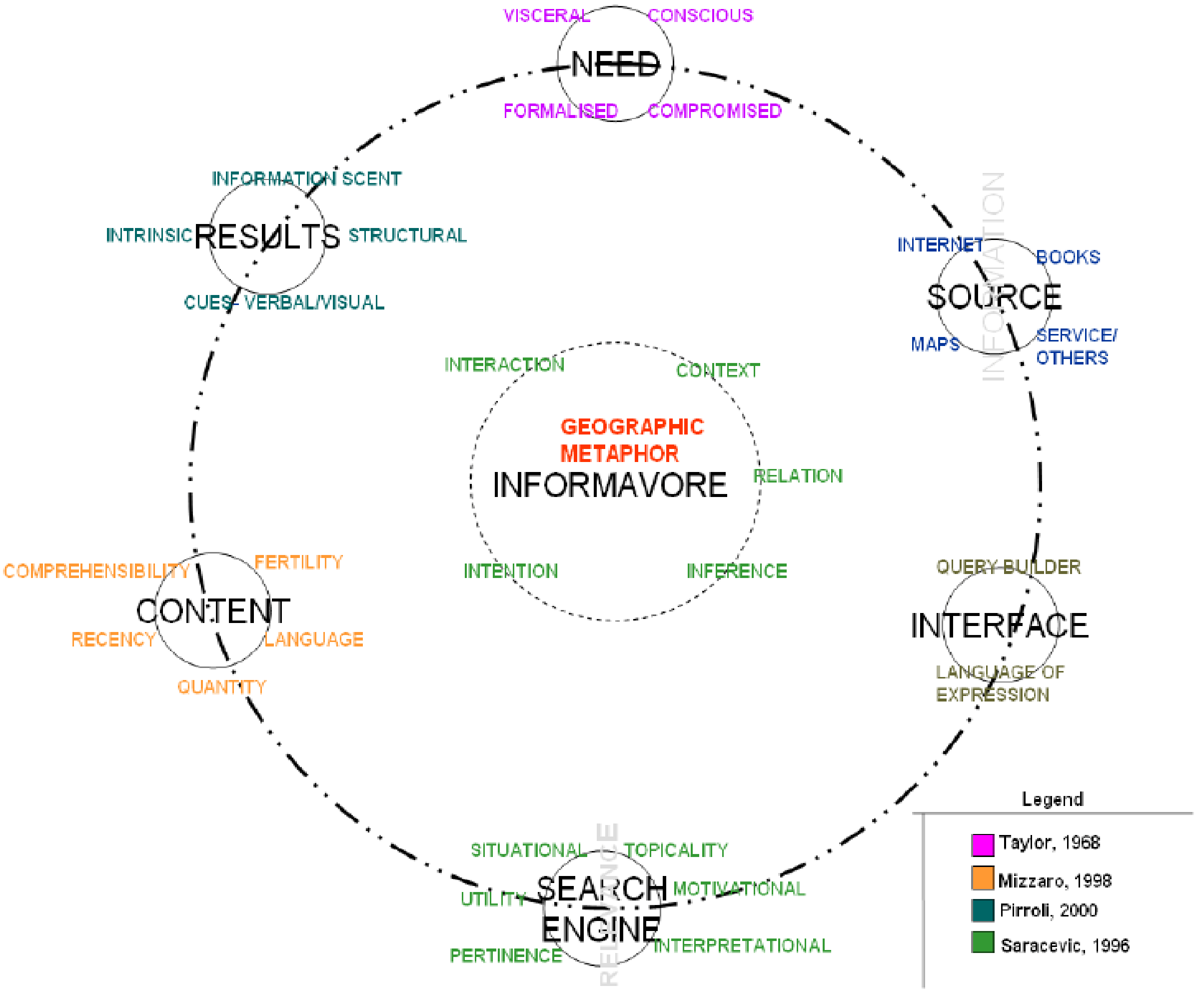


Figure 3.8: An episode of Spatially Aware Information Seeking process influenced by the works of Bates (1986), Taylor(1968), Mizzaro(1998), Pirrolli(2000), Saracevic(1996). The model illustrates "an episode" based on berry picking model.

Figure 3.8 An episode of the Spatially Aware Information Seeking process influenced by the works of Bates (1986), Taylor(1968), Mizzaro(1998), Pirrolli(2000), Saracevic(1996). The model illustrates "*an episode*" based on berry picking model.

Intent is often expressed in the form of a "*need*"; Taylor (1968) argues that need can be categorized as "*Visceral*" (the actual but unexpressed need of information, corresponding to real information need), "*Conscious*" (the conscious description of the need, corresponding to perceived information need), "*Formalized*"(the expression of user's need, as it would be expressed without the presence of the IR system) and "*Compromised*" (the expression of user's need as presented to the IR system, corresponding to the request).

The number of iterations or episodes to achieve a user's final objective depends on geographic know-how and selection of an information source that delivers the precise information needed. The findings of this study reveal that individuals prefer to choose specific information source based on the scope and depth to which each of these serves their purpose.

Presuming that the selected information source serves the purpose the need can be expressed in brief. Keywords are identified by users that facilitate the best possible match from the collection to increase or decrease the number of iterations or episodes. Interfaces act as a medium for expressing their need by proactively engaging subjects with possible suggestions and feedback and provide a structural framework to the information seeking episode.

The structural framework built by the interface thus serves as an input for retrieval engine for the best match called the *search engine*. The best matched results are ordered based on their relevance. To date, a benchmark relevance criteria hasn't been adopted, systems define relevance to suit the interests they serve. Saracevic (1996) identifies some key relevance criteria namely: *situational, topicality, utility, motivational, pertinence, interpretational.*

The possibilities for automated match based on expressed need (keywords) varies. Perceived relevance is governed by the comprehensibility (*how much of the retrieved documents can be easily understood*), fertility (*how much of the retrieved documents will be useful for finding other documents e.g. number of reference of the documents*), recency (how much of the retrieved documents have to be recent), language (*in which language the retrieved documents have to be written*), quantity (*number of documents and their length a user wants*) of the content (Mizzaro, 1998).

The results obtained from the best possible match defines the course of the search process for it is governed by the visual and linguistic cues termed as "*information scent*" (Pirrolli, 2000). If the final objectives are achieved, subjects terminate the process, else the process continues and each episode defines a specific goal towards achieving their final objective.

During each of these episodes, subjects (informavores) tend to adopt a geographic metaphor which predominantly influences their search strategies and objectives. Numerous other factors influence the decision making process of informavores such as *intention, interactions, context, relation and inference* (Saracevic, 1996). Innitency of these factors influencing the spatially aware information seeking process cannot be ruled out and the order of occurrence.

The model illustrated in Figure 3.8 builds on *berry picking model* presented by Bates (1986) and incorporates significant contributions with respect to each of the components outlined. Thus, a comprehensive theoretical framework for spatially aware information seeking model has been established educing from earlier scientific contributions and experimental studies.

3.7 Summary

Differing design and user perspectives reflect the functional areas a system serves to address. Communication of these perspectives between the designers and developers is critical for the success of any system. This chapter presents the requirements study carried out and validated by involving subjects with varied experience and domain expertise. Different phases of requirements gathering and analysis have been constructed not only to reflect upon the design and development aspects of the SPIRIT system, but also with respect to its perceived usage and functionality.

It has been demonstrated that scenarios and illustration of fictious situations to individuals provided a feel of a real setting and created real empirical and qualitative data that could be exploited as a part of the design cycle. The results were obtained from analyzing the spatially aware information seeking behaviours of real users, with actual and real-life needs modelled in a laboratory setting.

Users were independent to choose and refine strategies to achieve their final objectives. As a part of an iterative design cycle, re-assessment exercises were carried out to validate the early stage requirements by expanding the focus group to other domains. Qualitative feedback received resulted in a complete set of data to be evaluated. Combining the qualitative and quantitative data collections methods provided subjective and objective insights with respect to individual's spatial awareness and information seeking. The experiments helped us enlist a detailed task structure of spatially aware information seeking process. The users envision a system with spatial intelligence capabilities that address their locational information needs.

In the latter half, a conceptual model for spatially aware information seeking is constructed based on scientific contributions and experimental studies. Understanding the complexities involved in the whole process, a framework is adopted with a set of functional components each serving to address a key functionality of the interface. Thus, laying the foundation for designing a spatially aware information retrieval system.

Chapter 4

Visualization approaches

“The greatest challenge to any thinker is stating the problem in a way that will allow a solution”

-Bertrand Russell

A key focus of this thesis is to investigate the degree to which various contemporary visualization techniques and principles can serve to deliver functional views that suit the needs of the users looking for information relating to a geographic location. The focus of this chapter is to present potential approaches to visualizing the results of geographic searches, based around the key functional requirements identified in Chapter 3. In Chapter 5 the prototype implemented in SPIRIT is described in detail, but the data available for visualisations are briefly discussed here.

Dykes *et al.* (2005, p.107) state that the “*motivation to develop novel graphic representation and interaction techniques can come from the data itself and our abilities to access and use it*”. Conflation of unstructured data retrieved from the internet through an extraction mechanism (retrieval) with geographic data sets (e.g. SABE Data or Digital Chart of the World) draws interests across different domains. “*Whilst technological advances and availability of new data drive a need for new views, Casner (1991) convincingly demonstrated that the same data needed to be represented in different ways using different views in order to effectively serve different information needs. The information needs that require these views are defined by different tasks*” (Dykes *et al.*, 2005, p.109). Furthermore, Fairbairn *et al.* (2000) state that there has been a constant change in graphic information representation and handling methods, including those defined as “*cartographic*”, and that there is a need to investigate new methods of representation and understanding of these representations.

Visualisations of the results of searches for information relating to a particular location must effectively integrate geo-referenced search results with background contextual information. Methods for delivering background contextual data in the SPIRIT project are briefly described below. The process of searching for geographically and contextually relevant documents is described in detail in Chapter 5. Essentially, geographic references in web pages must be identified, disambiguated and grounded to a given geographic extent (Clough *et al.*, 2004). For the purposes of visualisation this means that the results of a search are a set of documents, each of which may have one or more geographic extents associated with it, and these geographic extents may vary in both spatial extent and location.

4.1 Delivering background web mapping

One of the key elements of the user interface is the provision of interactive maps, with minimum required functionality including panning, zooming and selection for query and results formulation (See Chapter 3). Many examples of web mapping applications exist, for example: Multimap (www.multimap.co.uk), Map24 (www.map24.com) and Digimap (www.edina.digimap.ac.uk). These tools generally allow the user to pan, zoom and select different elements for display. Simple querying by using clicks on the map may also be allowed. Integration of such tools into the SPIRIT interface allows us to produce simple background maps suitable for querying, and to visualize search results as points or other symbols on maps. Within the SPIRIT project the open source platform Deegree was used to provide web mapping through a web-coverage server

4.1.1 Degree OpenGIS web servers

With the web as delivery platform, there has been growing interest in the Open Geospatial Consortium (www.opengis.org) which has created a number of specifications and recommendations designed to allow data from a variety of sources to be integrated through internet applications and delivered to the users as data or maps. Amongst these standards those for GML (http://www.opengis.org/techno/specs/01-068r3.pdf) and Web Feature Servers (http://www.opengis.org/techno/specs/02-058.pdf) are central to the SPIRIT interface.

The Geographic Markup Language is an XML based standard designed to facilitate exchange and representation of geographical data including both spatial and aspatial properties of features. Web Map Servers (WMS) deliver images on receipt of an HTTP request to a server, and allow a server to build a map as a set of layers and styles. Finally, Web Feature Servers (WFS) allow a user to make a query for spatial data which is served up as GML. By using a WFS integrated with a WMS with some layer in between performing visualization tasks any data which can be delivered from a WFS (in other words that can be represented using GML) can be delivered to a client in a graphic or picture format. e.g: GIF, JPEG, PNG and SVG.

Deegree (www.deegree.org) is an open-source Java product for the implementation of local and web based GIS applications. It's interface and architecture guarantee optimized interoperability by conforming to OGC standards. The implementation of the Deegree WFS is based on this architecture with a WMS, a WCAS (Catalog Service), a Web Coverage Service (WCS) and a Web Coordinate Transformation Service (WCTS). The available version of the Deegree WFS is based on the OGC web feature service implementation specification 1.0.0 (OGC WFS Specification WWW) and serves the request of GetCapabilites, GetFeature, DescribeFeatureType and Transaction (with some restrictions) (www.deegree.org).

Shape files can be directly connected by the Deegree WFS via a JDBC driver. Alternatively, shape files can be imported into an Oracle Spatial database, then the WFS can access the Oracle databases with the Spatial Extension (8.1.6 and higher) and use their access-mechanism for querying. Tests within SPIRIT established Deegree handled simple geometric objects of the type likely to be requested in visualization well. The Deegree WFS can output data in both XML and GML. Presently, the Deegree WFS can be run by posting an XML file to a servlet, the XML file includes the simple query specification. For example: all three geometric objects in a rectangular region can be queried by a simple bounding box.

4.1.2 Data for background mapping

A key requirement for the SPIRIT project was the delivery of relatively seamless mapping across a range of European countries. Since a relatively small number of pan-european geographic data sets exist some potential candidates for use in visualisations are briefly described here. Digital Chart of the World (DCW) has a global coverage at a source scale of 1:1,000,000 (http://www.maproom.psu.edu/dcw) and consists of points and polygons representing major communication, population centres and some other topographical data. (Clough *et al.,* 2004). Eurographics produced a data set called SABE (Seamless Administrative Boundaries of Europe) representing the administrative boundaries for the whole of Europe at scales of 1:100,000 and 1:10,00, 000 which includes political boundaries to the level of communes represented as their polygons or centroids (www.eurographics.org/products/SABE)(Clough *et al.,* 2004). Since these data only represent political boundaries they are thus in themselves not suitable for providing useful background information. However, as discussed in Chapter 5, they form an important source for the geoparsing process and the associated grounding of web documents with spatial extents. Furthermore, they are used in some of the visualisation tasks illustrated later in this thesis. Finally, TeleAtlas (www.teleatlas.com) produce a seamless dataset, MultiNet which details roads, streets, points of interests, land use etc.

Appendix-F shows that large scale datasets offer very varying feature classes, while at smaller scale particularly around 1:2500,000 there appears to be more homogeneity. In addition to these data sets a number of gazetteers with variable feature classes are also available.

4.2 Web interface constraints

With the web as the information delivery model, a number of constraints come into play during the design. A set of technical considerations were made for the SPIRIT project, which included (Syed *et al.,*2003):

- **Web page validation:** In an effort to achieve consistency and conformance with W3C standards all web pages should conform to W3C validation standards (http://validator.w3.org)

- **Accessibility of web pages:** compliance with web content accessibility guidelines. Some countries such as the U.K have disability discrimination act, which is a legal requirement.

- **Minimization and standardization of client-side scripting:** a common source of non-standard behaviour is the use of client-side scripting, for example through javascript which can have unpredictable effects when used with different browsers. Client-side scripting should be minimized and the standard version proposed by ECMA-290 used (www.ecma.ch).

- **Use of plug-in's:** plug-in's often provide a plethora of additional functionality to interfaces. Information relating to installation and enabling of such plug-in's should be clearly instructed to the end users.

- **Interface compatibility:** adherence to standards unfortunately does not guarantee that the interface will operate under a certain system configuration. The SPIRIT consortium adopted a set of operating systems and browser configurations for testing. Also a testing suite of stable machines (or virtual machines) were used to document compatibility.

4.3 Ideation-Concepts for visualizing spatially aware information retrieved from the internet.

The main focus of this chapter is to present the initial concepts developed in the ideation phase of the development of the SPIRIT interface. In order to address a range of scenarios, consistent with the conceptual spatially aware information seeking model presented in Chapter 3, a series of initial interface designs were developed. These interface designs seek to incorporate key functional requirements (§ref. Table 3.22; 3.23) gathered during the requirements gathering process discussed in Chapter 3. The user interface is at the epicenter of the SPIRIT system, brokering the information flow between the key tasks of the retrieval system (Ontology, Ranking, Retrieval Components). The nature of these functional interactions is described in Chapter 5, whilst we focus here on presenting the initial set of interface designs used in the design process for SPIRIT.

Dykes *et al.,* (2005, p.110) state that "*advocates of task-driven tools do not only look at data but also look for something 'interesting', such configurations that may contribute to better understanding of the data or underlying phenomena.. While an explorer may not be aware of their tasks or know their specific outcomes, those designing tools that support ideation must consider them explicitly and deliberately to instruments that it can assist in the observation of distributions, behaviours, expose patterns and facilitate detection of relationships. Taking into account the concurrency of exploratory tasks it may be inappropriate to follow the approach of building separate graphical realizations for each task. Instead one should try to design and support a range of tasks, possibly through various methods of interactions*".

In this spirit, the tasks related to the SPIRIT interface were analysed and a range of interactions likely to be performed with the system in each search session or search episode identified. Table 4.1 details the phases of interactions with the SPIRIT interface. Furthermore, Fu *et al.* (2003) and Balley *et al.* (2002) detail key functional requirements of the SPIRIT interface.

Table 4.1: Task Analysis for the SPIRIT interface (Syed *et al.,* 2002)

S.No	Phase	Description
1.	Query formulation	The user will be able to specify a query relating a theme and a location, possibly related by a spatial relationship.
2.	Query disambiguation	ambiguous places or concepts are identified by the system automatically and a range of alternatives or choices are presented to the user
3.	Query expansion	a user query will be automatically expanded by adding terms appropriately
4.	Result presentation	results will be presented based on the user's needs.
5.	Search refinement	the user will be able to select a subset of the results to refine their search
6.	Feature control	the user will have the option to turn on and off features of the system, for example: query disambiguation, spatial indexing and ranking types

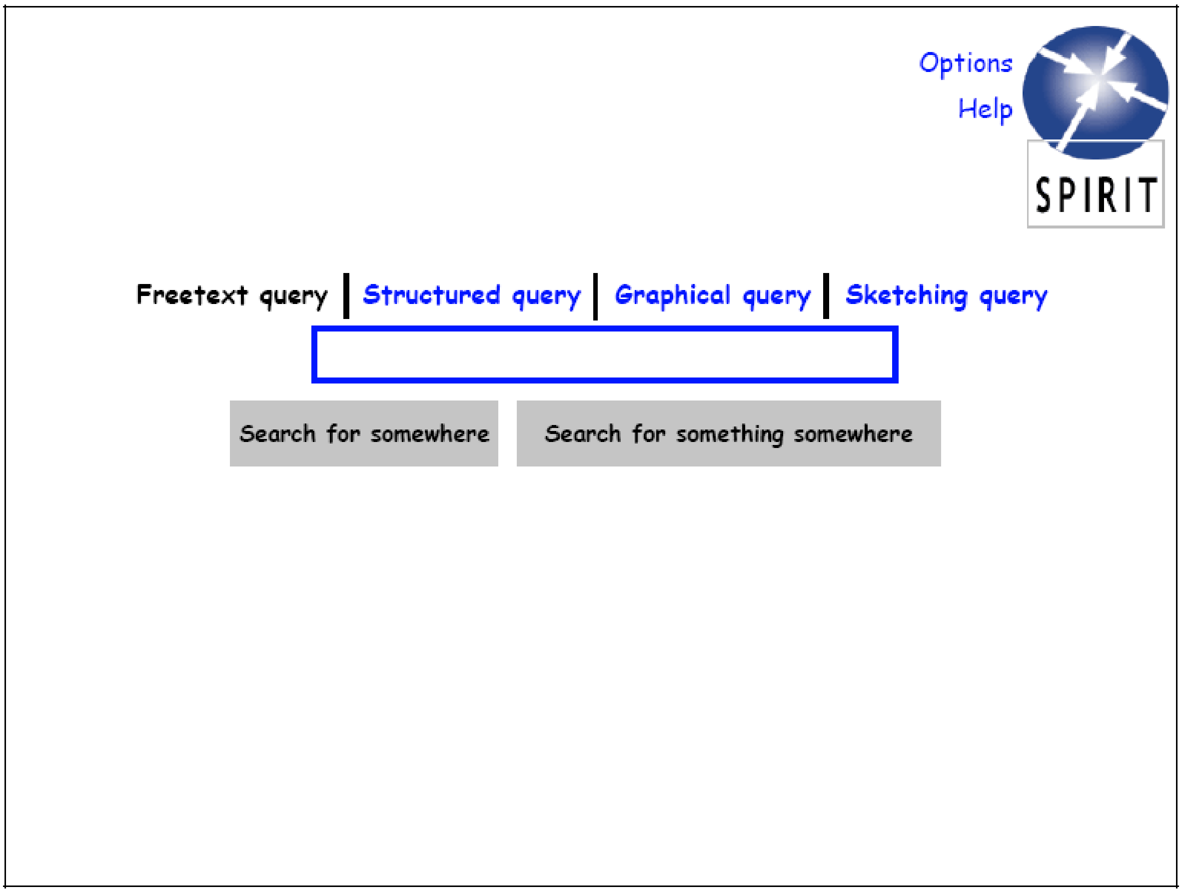


Figure 4.1: Free text querying: A design template(Syed ***et al.*****,2002)**

4.3.1 Query formulations

The first task of the interface is to allow the user to specify a query, representing their information need. A commonly used interface is a simple free text querying interface as shown in Fig 4.1, where users express their needs in natural language. Most search engines such as Google (www.google.com) and Yahoo (www.yahoo.com) have adopted such a simple interface because of its perceived ease of use. The main requirement for the interface stemming from the Ontology report of the SPIRIT project (Fu *et al.,* 2003) relates to the nature of the queries that the user makes. A typology of queries is presented which basically consists of a combination of a place name or place type with or without an aspatial entity. These terms may be related by a semantic relationship or a spatial relationship. The typology is characterized as a relationship between "*something*" and "*somewhere*", as shown in Fig 4.2. In formulating such queries there are three basic tasks that need to be performed:

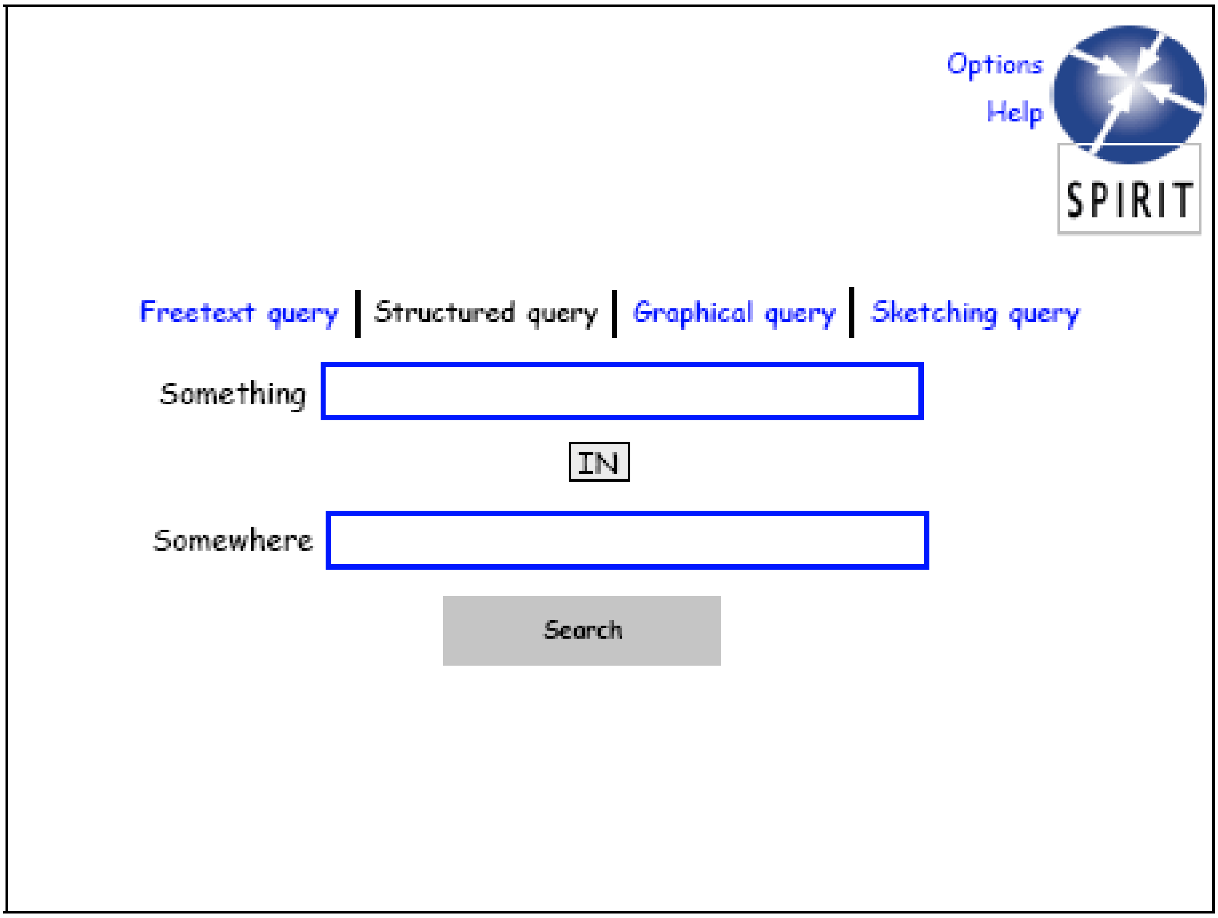


Figure 4.2: Structured query interface: A design template(Syed *et al.***,2002)**

- Identify the geographic extent of the query. The geographic extent of a query requires us to provide a location in space through some referencing system. In turn, this implies that some mapping is available between different reference systems- for example mapping an address to latitude and longitude. Geographic extents may in principle be entered either as text or through some form of interaction with a map.

- Identify the nature of the aspatial concept related to the place. The entering of a concept is a trivial task which can be achieved either through the entering of a text, or in the case of a domain specific ontology through the selection of a node on the ontology. It should be noted, however, that association/assignment of a word to a particular place in a concept hierarchy is non-trivial, presumably because a word may relate to multiple concepts.

- Specify the nature of the relationship between the place and the thing. The nature of the relationship between objects and places (*whether it is to be spatial or not*) is more complex. Many synonyms exist for descriptions of relationships and providing an exhaustive library is likely to be difficult. Simple spatial relationships often do not conform to, or even exist in, our mental maps - for example: is London north or south of Paris? Fu *et al.* (2003) suggest that interactive components should be provided to facilitate the input of spatial relationships but it will be necessary to consider usage issues when deciding a spectrum of functionalities for expressing relationships. The exact functionality required for query formulation needs to be well defined.

The potential complexity of the possible query formulations described above suggests that a range of interfaces incorporating differing levels of functionality and trading off ease of use for complexity should be developed. A key element identified by Balley *et al.,* (2002) is the provision of specifying location using an interactive map, as shown in Fig 4.3. This interface facilitates users to draw a region of interest on the map and specify the spatial relationships for the selected region with respect to a specific concept.

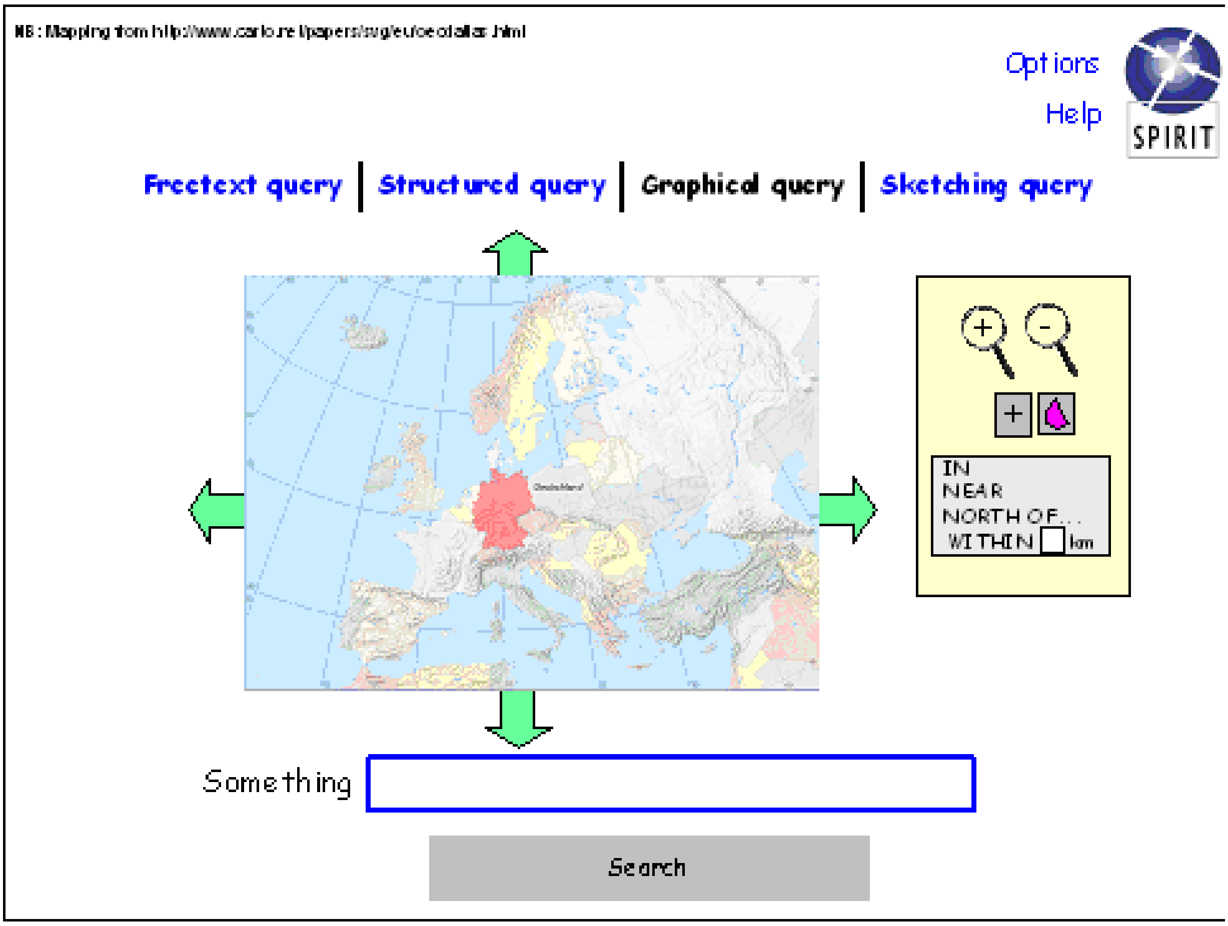


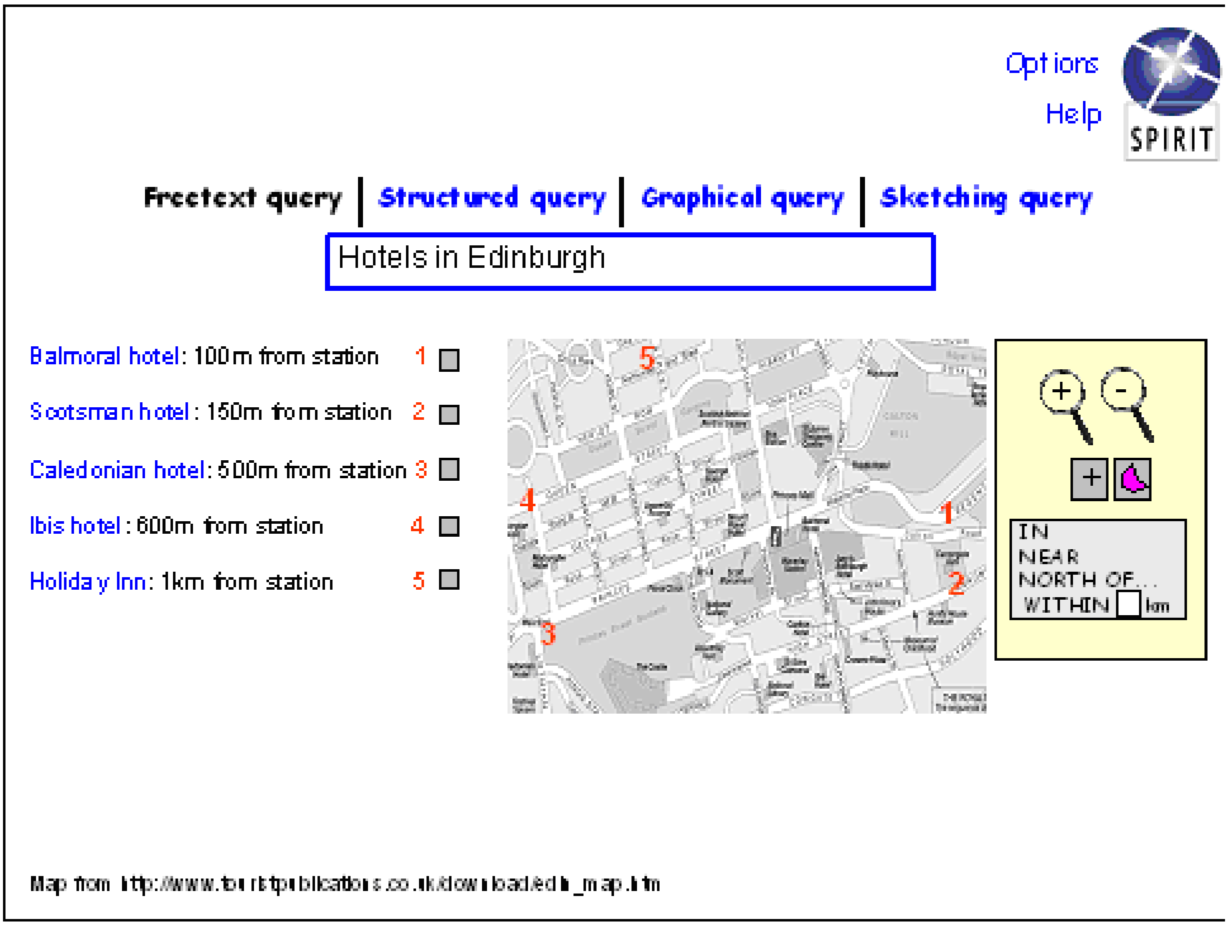
Figure 4.3: Graphical query interface using web mapping. : A design template(Syed *et al.*,2002)

4.3.2 Query disambiguation

Balley *et al.* (2002) and Fu *et al.* (2003) suggest that automatic identification of ambiguous place names prompts users with a list of choices to disambiguate ambiguous place names. Thus, for a query such as “Museums near London”, the system should prompt the users to interactively select from a list of possible options such as:

- London, United Kingdom.
- London, Ontario, Canada.

As many place names are ambiguous, it is likely that provision of such functionality may prove frustrating as users will be prompted to disambiguate most place names specified at the interface. Thus some consideration should also be given to weighting the likelihood of a query being ambiguous before a user is prompted to refine it.



Hotels in Edinburgh

Balmoral hotel: 100m from station 1 ☐

Scotsman hotel: 150m from station 2 ☐

Caledonian hotel: 500m from station 3 ☐

Ibis hotel: 600m from station 4 ☐

Holiday Inn: 1km from station 5 ☐

Map from http://www.touristpublications.co.uk/download/edh_map.htm

Figure 4.4: Map based presentation of the results: A design template (Syed *et al.*,2002)

4.3.3 Query expansion

Query expansion is suggested in the specification of system functionality as allowing the system to expand a query to include places related by some spatial relationship or alternative place names (thus a query related to London might also use place names found within London (e.g. Westminster)). Since automated expansion is likely to take place after a query has been entered then consideration of expansion is only likely to occur when the user is presented with results. The interface should make clear the expanded query used to obtain the results and allow the user to refine this if required.

Queries such as "*cities on the mediterranean coast of France*" and "*hotels in and around St.Maxime*" cannot be grounded into geographical features to serve as a basis for hierarchical query expansion, but have a reference to the feature type for the relevant scope of interest instead. In the latter query, the term "*cities*" are not only feature type, but also a conceptual illustration of a specific region (a morphological variant). Based on the conceptual illustration, textual similarity may or may not help us in identifying the most relevant set. The interface should not only facilitate geographical query expansion with reference to features but also with feature types. Furthermore, a query such as "*accommodation near London*" has varied interpretations based on the semantics of proximity relationships. The user is interested in accommodations that are only located in London or the user is interested in accommodations that are both inside london and in its vicinity. The system should be in a position to accomodate query expansion process based on the geographic scope intended by the users interactively.

4.3.4 Result presentation

Any search with SPIRIT (or other spatially aware search engines) is likely to return a set of documents ranked according to some combination of spatial and semantic similarity. This is of course one of the primary intentions of SPIRIT and in this section we introduce a range of potential visualization approaches which utilize the spatial nature of search results, and in some cases their conceptual and spatial ranking.

4.3.4.1 Map based representation

Figure 4.4 illustrates the interface design for simple map-based presentation of the results. A map with numbered links is displayed alongside a list of relevant documents. The document list provides links to web resources which may consist of one or more pages, and thus may include an intermediate step. By selecting an item from the list, the search can be refined using the spatial location of this item, and the map manipulation tools used in normal map query formulation remain available for query refinement. The map itself is zoomable and the maximum extent is defined by the bounding rectangle of the retrieved documents. It is also envisaged that a simple text based results presentation be made available, similar to that which users are familiar with through current web search engines. Documents in the document list will be ranked according to the measures developed to describe semantic or spatial similarity or some combination of both – that is to say a relevance ranked list of documents. See Chapter 5 for more on methods used in relevance ranking in SPIRIT.

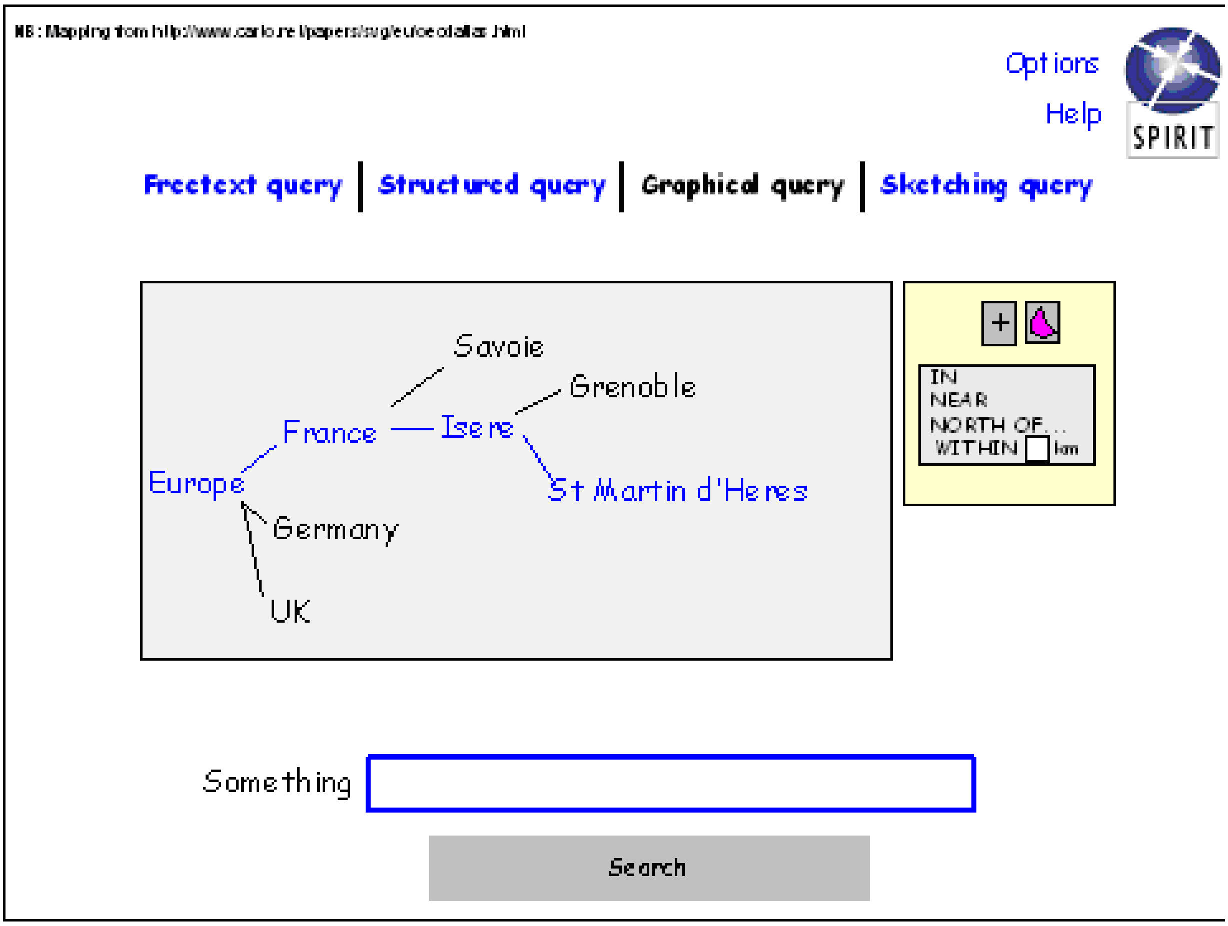


Figure 4.5: Ontology based traversal of query space: A design template (Syed *et al.*,2002)

4.3.4.2 Ontology based traversal of the query space

Ontology based traversal of the query space is a tangible choice to explore result sets. Alani (2003) exploited the potential of TouchGraph visualization (http://www.touchgraph.com) in visualizing ontologies. Touchgraph provides highly interactive environment with fast rendering, pan and zoom capability, locality control, etc. Using the TouchGraph open source Java environment, documents can be placed at nodes with the branches linking nodes from the ontology. For aesthetic reasons, a "*spring-layout technique*" is incorporated whereby nodes repel each other while edges (connections) attract, placing semantically similar nodes close to each other. Fig 4.5 illustrates the interface design for a potential technique of traversing ontologies with in the query space.

The next few visualizations illustrate a number of potential techniques for graphically displaying the results of SPIRIT searches, which all attempt to display a large number of documents through a variety of symbolizations. The visualizations themselves are once again intended to be illustrative and to provide techniques for summarizing information which will meet the needs of the users with different cognitive understandings of semantic and spatial similarities.

4.3.4.3 Cartograms

"A cartogram is a map or diagram showing geographical & statistical information. More specifically, the word cartogram has been defined as a 'combination of map and graph" (Wilkie, 1976, p.1).

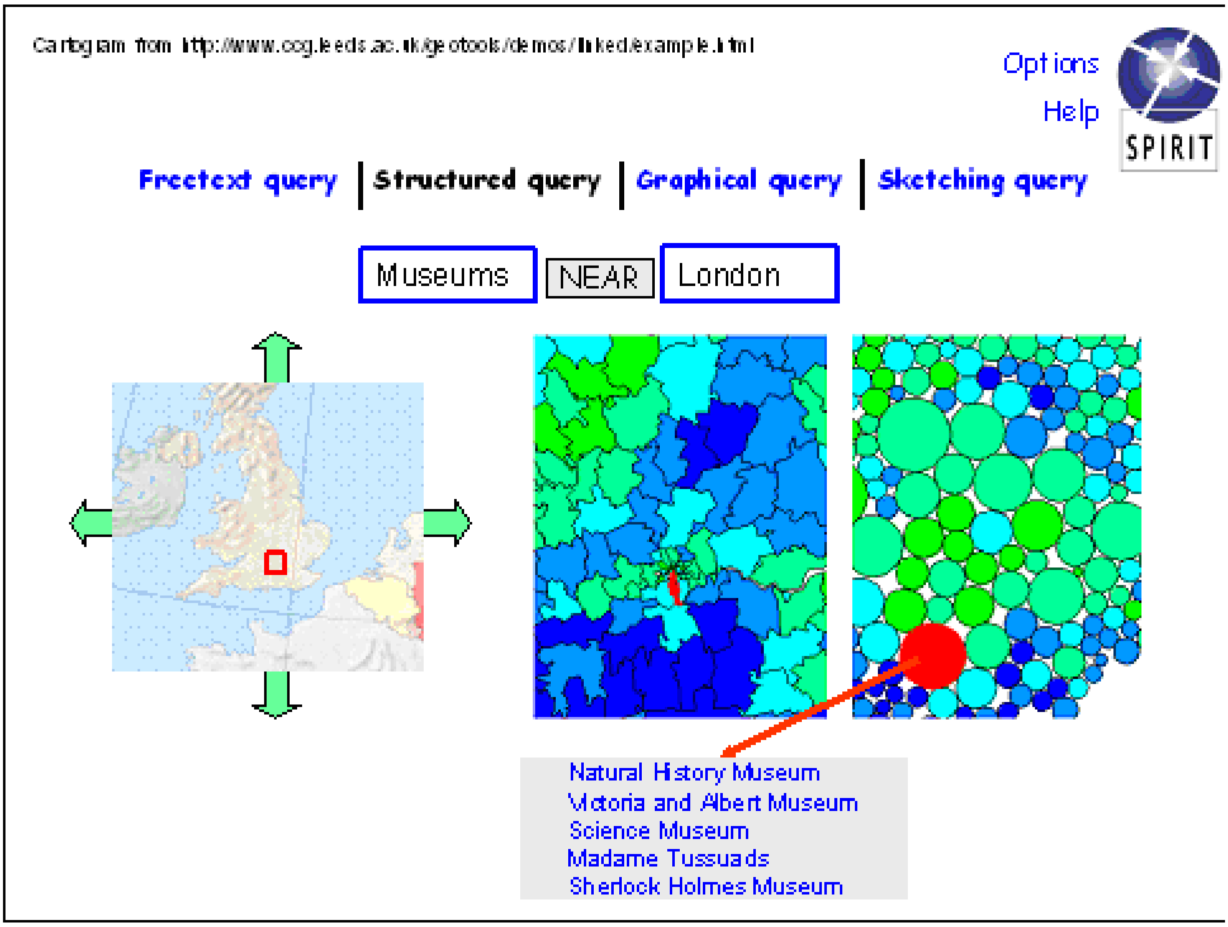


Figure 4.6: Results presentation using cartograms- note that the centre map is used to illustrate the concept, and that in implementation a topographic map would be displayed here: A design template (Syed *et al.*, 2002)

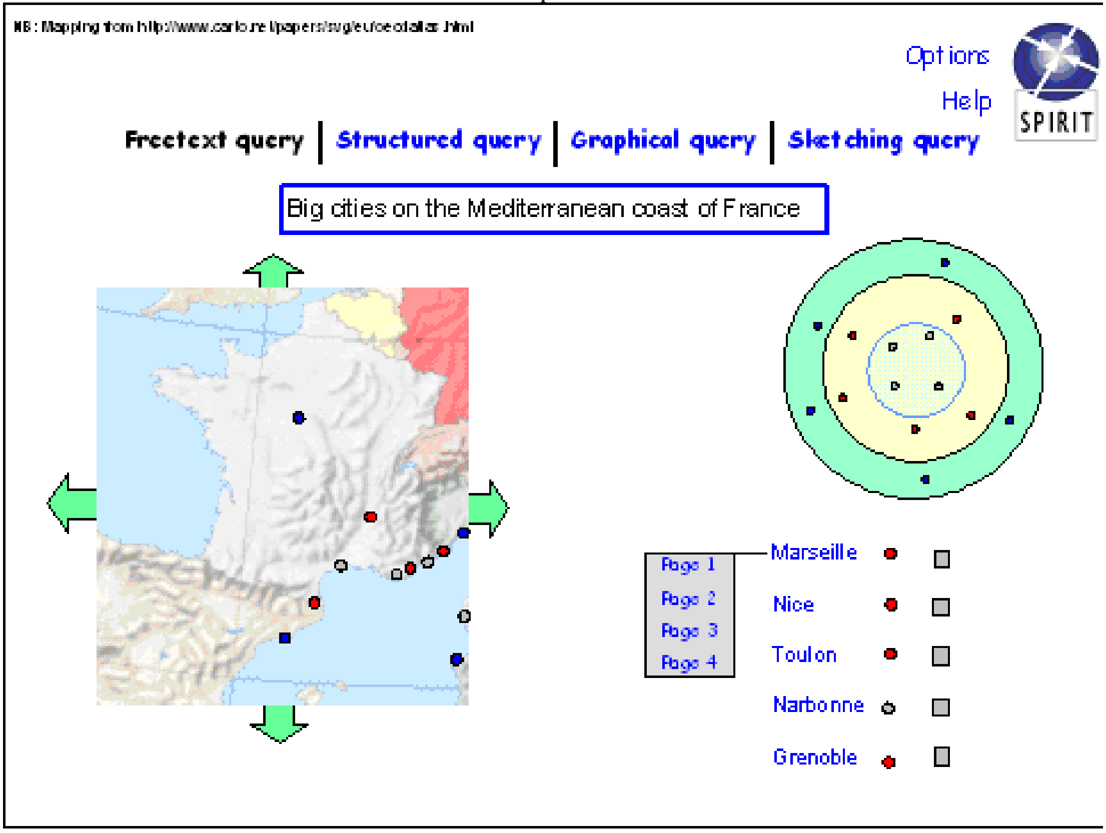
Fig 4.6 shows the use of circular cartograms to display the result of a spatial search for "*museums near London, United Kingdom*". The central map shows political units as stored in the SABE dataset. This map would in reality be transparent to the users, except for the highlighted polygon and more meaningful data, for example a simple topographic map, displayed here. On mouse over either on the topographic map or the cartogram, selected polygons and their corresponding cartograms are highlighted using a technique widely known as "*brushing and linking*" (Martin and Ward 1995; Monmonier, 1988,1989).

The size of the circle in the circular cartograms can be varied according to some weight associated with the polygon, which might in this case be the number of relevant documents contained in the spatial area or the overall semantic or spatial relevance of the documents in this area, or some other meaningful score. As the elements of the cartogram are selected so relevant documents are displayed in the window below. Finally, an orientation map on the left allows the user to pan and zoom within the linked cartograms.

Such visualizations assume some linkage exists between the data used in visualizing search results and the method of allocating search results.

4.3.4.4 Target Metaphor

Fig 4.7 illustrates an interface design using a so-called "*target metaphor*". A target metaphor is a simple metaphor which suggests the degree of similarity of the results to a query. The target metaphor positions documents relative to a query footprint expressed by a used according to some spatial relationship (e.g. topology, direction or proximity) The bands within the target can further demonstrate the scope of the documents with respect to the specified geographic location. The target is linked to the map by brushing and relevant results are displayed when an element is moused over. Individual elements can be used to redefine the spatial or semantic context of the search.



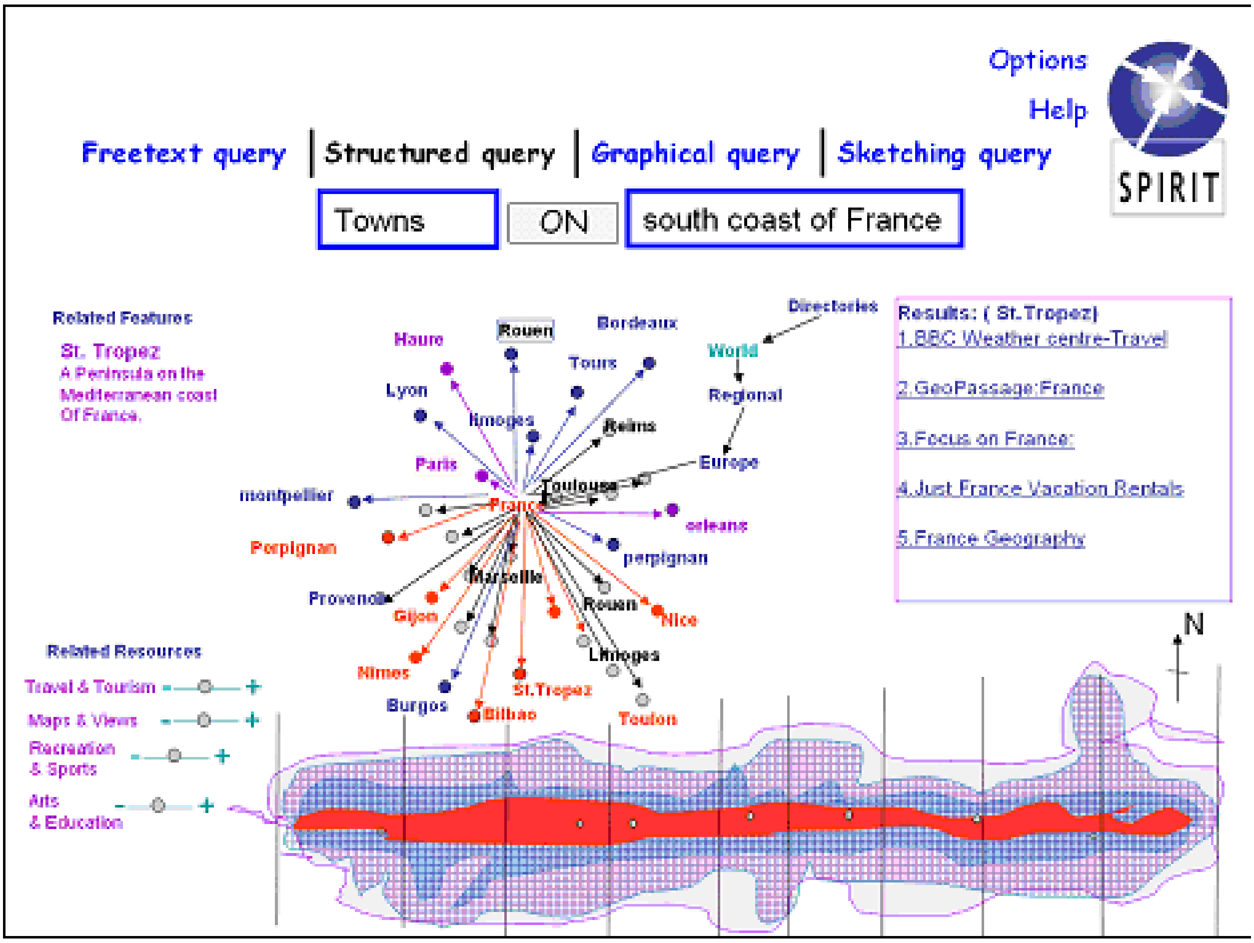
Big cities on the Mediterranean coast of France

Figure 4.7: Visualization of query results using a target metaphor: A design template (Syed *et al.*, 2002)

4.3.4.5 Touchgraph and ThemeRiver Visualization of Spatial Relations

The higher level of interactivity and explorability of Touchgraph can also be exploited to explore and illustrate the spatial relations within documents. Another visualization approach which demonstrates the spatial relationships of the documents is ThemeRiver (Havre *et al.*, 2002).

Combining these two visualization approaches it is envisaged that users may navigate through the hierarchy of place names and visualize the documents based on their physical location. On selection of a specific document, relevant spatial context and footprints of that document are displayed in the results. These documents can be further highlighted by references to a popular landmark. The pan feature gives a detailed description about the geography of the region selected by the user and their relative locations. Fig 4.8 illustrates a foreseen solution for a query "*Towns on south coast of France* " incorporating ideas from both the Touchgraph and ThemeRiver visualization techniques.



Towns

south coast of France

Results: (St.Tropez)

1.BBC Weather centre-Travel

2.GeoPassage:France

3.Focus on France:

4.Just France Vacation Rentals

5.France Geography

Figure 4.8: Interface design for ThemeRiver visualization for results presentation (Syed *et al.*, 2002).

4.3.4.6 Statistical Approach

Fig 4.9 illustrates an interface design based on visualizing documents on a two-dimensional plot, with geographic and semantic relevance as axes of the planes. Document positions on the plane indicate their position in cartesian space that are also displayed on an accompanying map. Spatial and semantic relevance is suggested by the graphical variables colour and height (size). For a query such as "*Big cities on the Mediterranean coast of France*"; The BBC weather forecast document has higher relevance according to Google search engine, In contrast, SPIRIT could have the same document plotted on a two dimensional plane according to geographic and semantic relevance.

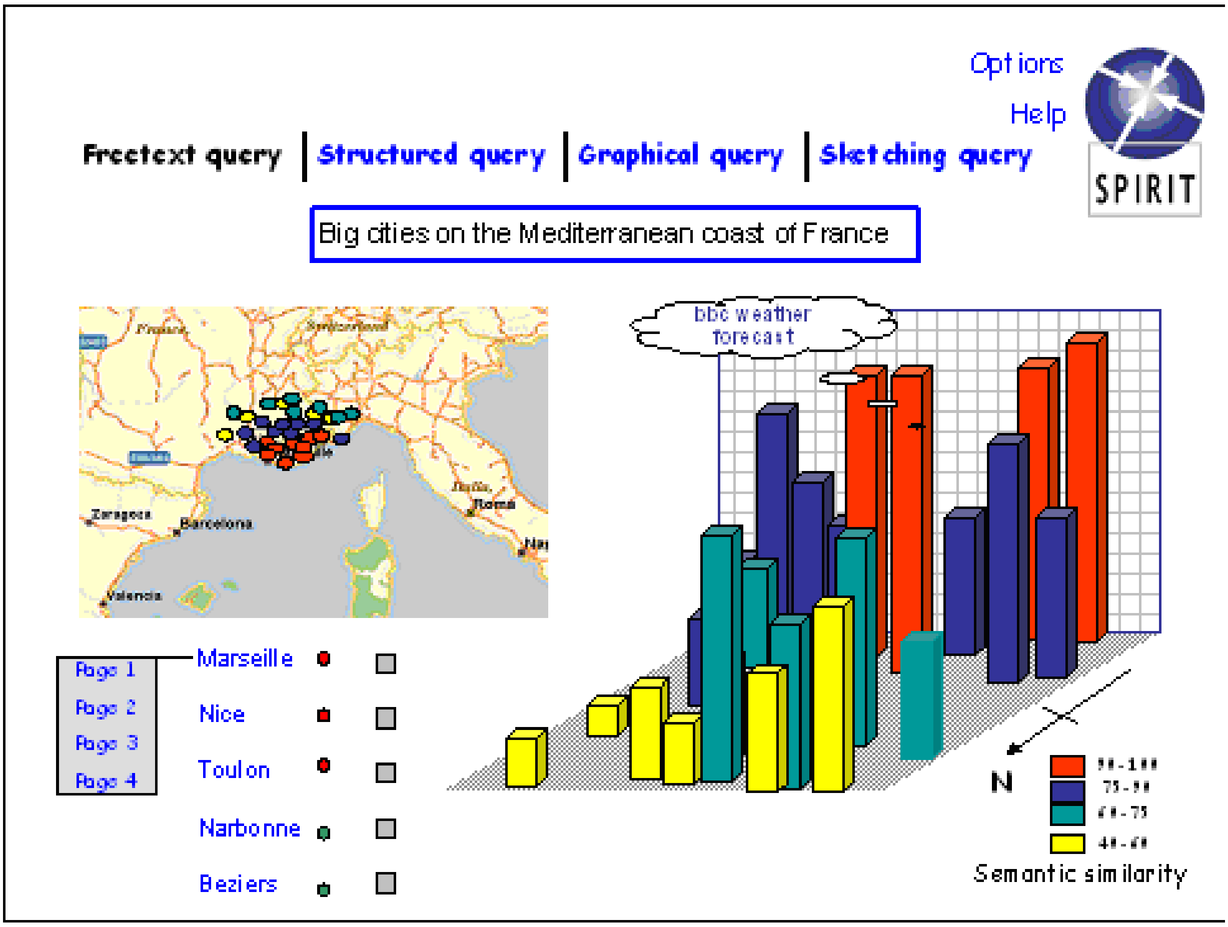


Figure 4.9: Query visualization using a statistical approach to illustrate spatial relevance vs distance from the centroid of expressed region of interest. Themes illustrate the semantic similarity of the documents.

4.3.4.7 Density Surface representation

Figure 4.9 demonstrates a statistical approach where in geographic and semantic relevance of documents are visualized as bar charts, and equal sub-intervals of the whole data are considered. This however does not consider local neighbourhood influences. Kernel density estimation considers such local neighbourhood influences, by centring a kernel function at each data point (in this case the geographic footprint of the document) and using a smooth kernel function to estimate the density over a region with the number of documents at a footprint used for the frequency at the datapoints. Moreover, kernel density estimators smooth out the contribution of each observed data point (geographic footprint of a document) over a local neighbourhood of that data point. Influence on a data point (geographic footprint) X(i) and X(n) depends on the shape of the kernel function adopted and primarily its bandwidth. (Hwang *et al.*, 1994)

4.3.4.8 Footprint based representation

For a web page referring to multiple geographic locations and varied contextual information, the geographic scope of the web page is complex. Figure 4.10 illustrates one such case where in a web page refers to locations such as *Scotland, Nepal, Alps, Norway, Glenmore, Greenland;* footprints of which are different and so the geographic scope of that document is relatively different. A possible solution would be to represent these footprints individually on a map giving an overview of the extent/scope of a web page and the underlying relationships between the same footprints occurring in different web pages.

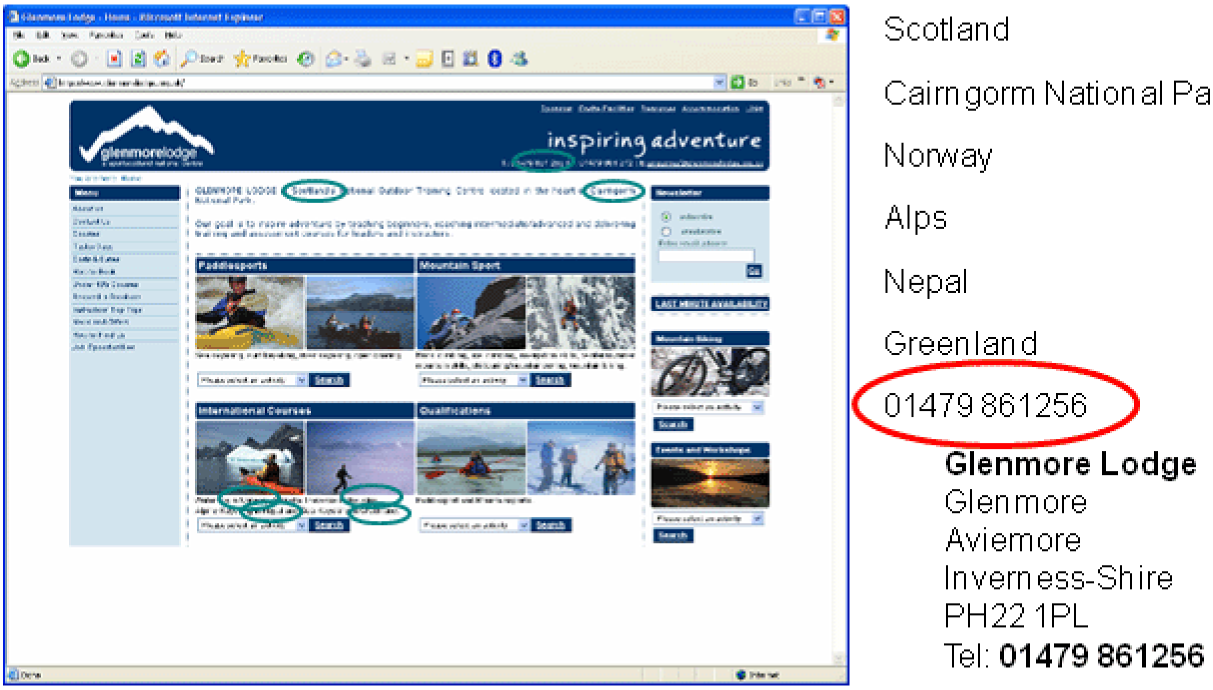


Figure 4.10: An example of geographic references on web page referring to various geographic locations around the world along with addresses, telephone numbers and postal codes. Snap shot of Glen More lodge web site (source: http://www.glenmorelodge.org.uk/)

4.3.4.9 Spatial Activity map

Identifying additional geographic information surrounding a geographic reference or geographic footprint which are specifically derived from web pages can serve to be extremely useful in defining activities or resources related to these web pages or locations. This is one of the stages of information extraction can be termed “contextual information gathering”. Almost all human activities are rooted at a specific geographic location. Thus, additional information gathered during the contextual information gathering phase can be used to build “spatial activity maps”. However, since individual web documents often contain relatively little distinct information, the task of defining a theme or activity is challenging. A possible approach to representing such information could be through the use of symbology to illustrate this additional information related to resources or activities at a specific geographic location. The success of this approach is very dependent on how the geographic ontology and associated metadata are constructed.

Each of the above mentioned approaches (4.3.4.1 to 4.3.4.9) to results presentation serves different purposes addressing a variety of potential user tasks. Many of these share some common characteristics with contemporary visualization techniques. A number of points should be noted about these interface designs for result presentation:

- All of the user interfaces are designed without recourse to the use of scroll bars to display more information, particularly in the form of long lists, addressing the web interface constraints and serving the standard resolutions of 800x600.

- Interface designs all include some provisions for result refinement through reformulating a query. In some cases user may select from a retrieved documents and use their spatial context as a new location.

- All the interface designs display many documents, in some cases aggregating these documents into spatial or semantic units. Such approaches, which we believe are essential to demonstrate visualization techniques in SPIRIT, place requirements (discussed in chapter 3) on the nature of the retrieval results passed to the interface.

- Many of the designs use both spatial and semantic aspects in visualizing results. Once more this will place requirements on the information retrieved.

4.3.5 Search Refinement

The search Refinement phase is one of the most prominent natural phases of search process as a whole, it is when user's try out a range of possible query refinements or expansions to close the search loop which may ends either in the success or failure of the user's search. Most of the existing search engines have suffered from the common malady of pigeonholing or stereotyping, as soon as supplementary terms are added, the query quickly becomes too specific to find the desired page(s). By allowing some interaction, the query replacement button seeks to alleviate this problem (when coupled with the right terms). Query refinement or expansion can be either done manually and interactively by the searcher (interactive query refinement or expansion) or automatically (automatic query expansion). *"Intuitively interactive query expansion should produce better results than automatic, however this is not consistently so* (Beaulieu, 1997; Koenemann and Belkin, 1996; Ruthven 2003)" (Vechtomova and Karamuftuoglu, 2004).

In SPIRIT, we intend to adopt intuitively interactive query refinement or expansion, where supplementary terms are added to refine the collection of documents retrieved from the initial information request. Lau and Horvitz (1998) present different patterns of the search refinement process illustrated in Table: 4.2.

Table 4.2: Analysis of Patterns in Search Refinement Process (Lau and Horvitz, 1998)

1.	New query: A query for a topic not previously searched for by the user during the process within the scope of the web collection of the data set available.
2.	Generalization: A query on the same topics as the previous query, but seeking general information than the previous query.
3.	Specialization: A query on the same topic as the previous query, but seeking more specific information than the previous query.
4.	Reformulation: A query on the same topic that can be viewed as neither a generalization nor a specialization, but a reformulation of the prior query.
5.	Interruption: A query on a topic search on earlier by a user that has been interrupted by a search on another topic.
6.	Request for additional results: A request for another set of results on the same query from the search service.

4.3.6 Feature Control

A user centric environment should facilitate users to interactively select through preferences or choices based on which the results can be retrieved, these preferences can be either of geographic relevance techniques or spatial indexing technique or number of results to be displayed per page etc... Provision of such feature control facility for geographic information retrieval engines and for personalization of the results would serve users needs. It is envisaged that such a feature control facility should be provided to the users to define their preferences.

4.4 Characteristics of the results and the visualization requirements

During the course of this Chapter, we have focussed on methods for formulating queries to and representing the results of a query from a spatial search engine. With a single web page referring to many geographic locations, the nature of the footprint obtained from the results may vary from well-defined points to collections of polygons and visualizing such footprints as points or collection of points may make it difficult to convey meaningful information to the users, for instance the identification of relevant documents. It is also evident that the result sets may have identical footprints which cannot simply be stacked on the top of each other. These difficulties in visualizing the requirements identified demand that multiple perspectives be adopted to best serve user needs. Table 4.3 illustrates the visualization requirements and the characteristics of the results.

Table 4.3: Visualization requirements and the characteristics of the results (Yang *et al.*, 2006; Purves and Yang, 2005)

1.	Represent documents and their associated footprints.
2.	Represent footprints associated to same location and their relation with other documents.
3.	Support level_of_details (LOD) visualization of search results.
4.	Allow users to interactively manipulate search results (e.g. Position identification of documents).
5.	Allow brushing and linking between footprints and documents to indicate the spatial location of documents.
6.	Support multiple views for the comparison of the representations of search results.

4.5 Summary

This chapter has focused on the development of possible approaches for querying and representation of information retrieved from the unstructured textual source i.e. internet. This chapter runs through query formulation, extraction and result representations. The key aspects of the representation discussed in this chapter are discussed below:

- Free text querying: Free text querying requires that queries be resolved into place names, spatial relationships and contexts. By providing two query buttons this parsing process may be simplified and it is suggested that the interface undertakes some initial simple parsing of the query using information available from the ontology. The use of natural language queries and the search results obtained using textual indices need to be investigated. (Syed *et al.*, 2003)

- Graphical querying: Adequate tools should be provided for the users to graphically express the region of interest and the corresponding context the user's are looking for. (Syed *et al.*, 2003).

- Ontology-based querying: The querying mechanism shown in Figure 4.9 can use the structure of the ontology itself. It is also suggested that the results can be presented using the ontology in aggregating it into spatial units. In both cases, these requirements imply that it be possible to use the ontology in visualization. It is suggested that the common use of SABE data set would be one means of accomplishing this, with some form of a simple look-up table allowing the interface to refer to ontological units (Syed *et al.*, 2003).

- Query disambiguation: Consideration should be given to intelligent query disambiguation using term frequencies and related spatial units (Syed *et al.*, 2003).

- Search Refinement: Where spatial query expansion occurs consideration should be given to the portrayal of such information for the users and the usage issues there in (Syed *et al.*, 2003).

- Spatial referencing of the results: Most visualizations presented in this chapter will require that spatial referencing of the results be passed to the interface along with the results. Consideration as to potential aggregation of results also implies that appropriate information for spatial aggregation be passed to the interface (Syed *et al.*, 2003).

- Semantic and spatial ranking of results: The system should be capable of ranking the results based on geography and context. The visualizations consider that more than one ranking technique are obtained from the geographic information retrieval engine (Syed *et al.*, 2003).

All of the visualization approaches presented address different purposes and reflect to the multidimensional ranking based on semantic and spatial aspects obtained from the geographic retrieval engine. These visualization have been designed for the web as the delivery platform and serving a standard resolution of 800x600 pixels.

Chapter 5

Prototype

"See things as you would have them instead of as they are"

--*Robert Collier*

The early stage requirements (cf. Chapter 3) presented key geographic information retrieval functionalities based on which a working prototype has been developed by the SPIRIT Consortium, which comprises of a number of components (as shown in Figure 5.1) interacting with each other.

Users are equipped with tools that enable them to formulate their information request using an interactive query interface which allows both textual and graphical queries. The results are ranked on the basis of their combined relevance to a specific geographic location and to a concept of interest.

Approximately 94 million web pages resulting from a crawl of the web form a collection of one terabyte. Based on this collection a set of documents (Joho and Sanderson, 2004) are characterized according to their geographical context and refer to parts of the United Kingdom, France, Germany and Switzerland. A multi-modal interface enables users to search through results of geotagged collections of about 90,000 documents with the help of a number of software components each serving specific functionalities. The SPIRIT prototype system is deployed in a distributed architecture using SOAP (Simple Object Access Protocol), with components existing on remote sites and communicating via defined functional interfaces. These functional components are dealt with in detail in the subsequent sections of this Chapter.

5.1 Functional Components

The key components which impart the spatial awareness functionality to the SPIRIT system as a whole are described in detail in this section. The roles played by these components can be categorised based on their functional interaction in two phases: The *pre-processing phase* and the *run-time phase*. Figure 5.1 illustrates each of the interacting system components, assigning them to the phases and highlighting them by colour coding.

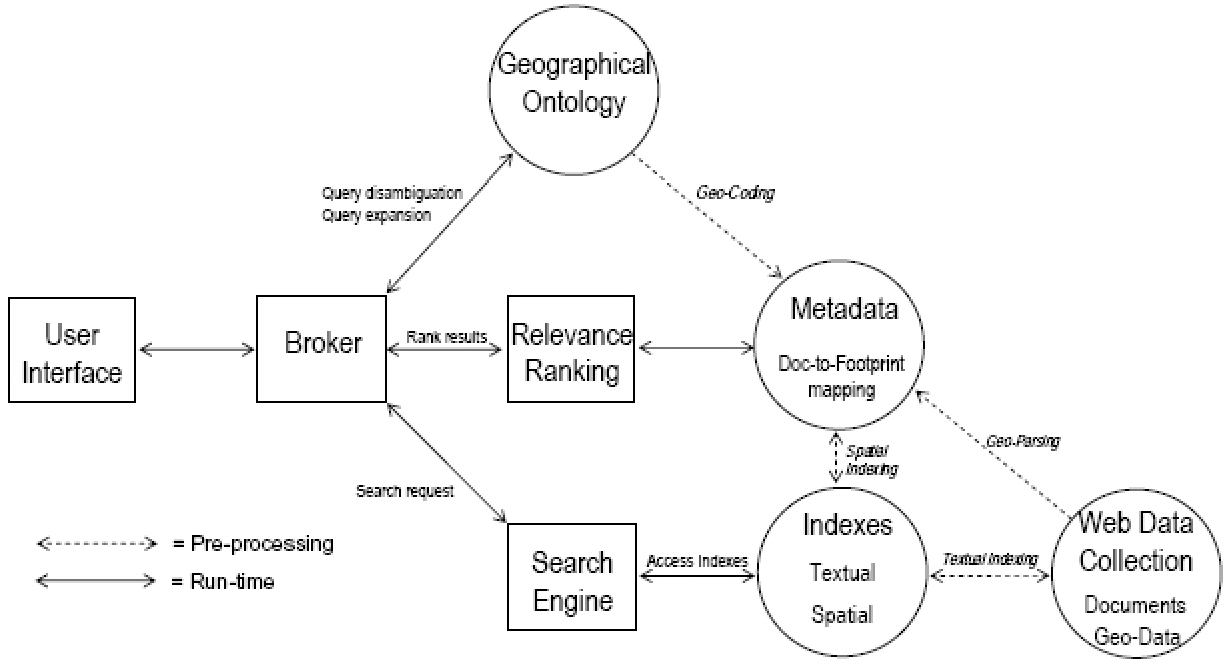


Figure 5.1: SPIRIT architecture showing run-time and pre-processing components and linkages (from Purves *et al.*, 2007, Figure 1).

5.1.1 Pre-processing

This phase involves the creation of searchable indexes which can be accessed by the run-time components of the system. During this process, the web documents are assigned spatial footprints based on the metadata describing both spatial and aspatial terms within a document.

5.1.1.1. Assigning spatial extents to web documents

Numerous authors (McCurley, 2001; Gravano *et al.*, 2003; Sanderson & Kohler, 2004; Himmelstein, 2005; Zhang *et al.*, 2006) experimentally support the notion that much of the information available on the web describes some geographic location. Hill *et al.* (1999) expressed the need for geographical reference of the spatial information within the web and assigning them a spatial extent or spatial coordinate, often termed ‘geographic footprint’. Larson (1996) and McCurley (2001) identified these two tasks as *geoparsing* and *geocoding* respectively. The overall process of geoparsing and geocoding is called geotagging (Purves *et al.*, 2007, p.13., Clough, 2005).

Geoparsing:

Parsing is performed using the GATE (General Architecture for Text Engineering) Information Extraction (IE) system (Cunningham *et al.*, 2002). The geographic terms expressed in a user’s query are identified using a combination of lists of known locations, organizations and people derived from gazetteers together with rules which capture elements of the surrounding context.

Gazetteers act as a simple, effective and language-independent lookup medium but they fail to distinguish between locations not used in geographical context (Mikheev *et al.*,1999). For e.g. ”Kansas City Shuffle” appearing in the lyrics of a song can be tracked down as a city within the U.S; or it can be taken to mean the organization Kansas City Transit (McCurley, 2001). Similarly, a movie name such as “Hyderabad blues” can be mistaken as a city of India. Hence, there is a need for an approach to apply context rules and additional name lists (proper names and commonly occurring terms) to filter out names in the gazetteer lists that are most likely to be used in a non-geographical sense.

Table 5.1: Summary of footprints (UIDs) identified by geoparsing for the SPIRIT collection (from Purves *et al.*, 2007; Table 1)

Country	**#Documents**	**Unique UIDs**	**Unique UIDs occurrence**	**Average UIDs per document**
United Kingdom	339,819	25,841	1,541,442	3.97
France	363,183	7,504	959,104	2.61
Germany	79,491	2,648	321,362	2.85
Switzerland	87,009	5,832	258,188	3.1

The SPIRIT prototype incorporates a gazetteer lookup, supported by the geographic ontology (described in Section 5.2.2) and was populated with two main sources of data (though only one of these was applicable outside of the U.K.) – the SABE (Seamless Administrative Boundaries of Europe) data set and the Ordnance Survey (1:50,000) scale gazetteer. Locations include regions such as villages, towns, cities, counties and places of interest represented spatially as points and polygons. An accuracy of 72% and 25% false positives for all annotations was demonstrated by using gazetteer lookup with additional context rules during the experimental studies carried out by Clough (2005) for a collection of 885, 502 web pages from the original one terabyte collection. This illustrates that web documents either contained no location or contained locations which could not be grounded. Table 5.1 presents a summary of footprints (Unique IDs) extracted from the processed web pages. The number of UIDs indicates the total number of unique locations identified within the collection. Within the document collection a total of around 1.5 million references to UK place names were found. Given that around 340,000 pages refer to a UK place name the average number of UIDs per document was about four. It is important to note that these grounded UIDs included falsely grounded place names (e.g. unidentified instances of place names such as "Jack London") and that the distribution of place names is likely to be strongly biased towards a much smaller subset of locations. The most striking example of this bias is London which occurs in 112477 documents in the collection. That is to say that around one third of the documents in the UK collection have reference to London (as well as potentially other place names). The results are not surprising, since many administrative organizations in the UK are based in or near London (Purves *et al.*, 2007).

Geocoding:

Within the web documents there are multiple references to a single geographic location, which in these cases have to be replaced by a default location associated with location metadata. Several approaches have been proposed to determine a default location, including:

- *the most commonly occurring place names* (Smith and Mann, 2003);
- *by population of the place names* (Rauch *et al.*, 2003);
- *by semi-automatic extraction from the web* (Li *et al.*, 2003).

Each of these locations was assigned an appropriate bounding box, representing a spatial extent derived from polygonal data stored in the geographic ontology (described in Section 5.2.3). Since the overheads associated with passing polygons through the system were too high, around 89% of all place names were grounded correctly (Clough, 2005). This is actually a geoparsing result, as the geocoding result is analysed rather than the exact co-ordinates.

5.1.1.2 Building Document Indexes

SPIRIT adopts a hybrid indexing scheme to retrieve text documents with respect to geographic context. The text indexes are generally maintained as an inverted file structure (Salton & McGill, 1983) using GLASS Search engine[1]. Spatial access methods (Samet, 1990; Rigaux *et al.*, 2002) can be used for performing point range and interval queries. Some of the spatial access methods such as regular grids, quadtrees, R-trees, and the like can be used in conjunction with the inverted file structure to provide a hybrid indexing structure for addressing spatial queries. A uniform grid scheme as shown in Figure 5.2 is employed here as in some successful commercial GIS Systems (Vaid *et al.*, 2005). This index divides the entire footprint coverage of the document collection into a grid of rows and columns. For each cell of the grid, a list of document IDs was constructed using the document footprints that were a result of the geotagging process.

Based on the best match of the query terms, only those documents that intersect the geographical scope of the place name in the query (spatial extent of the query) are filtered. This has been achieved by matching the resulting spatially indexed document footprints with those documents that lie within the spatial extent of the query. This scheme is referred to as "T" (separate text and spatial indexes) in Table 5.2.

Among several other approaches is spatial-primary spatio-textual indexing (ST), an individual text index referring to documents whose spatial extent lies within the respective grid cell and is associated with each of the cells of the spatial index grid. At run-time (i.e. when a search is executed) those cells that intersect the spatial extent are determined and then only the corresponding text indexes are searched.

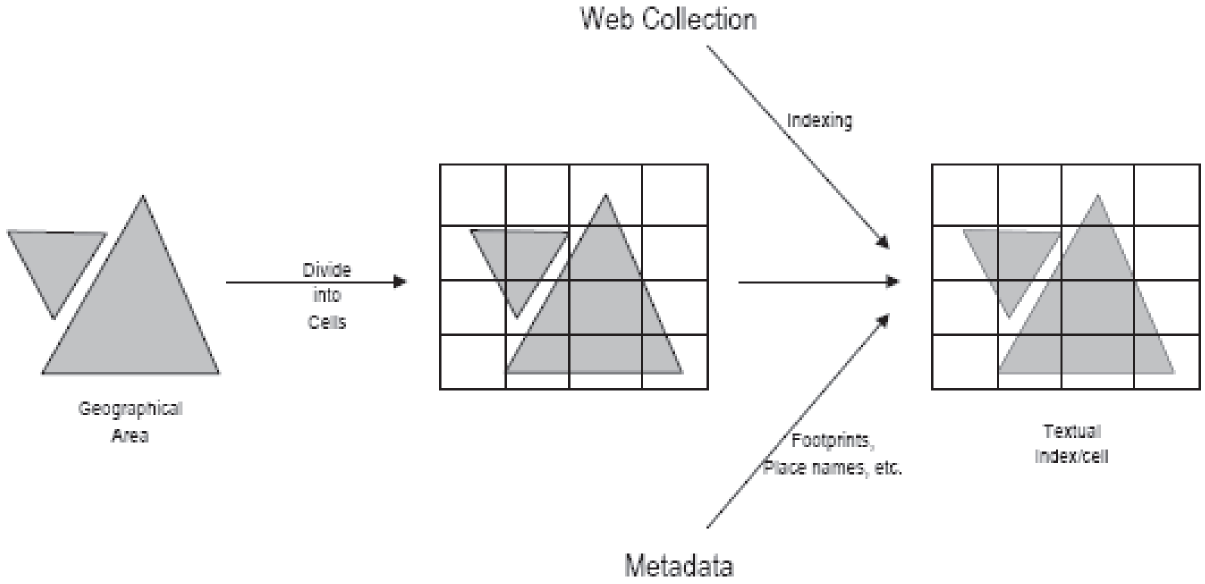


Figure 5.2: Spatial Indexing illustrating fixed grid scheme (Vaid *et al.*, 2005)

---

[1] For more details see http://dis.shef.ac.uk/mark/glass/

Table 5.2: Spatial Indexing Schemes (Vaid *et al.*, 2005)

1.	Pure Text (PT)	It employs the basic GLASS (Search Engine) text indexing procedure, which accesses a file-based lexicon using a binary search on the sorted index terms. These include all geographical and non-geographical terms.
2.	Space-primary spatio-textual indexing (ST)	In this approach the documents indexed are those whose footprints (spatial extent) intersect the cells. A UNIX file system is constructed bearing cell IDs, and the relevant text indexes are extracted by matching the cells with respect to the query.
3.	Text-primary spatio-textual indexing (TS)	Contrary to the earlier approach, the index scheme is built by modifying the document occurrences and lists in the GLASS index. For each indexed term, the occurrences list is segmented into cell-specific sub-lists, which refers to the footprint intersecting the respective cell. For ease of access, each occurrence file comprises of header data providing offsets of the start and end of each cell-specific sub-list.
4.	Separate text and spatial indexes (T) Also referred as Text indexing with spatial post-processing.	In the T indexing scheme the pure text index component is identical in structure to that of PT, while the spatial index consists of a table containing records with the structure [Cell_id, document_list].

The results are then merged and sorted based on the ranks of the documents from each of the intersecting cells. An alternative approach to this was implemented with text-primary indexing, where an inverted list text index is extended so that for each term the associated documents (containing the term) are grouped according to the spatial index cells to which they relate, as determined by the intersections of their document footprints (*spatial extents*) with the cell. Thus a form of spatial index is built for each list of documents for each term. This scheme is referred to as “TS” (text-primary spatio-textual Indexing; Vaid *et al.*, 2005).

A rather different approach has been adopted in Pure Text indexing (PT) where in all terms either geographical or non-geographical are indexed and a binary search is carried out to filter out the matching terms. Vaid *et al.* (2005) demonstrated that in ST and TS schemes, there is a penalty of multiple copies of indexes (multiple text indexes for ST and multiple spatial indexes for TS). Thus the overhead increases with the number of footprints that are selected during the geotagging and subsequently in the creation of the spatial index. This combined effect of multiple cells and footprints per documents constrains the index structure to coarse grid resolutions as the amount of total disk space in storing indexes becomes very high. Furthermore, Vaid *et al.* (2005) demonstrated that indexing TS required slightly more space than ST but exhibited better query response times. In contrast the T scheme resulted in longer query times (up to double) but with little storage overhead. With increasing spatial grid resolution, all spatio-textual schemes behaved the same way returning fewer documents reflecting the closer approximation of the grid cells to the query footprint with increasing resolution (Vaid *et al.*, 2005). Considering the above experimental results the text-primary indexing technique (TS) has been used in the prototype presented and the evaluation experiments are carried out for the same.

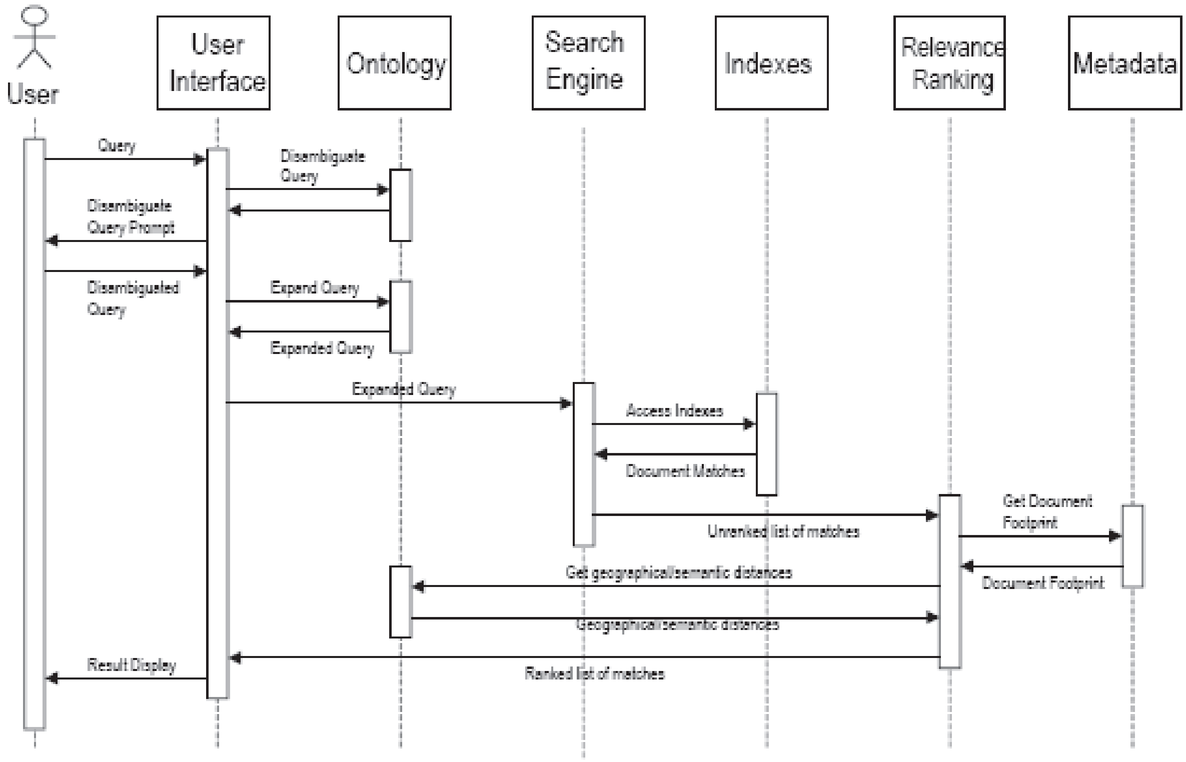


Figure 5.3: Diagram of run-time information flow (Syed *et al.*, 2003; Fig. 14)

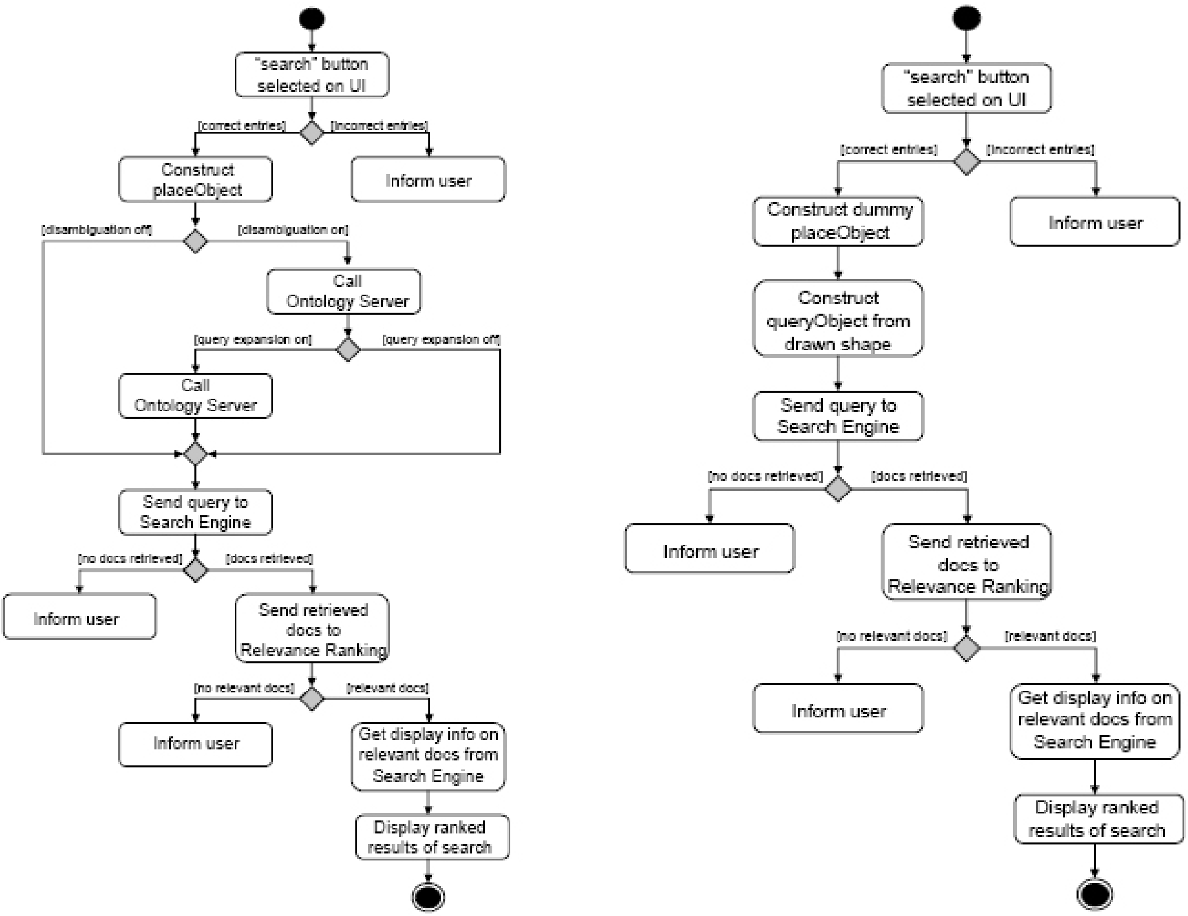


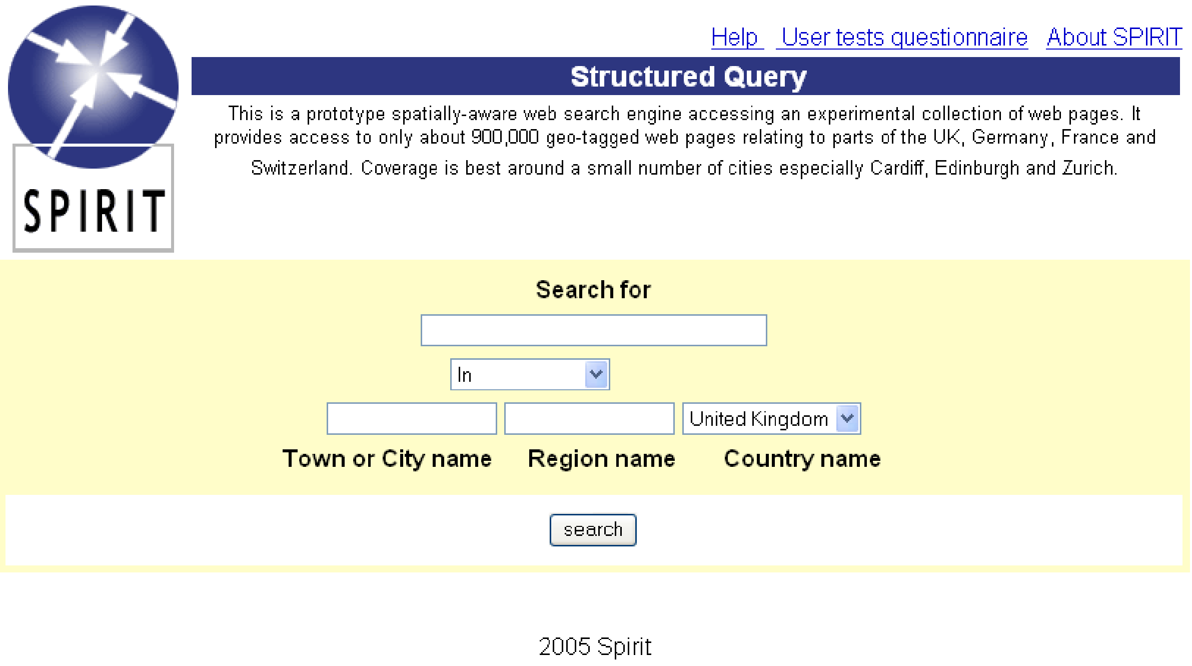
Figure 5.4: Workflow UML diagrams showing system interactions during search. a) For structured text query and b) for polygon or point based search (Purves and Yang, 2005; p. 5)

5.1.2 Run-time

The run-time operations involve a number of software components that interact with each other. Figure 5.3 illustrates a diagram of run-time information flow and Figure 5.4 illustrates a workflow UML diagram showing system interactions during search for different sub-components, namely a text based (structured) query interface and graphical query interface (point or polygon based query). Each of these components is discussed in detail in the following sections.

5.1.2.1 User interface

A multi-modal interface facilitating both textual and graphical query formulations serves to bridge information requests and the system. The SPIRIT query interface encompasses two sub-components, namely a text based query interface and a graphical query interface. Serving multiple purposes and interactions each of these interfaces enforces users to formulate their information requests as a triplet of <concept> <relationship> <location>. The text based interface (shown in Figure 5.5) is a simple extension of that of most search engines, enforcing text input in a structured form.



This is a prototype spatially-aware web search engine accessing an experimental collection of web pages. It provides access to only about 900,000 geo-tagged web pages relating to parts of the UK, Germany, France and Switzerland. Coverage is best around a small number of cities especially Cardiff, Edinburgh and Zurich.

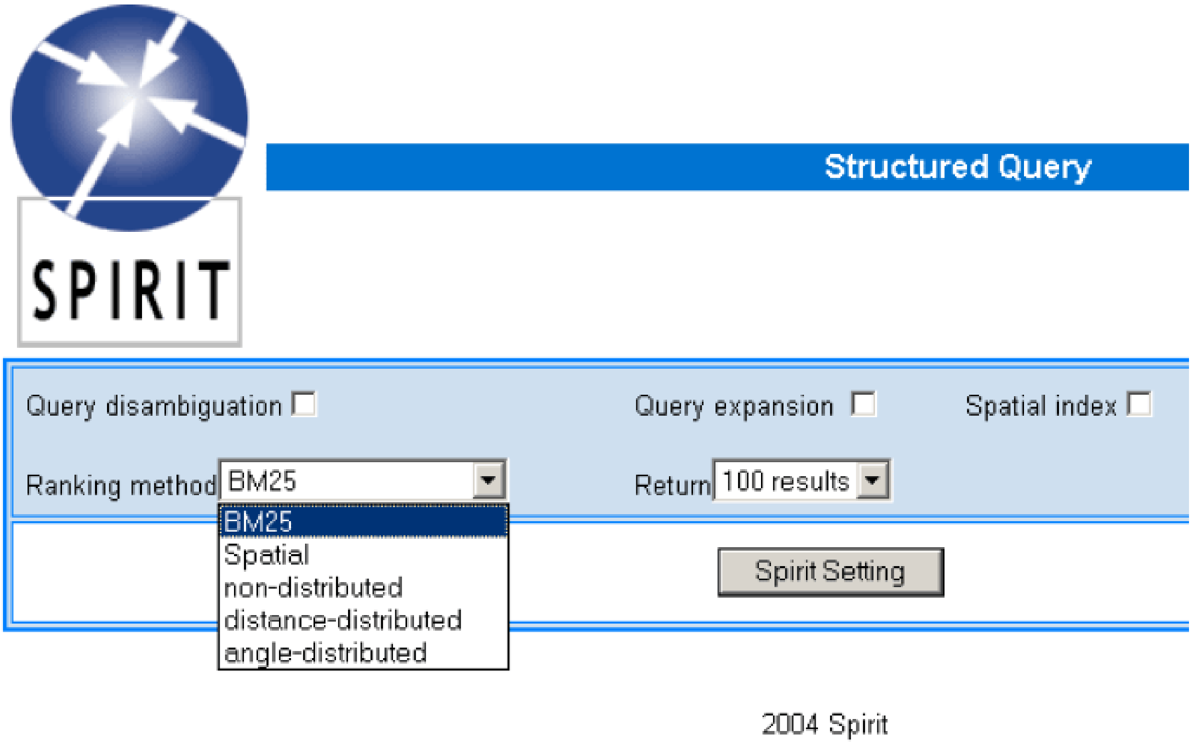
Figure 5.5: Structured query interface (also called Simple query interface) – starting point

The system can be pre-configured through a preference setup based on ranking type, query disambiguation, spatial indexing and number of results as shown in Figure 5.6. The graphical interface depicted in Figure 5.7 employs maps for users to define queries supporting basic interactive functionality, such as zooming and panning, enabling users to define query locations. The graphical interface offers a medium to define a query region without the knowledge of local place names (Syed *et al.*, 2003; Purves *et al.*, 2005; Purves *et al.*, 2007; Purves and Yang, 2006).

5.1.2.2 Query disambiguation

Many of the place names expressed in a user's query may be shared by multiple places (e.g. there are a number of places called "Newport" in the United Kingdom). A user based selective query disambiguation approach as illustrated in Figure 5.8 is adopted to pick up the appropriate footprint for later spatial processing. The geographic ontology (cf. 5.1.3) serves as a repository of place names with broader spatial contexts of the place, by utilising the containment (*part_of*) relationships encoded between places (Purves *et al.*, 2007). For example, for a query involving "Newport", a user will be prompted to select which "Newport" is intended from a menu of hierarchy information obtained from the geographic ontology:

- UK, Wales, Newport
- UK, England, Essex, Uttlesford, Newport
- UK, England, Leicestershire, Melton, Newport
- UK, England, Devon, North Devon, Newport



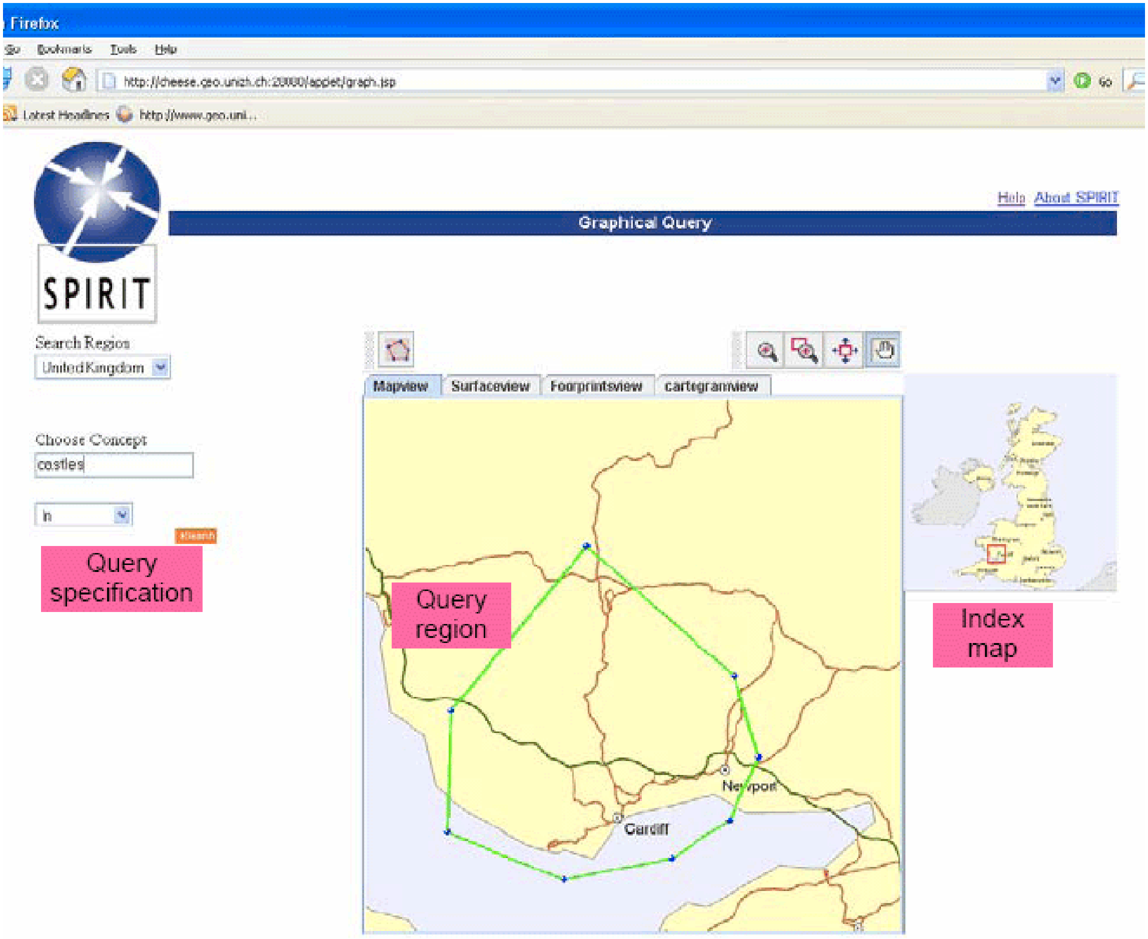
Query disambiguation ☐ Query expansion ☐ Spatial index ☐

Ranking method BM25 Return 100 results

Figure 5.6: User preference setup interface illustrating different preferences: Query disambiguation, Query expansion, Spatial indexing, Ranking method, and Number of Results to be presented.

5.1.2.3 Query Expansion

Unlike traditional query expansion where the process is about supplementing a query with additional terms to support the user's information needs, the *spatial query expansion* process is used to generate a geometric footprint. Geographic prepositions and place names that are common proponents of a user's query are used to generate this footprint. Fu *et al.* (2005) state that this approach allows us to perform more accurate spatial relevance calculations by analysing the query footprint and the document footprint. This does not happen with term-based expansion. Fuzzy terms such as "near" are difficult to interpret and can vary with respect to different user intentions as well as depending on the types of spatial and non-spatial terms used in the query. This has been resolved by adopting a query footprint generation process that accounts for the potentials of vagueness caused by geographic prepositions.



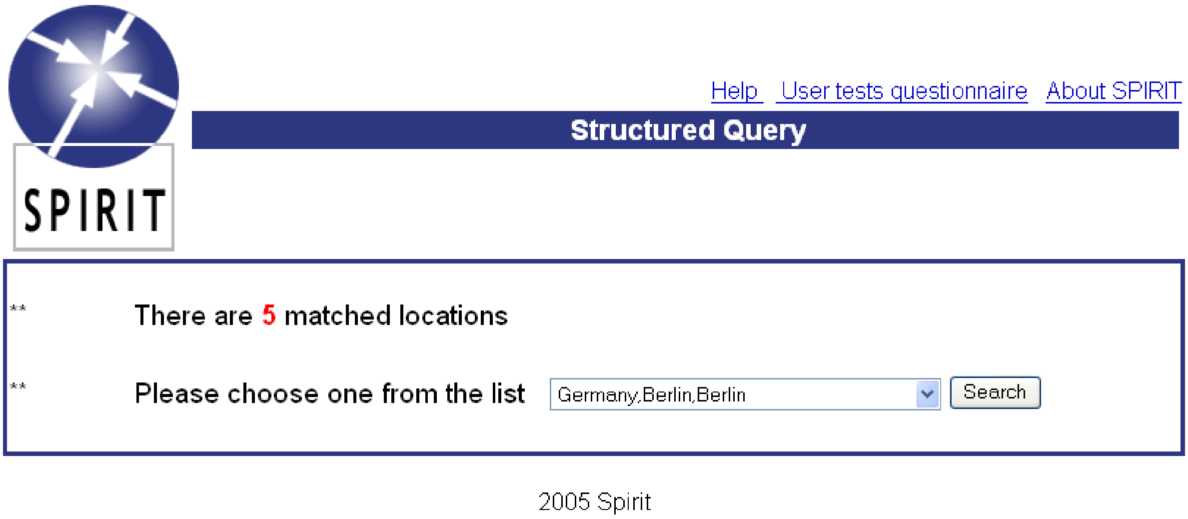
Search Region
United Kingdom

Choose Concept
castles

in

Figure 5.7: Query Specification by Polygon – In this diagram a polygon is used to specify the area in which the search will be made, for example: Looking for castles within some distance of Cardiff (Purves *et al.*, **2005).**

A range of geographic prepositions are currently supported, including *in, near, outside, north of, east of, west of*, *south of* and *distance from*. In an effort to experimentally validate the precision of the results, a set of queries was performed with spatial query expansion "ON" and "OFF". In the "ON" state, the spatial and textual index of the web collection and relevance ranking was performed using distributed ranking methods proposed by van Kreveld *et al.* (2004). When it was turned "OFF" all query terms (including spatial and non-spatial ones) were sent to the search component to perform a textual based search using the BM25 ranking method (Robertson *et al.*, 1995). The results show that the footprint based spatial query expansion approach considerably improves the search results when a query involves a fuzzy spatial relationship. Furthermore, the proposed method works efficiently using realistic ontologies in a distributed spatial search environment.



** There are 5 matched locations

** Please choose one from the list

2005 Spirit

Figure 5.8: Place name disambiguation interface

5.1.3 Geographic Ontology

The ontology component of SPIRIT is a comprehensive knowledge repository of places within the geographic coverage of the search engine. It augments geographic annotation of the existing web resources that describe or illustrate a geographic location. It models the vocabulary, spatial structure, the geographic extent, semantic and spatial relations (such as part of, containment, association, and proximity) between them, for the region covered by the search. Some of the key requirements specific to the SPIRIT ontology are listed in Table 5.3. Figure 5.9 illustrates the role of the geographic ontology (Jones *et al.*, 2003; Fu *et al.*, 2003; Jones *et al.*, 2004; Abdelmoty *et al.*, 2005; Fu *et al.*, 2005).

Table 5.3: Requirements for the SPIRIT ontology (Fu *et al.*, 2003, p.4)

1.	Recognize the presence of a place name or spatial relationship in a query or a document.
2.	Find web resources that contain alternative versions of a user specified name.
3.	Find web resources that refer to places that are inside or nearby to a specific location.
4.	Distinguish between different types of places.
5.	Perform efficient indexing of web resources to find quickly resources relating to a particular region of space.
6.	Perform relevance ranking with regard to geographic space as well as to non-geographic factors.

The geographic ontology is at the core of deriving spatial knowledge with respect to place names in the geotagging process. It maintains geometric footprints associated with a place. In its interactions with several components such as the user interface, metadata extraction, ranking component and search component these geometric footprints are put in use. The geometric footprints are in the form of *points, bounding boxes* or *polygons* serving different purposes. As explained in the previous section, these purposes include the identification of possible ambiguities of a place name to generate a query footprint that reflects the region of space to which the query refers to, as well as the expansion of a query by demonstrating possible geometric interpretations of the place name and the corresponding geographic prepositions expressed in the query.

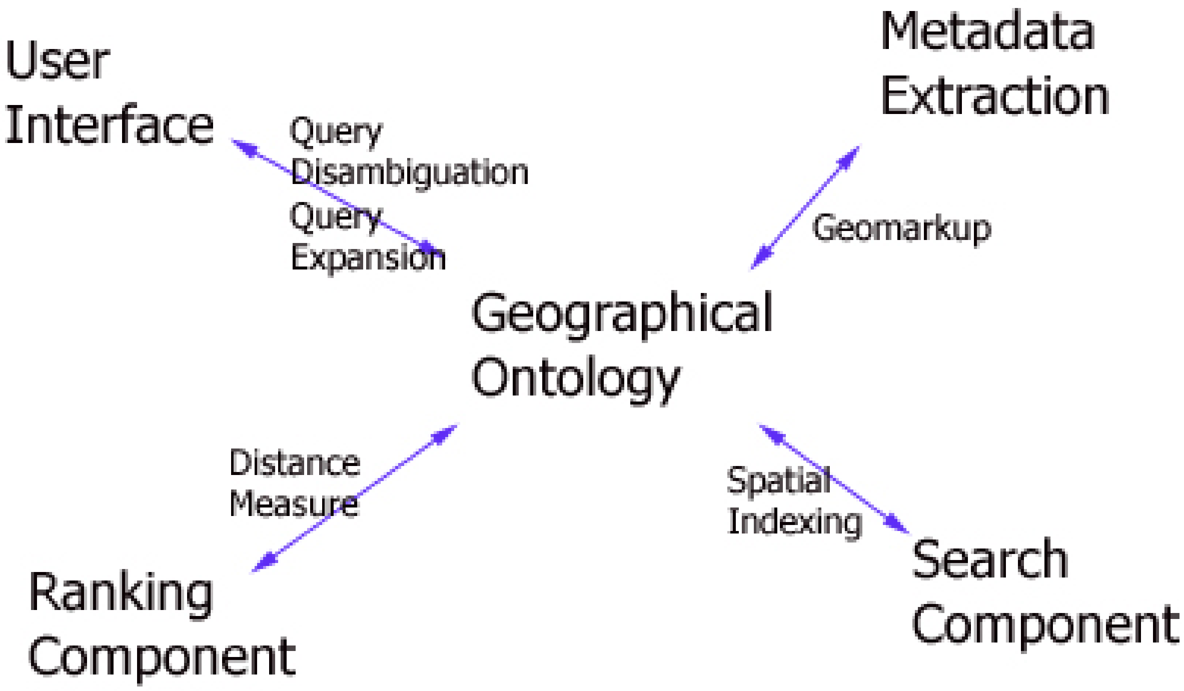


Figure 5.9: Role of the geographic ontology

Hence, the geographic ontology serves the fundamental task of interpretation and transformation of user requests. In addition it recognizes geographic terminologies and encoded vocabulary of place names applicable to a specific region. Its ability to recognise multiple geographic place name variants such as “Abertawe” and “Swansea” facilitates the intelligence to interpret user queries (Fu *et al.*, 2005). A small number of place names such as the “Central Europe”, the “British Midlands”, the “Swiss Mittelland” or the “American Midwest” are special in that they cannot be accurately assigned to a particular place. Since these place names refer to regions, they are called imprecise regions. In order to make these place names known to the geographic ontology, their boundaries first had to be delineated somehow. In SPIRIT the boundaries of these regions were derived by first mining the web for documents mentioning such regions, then geoparsing and georeferencing occurrences of other place names in such documents based on the assumption that occurrences of place names in documents are likely to be spatially autocorrelated. Such operation permits the creation of a large pool of candidate locations for imprecise regions. Finally, boundaries for the regions could be derived from the candidate locations using a variety of techniques developed by SPIRIT researchers (Arampatzis *et al.*, 2006; Purves *et al.*, 2005).

The geographic ontology is used both at the pre-processing stage and at run-time to support spatial search. During pre-processing, the geographic ontology helps primarily in geo-parsing of web documents and organization of spatial indexes (cf. Section 5.1.1). At run-time (discussed in Section 5.1.2), the geographic ontology helps disambiguate place names in the query, spatially expand the query and provide data to enable ranking of retrieved documents (described in Section 5.1.5).

5.1.4 Search Engine

Sifting through an unstructured 500,000 pages of web collection, the search engine (GLASS) facilitates the core information retrieval functionality accessing the pre-processed indexes (spatial and textual) to obtain the best match to the user’s query. The process is initiated by stemming (using Porter Stemmer; Porter, 1980) the thematic terms and stop words (e.g., “of”, “the”, “he”, “that”). Using a similar measure based on co-occurring terms, the keywords are matched with the inverted index document collection. A relevance set is constructed by querying one or more terms from the indexed document collection after computing a similarity score. The similarity measures used in the GLASS engine are based on the BM25 weighting scheme (BM stands for Best Match; Robertson *et al.*, 1998). The same weighting scheme is also found in the Okapi probabilistic IR system, which estimates term frequency as a Poisson distribution and takes into account inverse document frequency and document length. BM25 is based on three sources of weighting which have been shown to be useful for different retrieval tasks (Joho, 2004; Sparck Jones and Willet, 1997):

- **Document frequency:** terms occurring in only a few documents are likely to be more useful than terms occurring in many documents.

- **Term frequency:** the more frequent a term appears in a document the more important it is likely to be for that document.

- **Document length:** a term occurring the same number of times in a short document than in a longer one is likely to be more important in the shorter one.

The resulting set of candidate documents R are retrieved by mapping or intersection of the set of documents which lie within the query footprint and those which contain the thematic terms matched as discussed in the earlier sections. A varying set of candidate documents R may be obtained illustrating different order based on the type of indexing adopted such as T, ST or TS (cf. Section 5.1.1.2). This candidate set comprises of 1,000 documents obtained from the core search engine, which are forwarded to the relevance ranking component for re-ranking based on geographic relevance. Only key functionalities of the search engine are explained here. An in-depth study about search engines can be obtained from Joho *et al.* (2004), Sanderson & Joho (2004).

5.1.5 Ranking

Rank has evolved as a measure of relevance delivered by the system to the information request. "*A rank is a score or distance to the query and the results are sorted by this score or distance*" (van Kreveld *et al.*, 2004). A variety of ranking mechanisms varying such as PageRank (used by Google) or Best Match are adopted by commercial search engines, which are proprietary to them. While commercial search engines focus on textual relevance only, the resulting candidate documents are ranked according to both textual and spatial relevance in the SPIRIT prototype. For a query such as "*camping near Merdinger, Schwarzwald*" the documents would ideally have a score for the query term "camping" and a score for the proximity to "Merdinger, Schwarzwald", implying that the results can be mapped on to a point in a 2-dimensional space, where both axes represent a score.

From a geo-spatial perspective, each document in the web document collection is represented by a bag of footprints following the grounding of the web locations to places, and a query is also represented as a footprint. Depending on the spatial relationships (*geographic prepositions*) used in the original query, different formulae are used to calculate footprint similarity scores between query and document footprints (Table 5.4; van Kreveld *et al.*, 2005). When all footprints in a document are assigned a similarity score with respect to the query footprint, a spatial similarity score for the document can be calculated.

SPIRIT incorporates a geographically distributed ranking (van Kreveld *et al.*, 2005) mechanism which requires to define a spatial score based on the use of geographic locations associated with the web documents explicitly in two coordinates. In addition to this, there is also a term score. For the query "Castles in Scotland" documents referring to two castles located in different directions but with the same distance from the centroid of the query region in Scotland are ranked consecutively. Different ranking possibilities exist based on the combination of textual terms, spatial terms, or a textual term and metadata information. SPIRIT adopts two basic distributed ranking methods (van Kreveld *et al.*, 2004, 2005):

- *Proximity to query: points close to the Query Q are favoured.*
- *High spreading: Points farther away from already ranked points are favoured.*

Furthermore, these two basic methods are extended by using staircase enforcements and limited windows methods discussed in van Kreveld *et al.* (2004, 2005). Experimental studies have been carried out to indicate that both requirements for a good ranking, and a small distance to query as well as high spreading can be obtained simultaneously. The experiments revealed that staircase enforcement methods outperform the limited windows approach.

For a query Q_i, the following procedure is adopted to produce the relevance ranking of documents:

- For every document footprint, a footprint similarity score is produced with respect to the query footprint and connector.
- For every document, a document spatial similarity score is produced based on the footprint similarity scores of all the footprints contained in the document.
- Document spatial similarity scores are usually combined with textual BM25 scores into a document similarity score.
- Documents are ranked in descending order of their document similarity scores.

Table 5.4: Formulae used for footprint similarity score and geographic prepositions used in the query (van Kreveld *et al.*, 2004)

1.	*Inside*	Binary operator defined between a query's bounding-box and a document's footprint (MBR or centroid). Coordinates are checked for containment.
2.	Near	$Near(a,b) = exp(-L * D(a, b))$, where a and b are the centroids of a query's and a document's footprint, $D(a, b)$ is their Euclidean distance. Thus, proximity scores decrease exponentially from 1 to 0 with increasing distance. L controls the rate of decrease, or, in real-world terms, "how far is far", and can for example be a function of the query footprint; thus for example, near things can be further from large objects.
3.	*North-of, south-of, east-of, west-of*	Assuming that *a* and *b* are the centroids of a document footprint and a query footprint respectively, and that *psi* is the angle of the vector *ba* from the positive x-axis with the origin assumed on point *b*. For *north-of*, if *psi* >= 180 or *psi* <= 0 then *north-of* (a, b) = 0, otherwise *north-of* (a, b) = 1 - ¦90 - psi¦ / 90. The other directional operators are calculated in a similar manner. Proximity is also taken into account, and to obtain the final score they are multiplied with *near*(a, b).

Table 5.4 details the formulae used to calculate the footprint similarity scores depending on the geographic preposition used in the original query. Prior to the combination of textual and spatial relevance scores to generate a single ranking, they are normalized into the range [0,1]. BM25 scores can be quite unpredictable in their range and they are currently normalized linearly by dividing with

highest document score for the query. Figure 5.10 illustrates a plot of both spatial and BM25 scores for 8 documents. Table 5.5 illustrates some of the methods adopted to combine both spatial and BM25 score into one. A detailed study of the algorithms and ranking mechanisms are reported in van Kreveld *et al.* (2005).

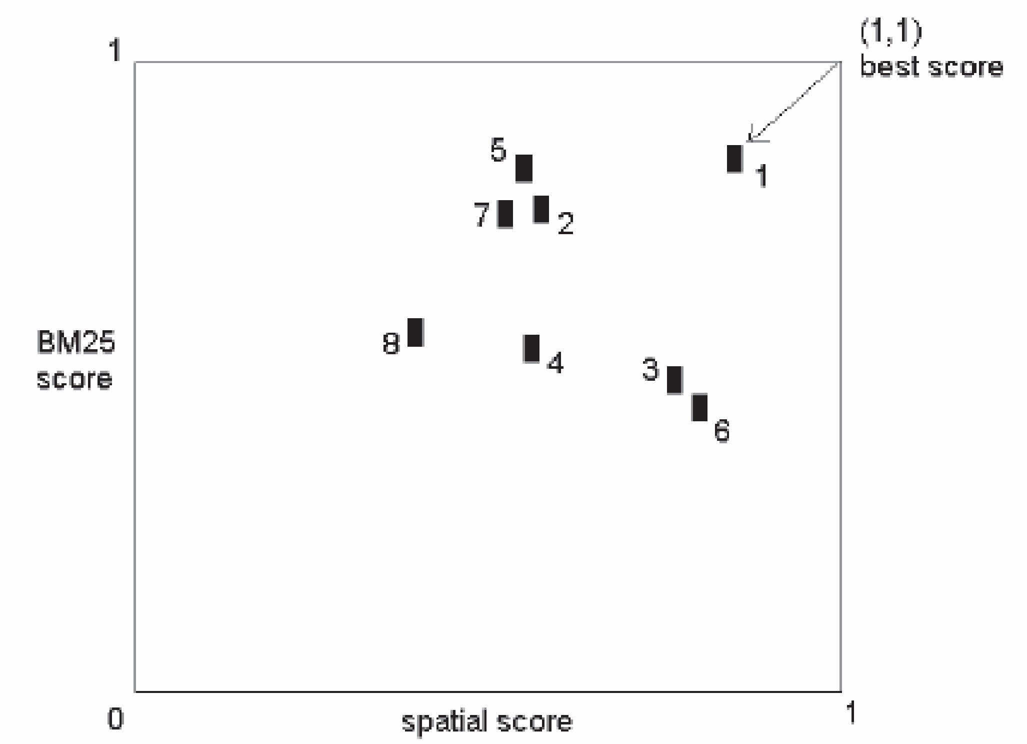


Figure 5.10: Plot of BM25 (textual) against spatial document scores and associated ranking for a distributed method (van Kreveld *et al.*, 2004, Arampatzis *et al.*, 2006)

Table 5.5: Methods adopted to combine spatial and BM25 scores (van Kreveld *et al.*, 2005)

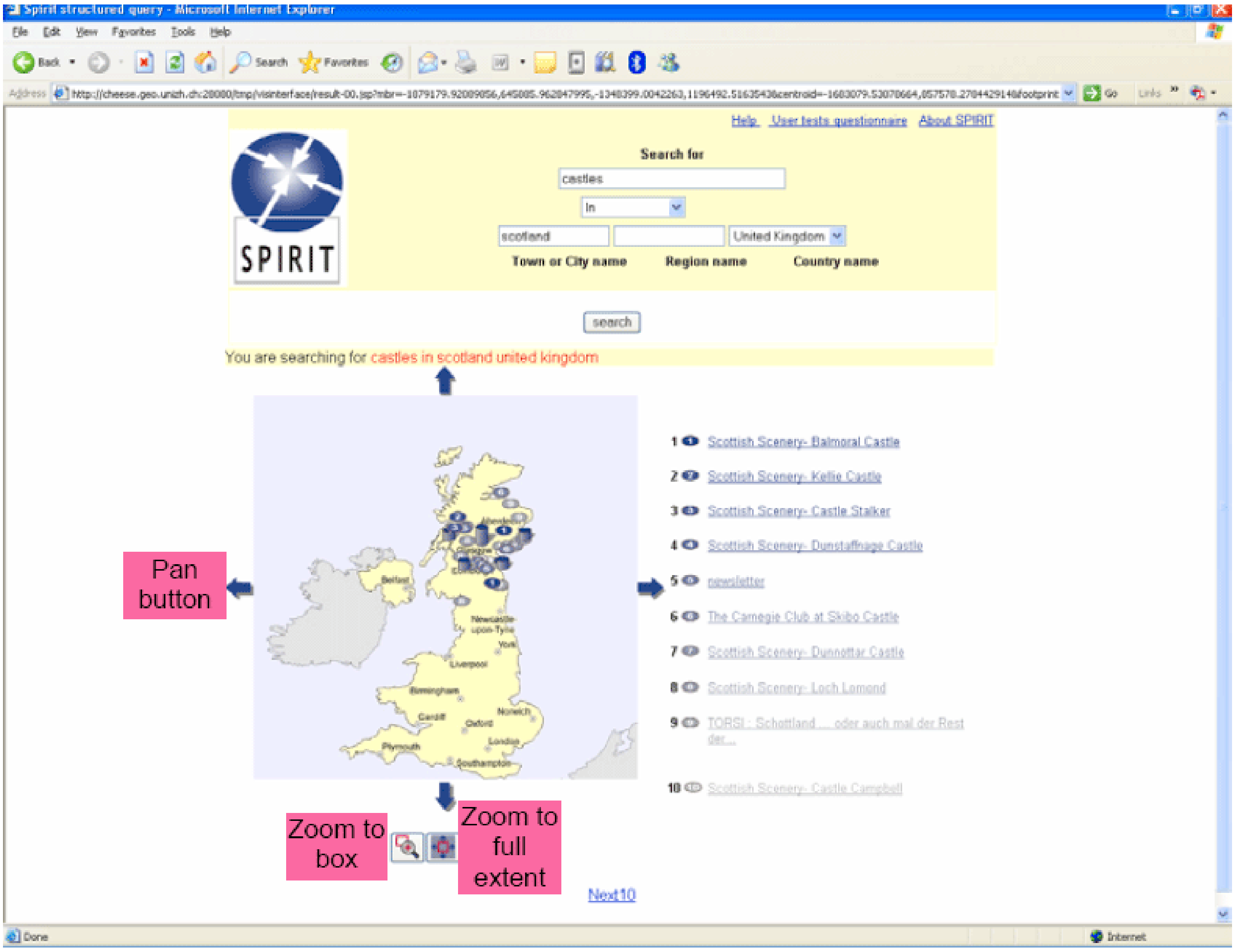
1.	Non-distributed Method	This method ranks the documents in ascending order of their Euclidean distance from point (1,1) that is assumed to be the most relevant possible document.
2.	Distributed Method	This method tries to de-cluster documents that have almost the same score components. Figure 5.10 demonstrates the distributed method.
3.	Angle-based Distributed Method	This method tries to de-cluster documents that have almost the same score components on the angle of a yet unranked point to the ones already ranked.
4.	Distance-based Distributed Method	This method tries to de-cluster documents that have almost the same score components on the distance from the already ranked documents.

5.1.6 Metadata

The metadata component interacts with the geographic ontology to recognise the presence of place names in a document for geotagging of the web documents. The metadata component attempts to associate the documents of the web data collection with one or more document footprints representing the regions of space to which individual documents relate. In SPIRIT the resulting geotagged collection consists of about 900,000 documents that refer to parts of the UK, France, Germany and Switzerland. This is used in generating the spatial index and in spatially aware relevance ranking (the ranking component) (Purves *et al.*, 2007).

5.1.7 Results Display

An elemental approach to the representation of retrieved information is map based representation of the document footprints and corresponding textual lists ordered by the combined ranking. By employing key cartographic visualization techniques such as brushing and linking, correlation between the ranked snippets (textual listing of the URLs) and the document footprints is achieved. Figure 5.11 presents a snapshot of the simple map based representation of the results.



Search for

castles

In

scotland | | United Kingdom

Town or City name | Region name | Country name

You are searching for castles in scotland united kingdom

1 Scottish Scenery- Balmoral Castle
2 Scottish Scenery- Kellie Castle
3 Scottish Scenery- Castle Stalker
4 Scottish Scenery- Dunstaffnage Castle
5 newsletter
6 The Carnegie Club at Skibo Castle
7 Scottish Scenery- Dunnottar Castle
8 Scottish Scenery- Loch Lomond
9 TORSI : Schottland .... oder auch mal der Rest der...
10 Scottish Scenery- Castle Campbell

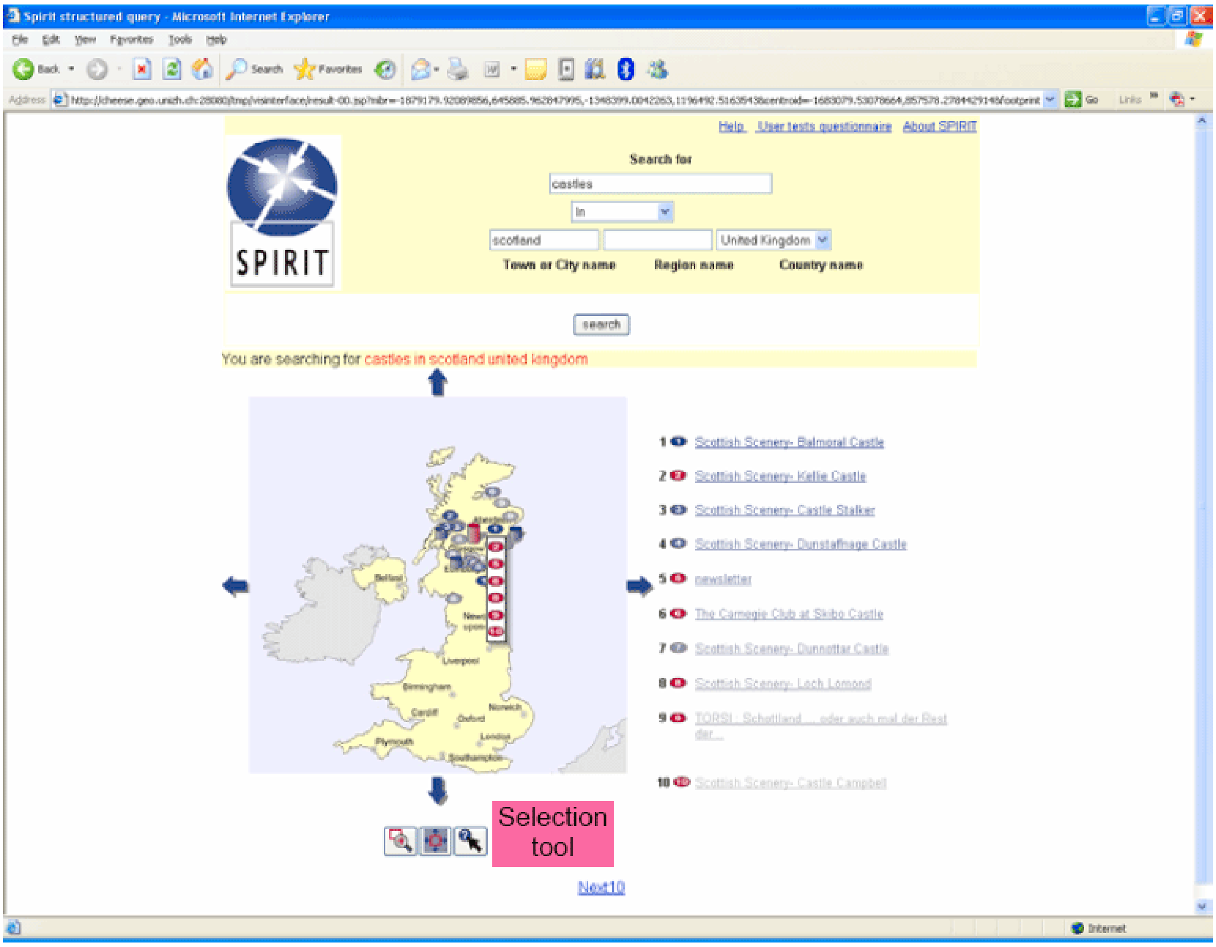
Next10

Figure 5.11: Results presentation of the structured query interface. Each of the elements of this interface are described in pink boxes (Purves *et al.*, 2005).

Brushing and Linking (as shown in Figure 5.12 and Figure 5.13) were implemented through Java Script with mouse-over events on the map or the list of returned documents resulting in the corresponding link being highlighted by colour change to red. Clicking on the list, or on a map symbol (stack) when the select tool is highlighted (as opposed to zoom) results in opening the cached version of the web pages (as stored in the SPIRIT collection) to be displayed.

It is non-trivial to display multiple query results, sharing similar footprints at the same location. Two different approaches are explored, with simple map based representation adopting stacks or cylinders, whose height is indicative of the number of documents in them and all subsequent documents highlighted in the ranked list. Figures 5.12 and 5.13 illustrate an example of simple and graphical map based displays. The ranked list is colour coded to indicate the relevance of individual document footprints. After a number of iterations, this colour scheme was chosen to minimise conflict with elements displayed by the map and to maximise readability.

Unlike the simple map based representation, the graphical map based representation employs a different approach. The results sharing similar footprints are represented as a 'spider' graph of connected circles shown in red in Figure 5.13. Possible approaches were investigated to better reflect the positioning of the footprints within a document based on directional aspects, rather than representing the centroid of the footprints of the documents as a single orb.



Search for: castles

In

Town or City name: scotland

Region name:

Country name: United Kingdom

You are searching for castles in scotland united kingdom

1 Scottish Scenery- Balmoral Castle

2 Scottish Scenery- Kellie Castle

3 Scottish Scenery- Castle Stalker

4 Scottish Scenery- Dunstaffnage Castle

5 newsletter

6 The Carnegie Club at Skibo Castle

7 Scottish Scenery- Dunnottar Castle

8 Scottish Scenery- Loch Lomond

9 TORSI : Schottland ... oder auch mal der Rest der...

10 Scottish Scenery- Castle Campbell

Figure 5.12: Brushing and linking results for multiple documents with identical footprints (Purves *et al.*, 2005).

To serve the geographic metaphor of the users, maps have been exploited as a medium to transmit the information emphasizing its spatial aspects. However, without adequate background context, they may be of little value of the map representations to the users. In choosing background maps (i.e. base maps or reference maps) there is the requirement of appropriate granularity relative to the results, that is, the detail presented on the background map should correspond to the detail with which an object (i.e. a search result) is located on the map. Furthermore, consistency in the symbolization used in mapping should be such that the foreground (in this case the search results) are not obscured.

This consistency is achieved by mapping the granularity of indexes to that of the background map context, which is relatively coarse, meaning that in most locations documents are not positioned with coordinates that are meaningful beyond the town level. The data for the four test countries France, Germany, Switzerland and the United Kingdom have been indexed. Based on the earlier analysis, there is considerable mismatch when compared with the provision of higher resolution data sets. For each of these countries four levels of details (scales) of background mapping are developed, with gradually increasing amount of detail (as shown in Figure 5.14). The national coverage data was built by using Digital Chart of the World. To ensure consistency among the data sets and representations, they have been transformed to the same, common (UTM) projection.

Table 5.6: Scales for mapping in different countries (nominal resolution 72 dpi) (Purves and Yang, 2005)

S.No	United Kingdom	France	Germany	Switzerland
1.	1:7,261,309	1:7,155,012	1:5,995,273	1:2,725,523
2.	1:3,758,825	1:3,546,627	1:3,103,460	1:1,410,856
3.	1:1,517,426	1:1,431,713	1:626,163	1:569,539
4.	1:455,118	1:429,428	1:375,767	

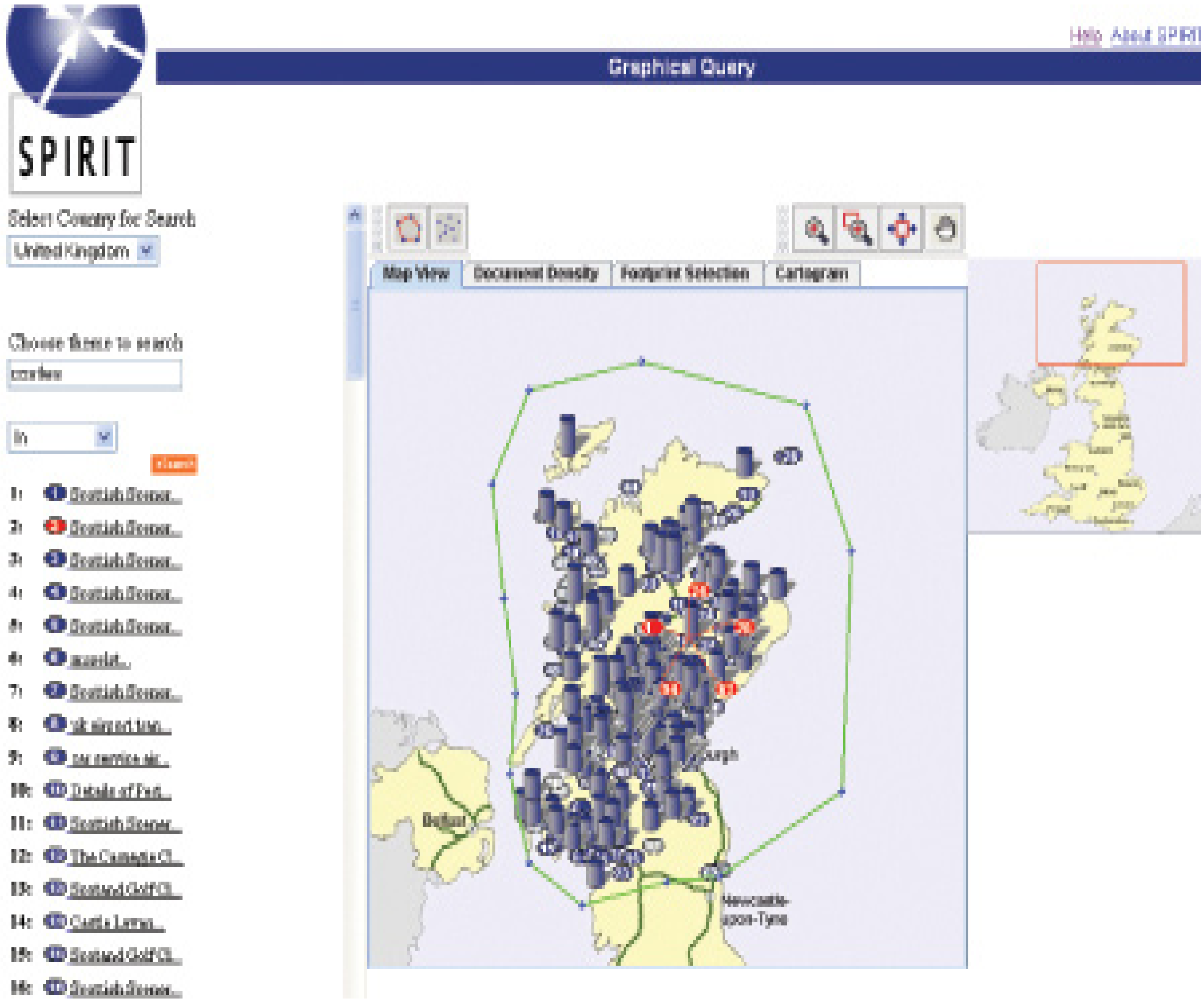


Figure 5.13: The graphical interface for the presentation of results for a query such as "Castles in Scotland". The region of Scotland was selected by the user by defining the search polygon. Highlighting of document footprints that share the same location is achieved by drawing a 'spider' graph of circles (in red).

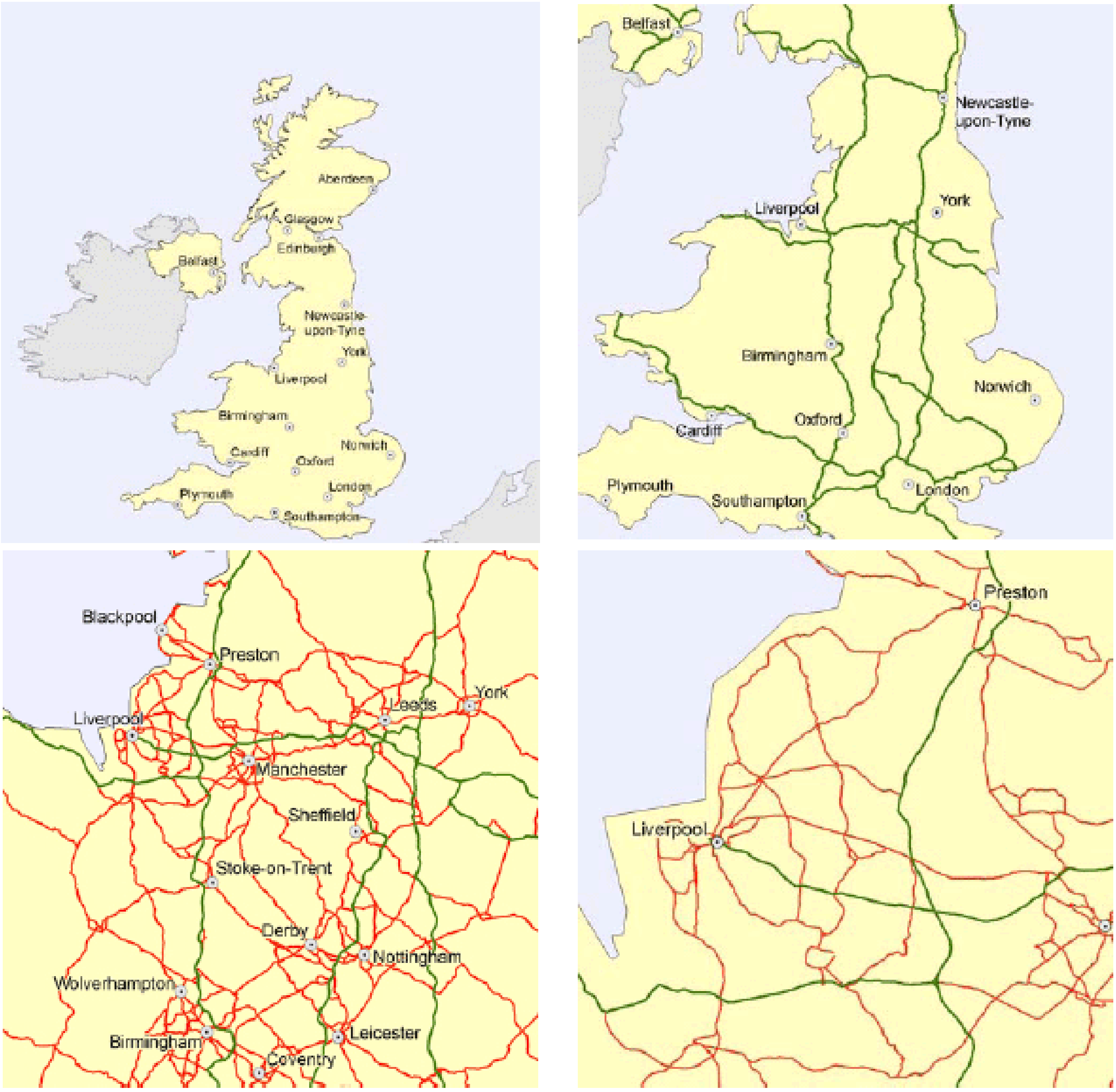


Figure 5.14: Background mapping for the United Kingdom (Purves *et al.*, 2005).

At the time when the SPIRIT prototype was developed, the typical minimum screen size that could be assumed was 800 x 600 pixels. Hence, in all cases the maps were designed to be displayed in a 400 x 400 pixel window, leaving room for the remaining parts of the user interface. Table 5.6 details the scale adopted for all four countries with a nominal resolution of 72dpi. Increasing scales were then established by an approximate doubling of image size in pixels. For two selected test regions in Edinburgh, Scotland and Zurich, Switzerland data with nominally higher granularity were provided using Tele Atlas data to achieve a detailed backdrop (as shown in Figure 5.15). Navigation through these levels has been facilitated through panning and zooming functionality. Simple tools for panning are provided through screen arrows and zooming was achieved by dragging a rectangular box. To switch back to the initial state, a reset button is provided to move to the original (full) extent.

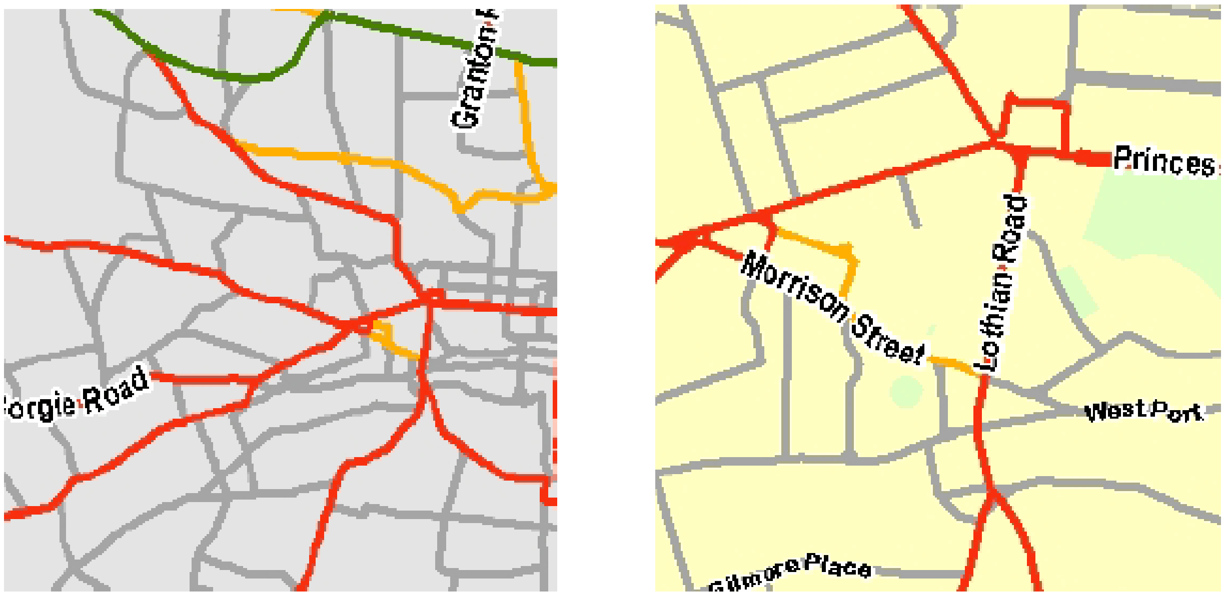


Figure 5.15: Tele Atlas Data set illustrating the map background for Edinburgh at the highest possible resolution (Purves *et al.*, 2005).

5.1.8 Broker Component

Each of the software components describd above exists at remote sites (at the universities of Cardiff, Sheffield, Utrecht and Zurich). They communicate via a functional interface called the broker component, which acts as a session manager, controlling and scheduling the information flow through the system as well as enabling the recording of all steps involved in processing a query for monitoring and evaluation purposes. The distributed architecture allowed each partner site to develop and maintain each of these components individually. For instance, the geographical ontology was developed in Cardiff, the search engine in Sheffield, the relevance ranking component in Utrecht, and the user interface and results presentation in Zurich. Communication between these components was achieved by using SOAP (Simple Object Access Protocol), a light weight communication protocol that facilitates interoperability using XML as the data serialization format (Purves *et al.*, 2007).

5.2 Summary

The SPIRIT prototype system presented in this chapter has addressed the need for tools to support the spatially aware information seeking process. Alternative interfaces such as the structured query interface and the cartographic interface serve to transmit the information request to the system. Of these two interfaces, the former presupposes that users have knowledge of appropriate place names with which to formulate their query, and the latter provides assistance to inexperienced users who have little knowledge of place names.

Map backdrops serve to address a number of tasks in a cartographic interface such as query formulation (selecting the regions of interest in the cartographic interface), query reformulation (identifying relevant documents to the user and re-submitting a query) and query visualization (displaying search results to the user).

The simplest possible representation of these results obtained from the SPIRIT engine is by displaying them against a map backdrop. Each of these documents are represented by their footprint's centroid, though they may have widely varying spatial extents (e.g. *the footprint of Switzerland vs. the footprint of the City of Zurich*). A document may have many geographic footprints, of which usually only a small number are in fact relevant to a query. For a query such as "*McDonald's in Zurich, Switzerland*" emphasizing on the locational element of the query, Zurich, the spatial extent of the query is limited to the region of Zurich. But, these documents may also refer to place names outside of Zurich (Switzerland) or Europe, occurring in the same web document. The relevance ranking component of the SPIRIT prototype gives individually identified footprints a relevance score before combining the thematic and spatial relevance score of the entire document to give an overall relevance score to the document used to assign its position in the ranked list (van Kreveld *et al.*, 2005). Only those footprints that are considered by the ranking component to be relevant to the query are then displayed on the map. This constitutes a major difference between a spatially aware search engine, where location is explicitly used in ranking. Documents searched are unstructured text, for a search engine where locations are geo-referenced but searched textually.

The limitation of the map based visualization is its inability to accommodate as many documents as possible in the ranked list. Although a relevant document may well be contained in the top ten, the ability of a map to convey information about many more footprints, and thus many more documents, to a user should be exploited. Since most documents have identical footprints zooming to any level of detail will not allow them to be visualized independently and instead we adopt two techniques to convey their locations to the user. Brushing and Highlighting are used to link result data sets (URLs) with corresponding footprints. However, footprints still obscure each other and must be schematically illustrated (e.g. through a different symbol) to better convey information to the user. Also, one cannot rule out the possibility that the footprints of different documents relating to different concepts overlap. For instance, points of interest at a certain geographical location might refer to concepts such as tourist attractions, museums, restaurants, castles etc. The weaknesses of these basic map based methods of results display suggest that other approaches should be investigated and used. The next chapter will therefore introduce a variety of more advanced visualization methods.

Chapter 6

Alternative Visualizations

"Visualization and belief in a pattern of reality, Activates the creative power of Realization"

- A.L.Linall, Jr.U.S.Editor

One of the primary objectives of this research is to deliver a variety of functional views based on the user's need. The most elementary graphical display of results is in the form of a textual listing, supported by relevant footprints from the document on a map (§ref. Figure 5.11, 5.12 & 5.13). Documents are simply represented by their footprint's (geographic extents) centroid, though they may have widely varying spatial extents (e.g. the footprint of Switzerland vs the footprint of Zürich). In this chapter a range of other implemented visualisations of the geographic information retrieved from the internet are presented, all of which aim to address user's differing geographic metaphors (Gould and McGranaghan, 1990; Kuhn, 1995; Rauterberg and Hof, 1995) and thus address the issues of spatially-aware information seeking presented in earlier chapters.

Footprints have evolved as a medium for cataloguing the geographic extent of information available on the web relating to a specific location. Given the cognitive and experiential aspects involved when interacting with the various representations during the spatially aware information retrieval process (Spink, 2002), the need for tools to support the process has been widely acknowledged in geovisualization with, for example, Kraak (2001, p.398) pointing out that such tools can be used "*... to explore data, and through that exploration to generate hypotheses, develop problem solutions and construct knowledge*". A range of possible representations and methods to visualize the results were presented in Chapter 4. Numerous factors (Table 4.6) have shaped the query and display methods that were implemented in this work.

While the emphasis is on spatialization of geographic information retrieved from the internet, the elementary display of the results is in the form of a textual listing, supported by relevant footprints from the document on a map, which is referred to as map based representation presented in chapter 5. Documents are simply represented by their footprint's (i.e. geographic extent's) centroid, though they may have widely varying spatial extents (e.g. the footprint of Switzerland vs the footprint of Zurich). Footprints have evolved as a medium of cataloguing the geographic extent of the information available on the web relating to a specific location. With map based representation as an elementary form of representing these footprints, documents which share similar footprints are represented as a stack and regions which share multiple document footprints form clusters. Furthermore, the map based representation limits the ability to exhibit the underlying relationship between documents and their footprints. These characteristics of the result sets outlined in Table 4.6 streamlined the way results are visualized, to investigate alternative visual representations that help users with the tools to explore differing aspects of results set. The alternative representations presented in this chapter engages in investigating various representation that serve varying functionality. This chapter sets out to discuss various visualization techniques adopted for analysis and presentation of the data and is structured around these techniques as follows:

- Density surface representations
- Footprint based filtering
- Cartogram
- Spatial activity map (mock-up constructed with real datasets)
- Isoline representation (mockup constructed with real datasets)

6.1 Density surface representations

The spatial pattern of geographic extents (footprints) both within a document and the relations between documents makes it non-trivial to derive boundaries representing the nature of a results set. Thus, a surface is created representing imprecise regions that somehow represent the extent of the results retrieved with respect to a given query. The following special cases may occur describing the characteristics of the result set (geographic extents).

- **Equal probability:** any point has equal probability of being in any position or equivalently each small sub-area of the map has an equal chance of receiving a point (O'Sullivan and Unwin, 2003).

- **Independence:** the positioning of any point is independent of the positioning of other points (O'Sullivan and Unwin, 2003).

By collapsing the geographic extent of a footprint onto its centroid, rather than its geographic extent, we obtain a point location for each document footprint from which a density surface can be calculated.

6.1.1 Method

The point patterns obtained from the collapse of geographic extent can be effectively qualified and quantified in terms of their orientation, density, relative distance, ranking and degree of spatial randomness. Point pattern events under complete spatial randomness should display no signs of either first-order or second-order variations (O' Sullivan and Unwin, 2003).

O'Sullivan and Unwin (2003, p.79) proposed two interrelated approaches, "*based on point density and point separation: which are related to distinct aspects of spatial pattern of the first and second-order effects. Recall that first order effects are marked, absolute location, which is an important determinant of observation, and in a point pattern clear variations across space in the number of events per unit area are observed. When second-order effects are strong, there is interaction between locations, depending on the distance between them, and the relative location, is important. In point patterns such effects are manifested as reduced or increased distances between neighbouring or nearby events*".

Table 6.1: First and second order effects (O' Sullivan and Unwin, 2003)

	First order	Second order
1.	The assumption of equal probability cannot be satisfied.	The assumption of independence cannot be satisfied.

Table 6.1: First and second order effects (O' Sullivan and Unwin, 2003)

	First order	Second order
2.	When the first order effects are marked, absolute location is an important determinant of observations, and in a point pattern clear variations across space in the number of events per unit area are observed.	When second-order effects are strong, there is interaction between locations, depending on the distance between them, and relative location is important.

Kernel density estimation is a statistical method of creating a density surface model through the summation of weighted windows associated with each data point. The pattern has a density at any location in the study region, not just at locations where there is a data point (O'Sullivan and Unwin, 2003). Table 6.1 details the difference in first and second order effects which may influence the resulting density surfaces. The coverage of the documents obtained from the result set over a wide area means that the system is well suited to exploring document densities over space, since over large geographic extents the search results are likely to accurately reflect the distribution of documents in space. The resulting density based representation shows "*hot spots*" of incidence and by overlaying the surface on the document footprints may provide useful contextual information.

An intensity estimate at a point p can be expressed as

$$\hat{\lambda}_{\mathbf{p}} = \frac{\text{no. } [S \in C(\mathbf{p}, r)]}{\pi r^2}$$

where C(**p**, **r**) is a circle of radius **r** centered at the location of interest **p**.

O'Sullivan and Unwin (2003, p.87) state that "*many sophisticated variations weight nearby events more heavily than distant ones in estimating the local density. They are based on distance of the point to be estimated from events in the point pattern. These are always specified with a parameter that is equivalent to the simple bandwidth*".

Figure 6.1 shows the density surface visualization for castles in a user specified region (Scotland) by drawing a query polygon. The black dots indicate the positions of the documents. Table 6.2 details the justification for the use of density estimation for SPIRIT result sets. However, this visualization tool only supports a static representation of search results. A possible application of this tool is in the tourism domain, to explore why a single peak was so prominent and whether regions with less document density could be better represented on the web or the topographic conditions of the region are well described.

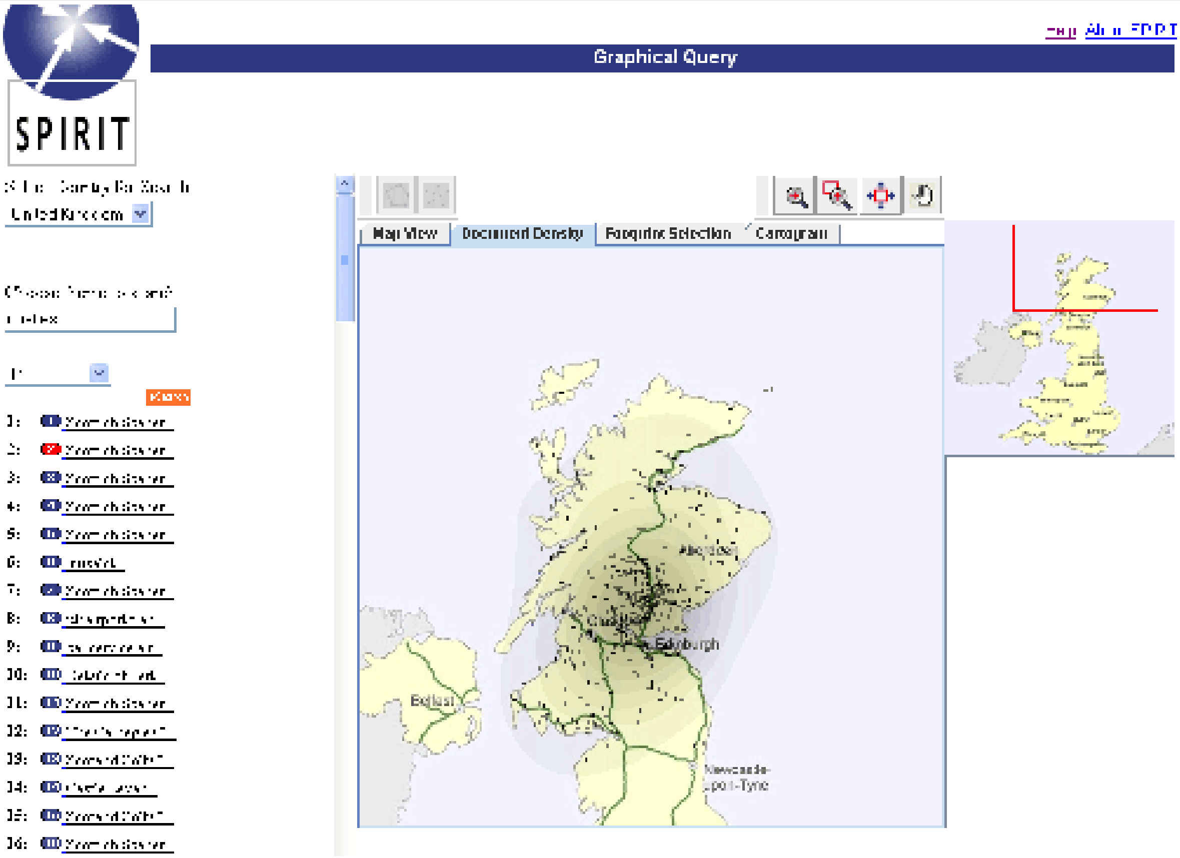


Figure 6.1: An example of a density surface representation for "castles" in a user specified region of Scotland, United Kingdom.

Table 6.2: Justification for the use of density estimation for collapsed geographic extents (footprint centroids) point data sets obtained from the SPIRIT engine.

	Pros	Cons
1.	The high density areas in data can be readily seen. Thus, allowing ease of exploration of document distribution across space	Requires parameters such as frequency and bandwidth to be set correctly.
2.	Summarizes large amounts of data.	Difficult to compare analyses from different times and regions.
3.	Gives quantified density measurements for all areas in the region in a single graphic view.	Doesn't use administrative boundaries, making it difficult to respond to these high density areas.
4.	Can infer a density volume at any location of the study area, not just a footprint centroid.	The density surface is calculated over a map view area rather than the political or administrative boundaries.
5.	Can create a surface representing imprecise regions and derive boundaries of imprecise regions, where a query is made representing such a region.	The density surface does not take into account the combined relevance of document footprints.

6.2 Footprint based filtering

A prominent characteristic of the results obtained from the SPIRIT engine is the geographic extent and shared position of the footprints. A hypothesis is built on these characteristics that the documents with a few or single scopes with narrow geographic extents may describe a single location and be thematically well defined, while documents with multiple and large geographic scopes may typically describe an activity or activities that might be carried out in a region or a journey through a region, facilitating descriptive information in general. Furthermore, locations with fewer documents referring to them are more unique, and thus may be of more interest.

6.2.1 Method

Provision to explore the properties of geographic extents of the footprints of individual documents and the relationships between documents is facilitated by interactive filtering, selection and highlighting. Unlike the density surface representation the geographic extent of the footprint is retained as bounding box rather than collapsing it into a point based on the centroid, to enrich the user's understanding of the geographic scopes of each of the footprints and the documents as a whole. Figures 6.2 and 6.3 demonstrate the interactive filtering technique, wherein users are equipped with tools to sift through documents through the size of the geographic extent of the footprints and the number of footprints per document. In addition to this, basic map functionalities such as zooming and panning are also supported.

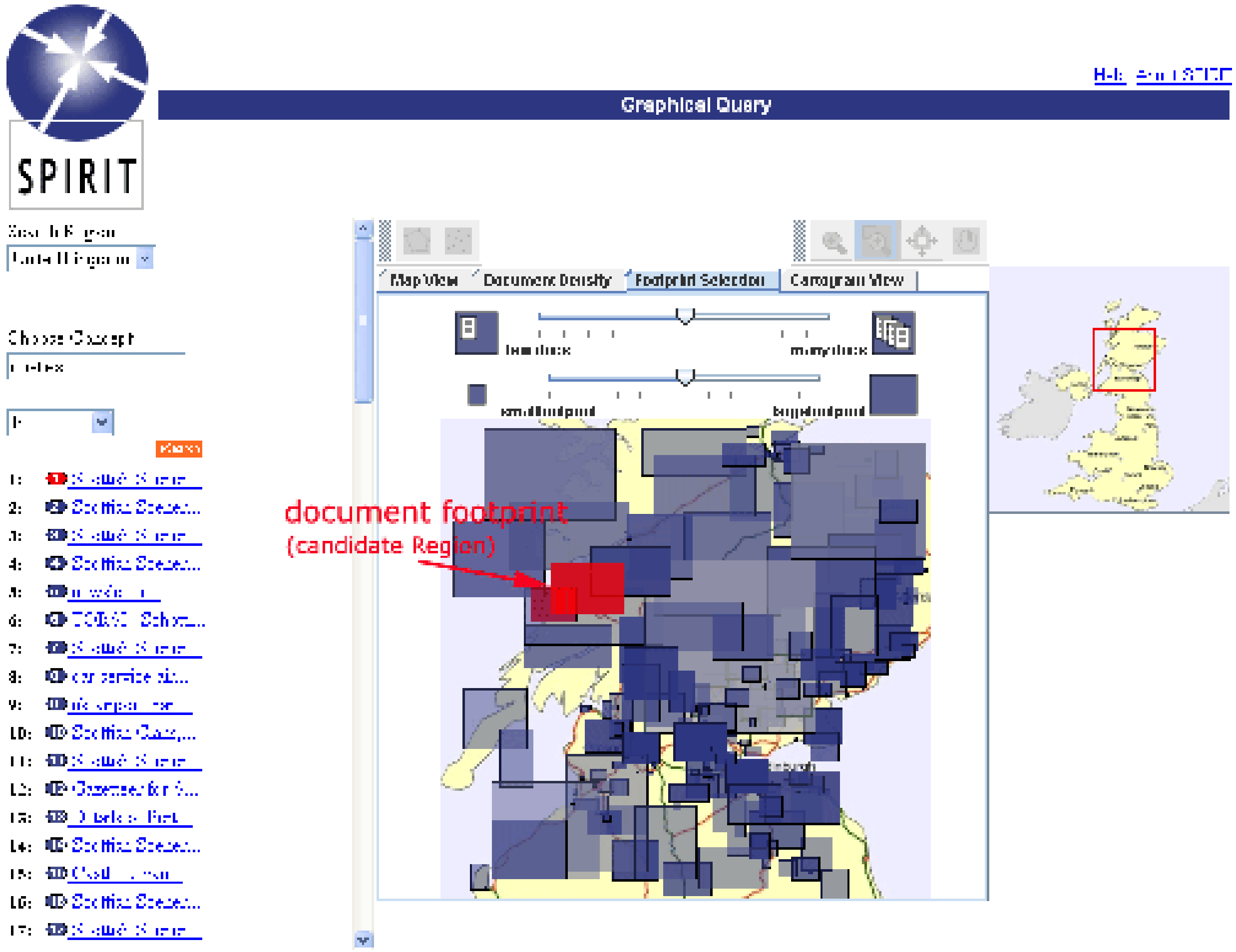


Figure 6.2: An illustration of footprint based filtering technique for a query "castles in Scotland", highlighting footprints within a document illustrated as document footprint (candidate region).

Varied perspectives are possible through the combination of the filtering mechanism provided by the two sliding bars to explore the search results, where documents and footprints are filtered based on the scale of refinement. The tool can thus be used to analyse web document coverage and nature as suggested by the density tool and also by users for identifying particular types of documents.

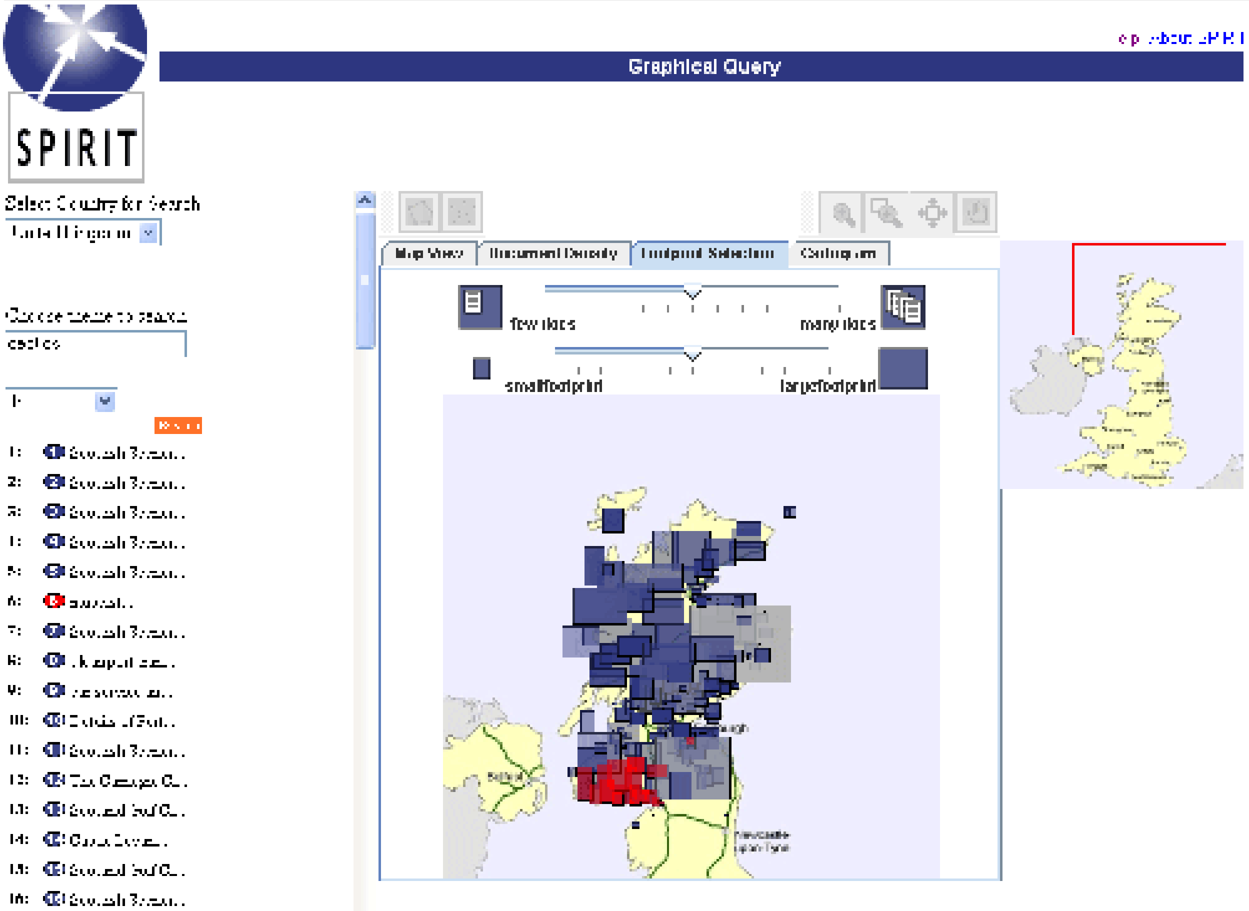


Figure 6.3: An illustration of footprint based filtering technique demonstrating interactive filtering for a user's selection of a few documents for a specific footprint (geographic extent); for the query "castles in Scotland".

Table 6.3 discusses justifications for the use of footprint based representations demonstrating the spatial extent of the footprint and its representation.

Table 6.3: Justification for the use of footprint based representation demonstrating the spatial extent of the footprints

	Pros	Cons
1.	Overviews of document and footprint coverage/spatial extent	Complexity in understanding and readability due to overlapping spatial extents with in a document.
2.	Allows the filtering of documents by extent and number of footprints	Represent documents whose true footprint is either a point or a polygon as a rectangle, which implies a high degree of generalization.

6.3 Cartograms

A density equalization approach widely known as cartograms or density-equalizing maps or value-by area maps, where each geographical unit's size is proportional to some thematic count or weight with respect to the area (Dorling, 1996; Dent, 1998). Subsequent usage of effective legend design and clear demarcation of each geographical unit facilitates visual distinction to communicate to the users the shapes of enumeration units and the related values of attributes. Some of the key constraints in the effectiveness of cartograms include preservation of geographic relationships and data that has sufficient variation to convey difference in the visual representation.

6.3.1 Method

Dorling's (1993) algorithm has been used to create a non-contiguous cartogram to represent the number of documents in a geographic unit which is directly proportional to the size of circles (Fig 6.4). The program runs an iterative procedure which moves overlapping symbols apart, while drawing neighbouring cases together. In doing so it attempts to retain the contiguity structure of the geographic space.

Brushing and linking facilitates exploration of the relationships between document counts and some pre-defined tessellation of space - in this case based on administrative units, retrieved from a Web Feature Server (WFS) on the fly, thus facilitating identification of regions and thus refinement of a search specific to a region.

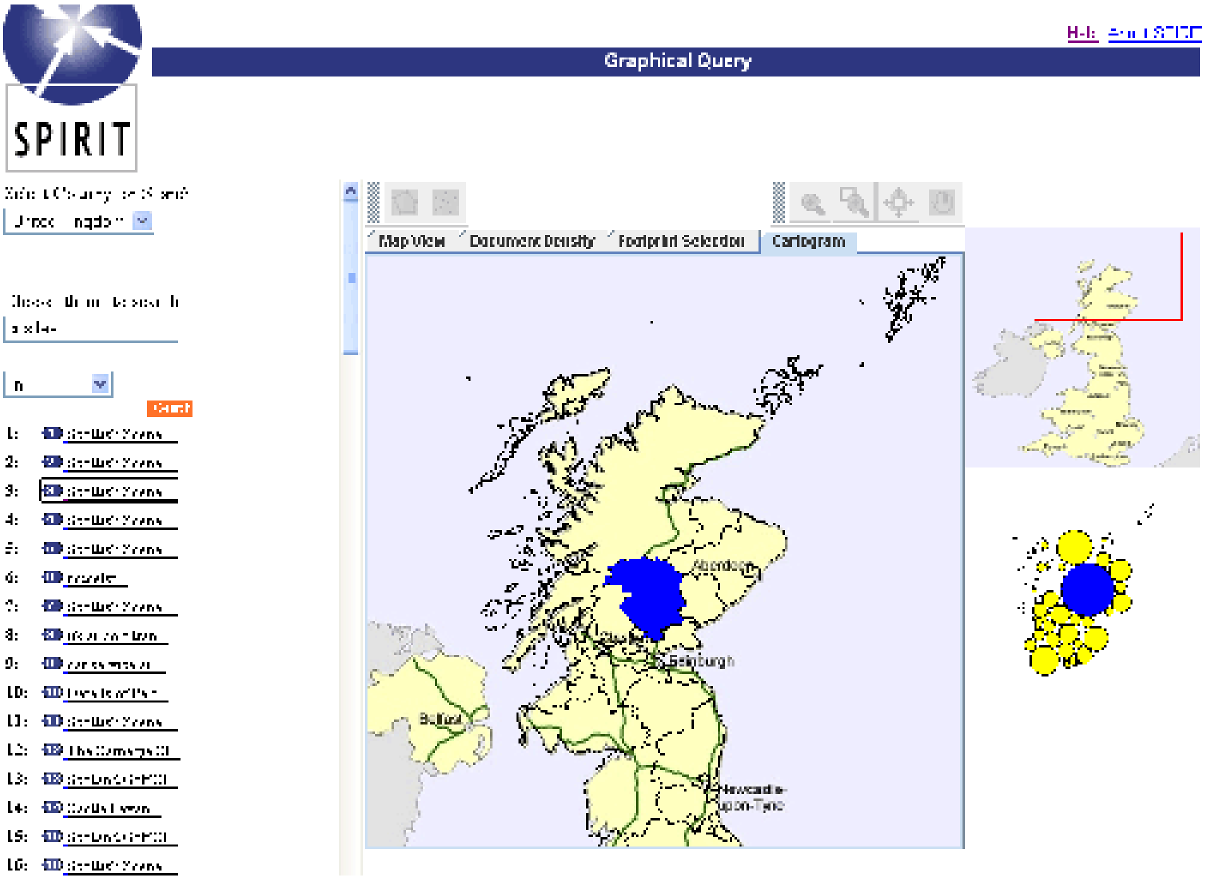


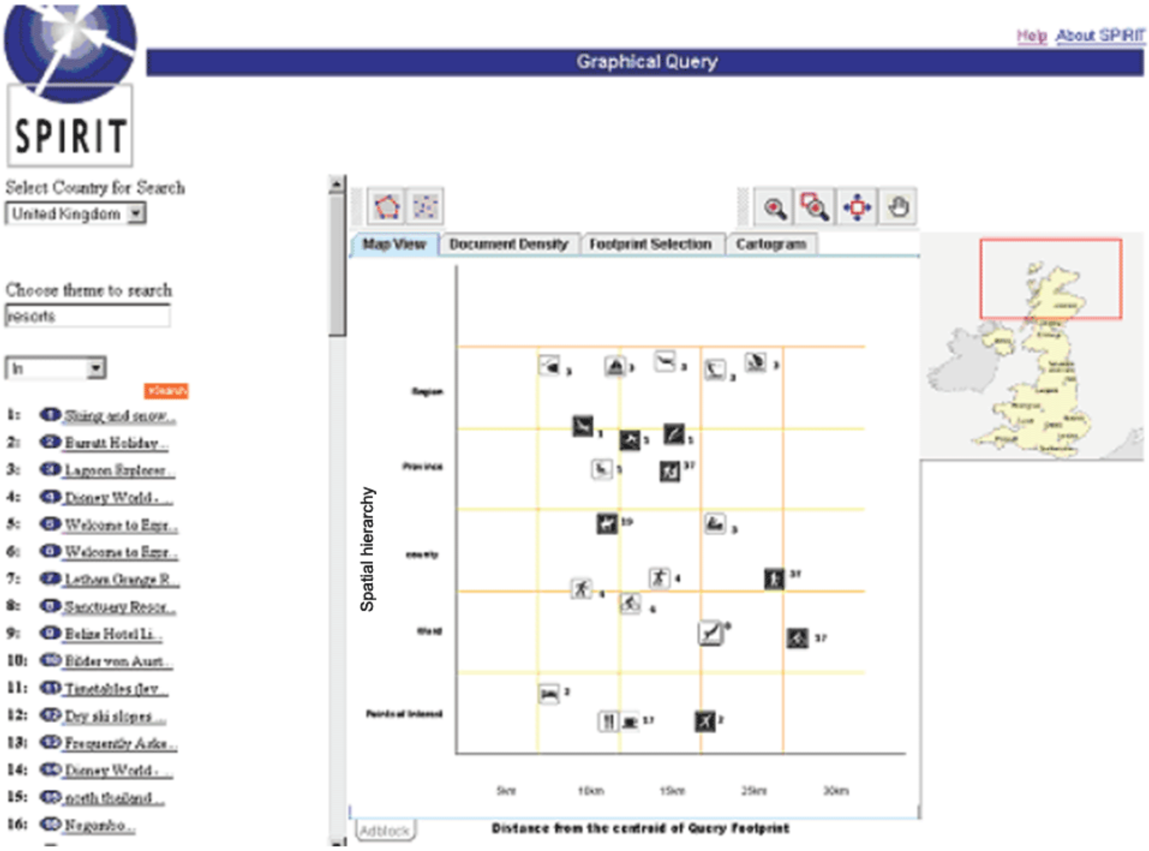
Figure 6.4: An illustration of cartogram based representation for a user's selection of a specific region for the query "castles in Scotland".

Table 6.4 Justification for the use of Dorling's Cartogram

	Pros	Cons
1.	Provides a collective summary of the frequency of the footprints in a specific geographic unit.	Fails to present a detailed perspective associated to footprints in a specific region.
2.	Theme based e.g. The population density of a specific geographic region.	Two variables can be used to demonstrate its representation.
3.	Creates more space where there are more documents.	Involves spatial transformation while maintaining limited degree of geographic accuracy.

6.4 Spatial Activity Maps / Plots

Activities like fishing, cycling, mountaineering or trekking are some of the common proponents expressed in the user queries. Taking part in such activities is governed by the availability of the resources or infrastructure at a specific geographic location.


Select Country for Search: United Kingdom

Choose theme to search: resorts

In

Figure 6.5: An illustration of spatial activity plot (a mock-up representation using real datasets) depicting spatial hierarchy on the Y-axis and distance from the centroid of expressed region of interest (candidate region) for a query "resorts in Scotland, United Kingdom".

There has been growing interest in integrating human activities and movements in space and time in the social sciences, covering wide areas such as migration, residential mobility, shopping, travel and commuting behaviour (Kwan and Lee, 2003 p.2). Integrating activities into visualizations should allow users to better understand the usage of resources available at specific geographic locations.

Spatial activity can be defined as an *enabled activity* at a specific geographic location that facilitates the usage of the resources necessary to carry out a set of activities of interest to the user. For instance in a query such as '*biking in Black Forest, Germany*', biking refers to an activity that a user wishes to perform at a specific geographic location, '*Black Forest, Germany*'. The results obtained from each document comprises the geographic location "*Black Forest of Germany*". It is evident from the example that relationships are often geographically localised. In association to this, conceptual information is available for each of these geographic locations, which helps us in defining whether an activity is enabled with respect to the availability of resources at a specific geographic location. A possible activity based representation would incorporate symbolisation of enabled activities and displaying their geographic footprints.

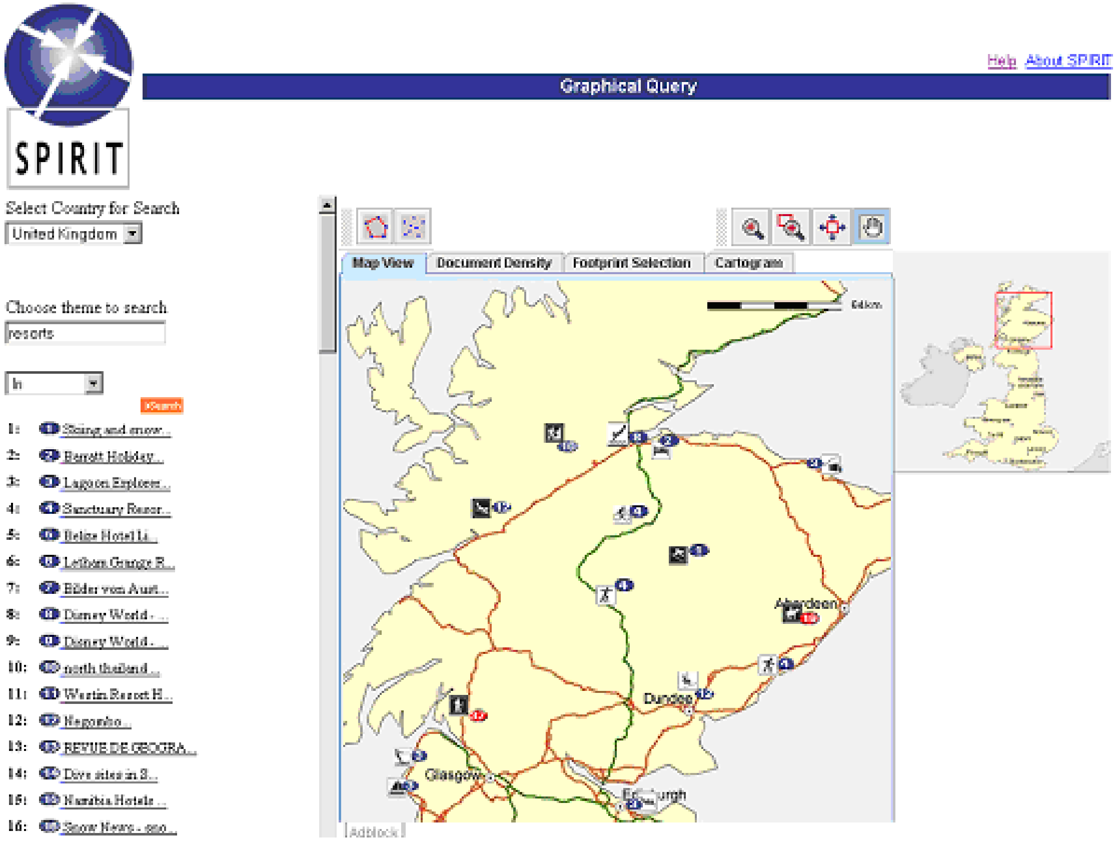


Figure 6.6: An illustration of spatial activity map using real datasets with symbols depicting the contextual details of each document for a query "resorts in Scotland, United Kingdom".

Cai & Xue (2006) state that "*an activity is the fundamental, meaningful unit of analysis to understand the interacting relationships among human minds, artefacts and the environment. Activity includes not only external resources (people, artefacts, settings) but also internal and mental processes (goals and beliefs). Such constituents of activity are not fixed but can dynamically change as conditions change. Context-awareness means that the systems actively construct and update a model of the ongoing activity by sensing, communicating and interpreting changing conditions, resources and processes. Context adaptation takes the activity as the context inorder to infer what actions are to be taken to ensure the success of the overall activity*".

All of the visualizations discussed earlier in this chapter have been developed to demonstrate different techniques of visualizing the result set. With conceptual and spatial aggregation, new visual representations can enable users to glean knowledge and gain deeper insights from the result sets. We chose the TouchGraph visualization technique to represent such relationships because of its ability to provide a high level of interactivity in exploring result sets. These visualizations aim to maintain uniformity in the representation of results based on containment and association using geographic prepositions expressed in the query triplet formulated by the users.

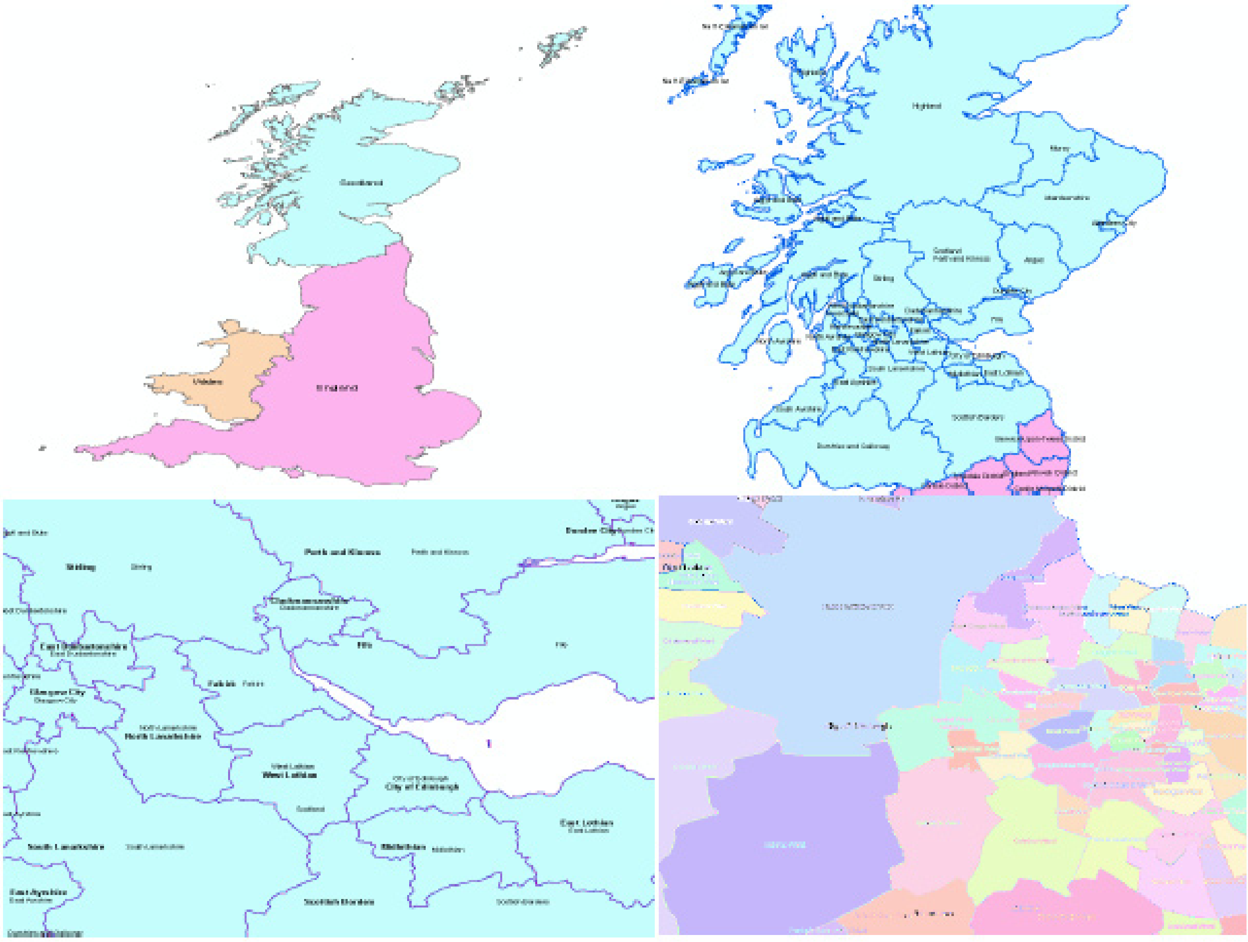


Figure 6.7: An example of spatial hierarchy and level of detail for the United Kingdom obtained from SABE dataset. a). Region (level-1) b)Province (Scotland- level 2) c). County (level-3) d). Ward (level-4)

Figure 6.7 and Figure 6.8 illustrate the data sets at different resolutions defining a certain level of hierarchy such as street level, ward level, county level etc of the SABE dataset adopted for background mapping. The scale of the representative datasets defines the granularity to which the results sets can be mapped to some extent. Figure 6.7 details the level of detail upto ward level for regions within Scotland north of United Kingdom. Mapping of the spatial extents of the footprints with reference to the available dataset enables users to glean knowledge and gain deeper insights from the result sets.

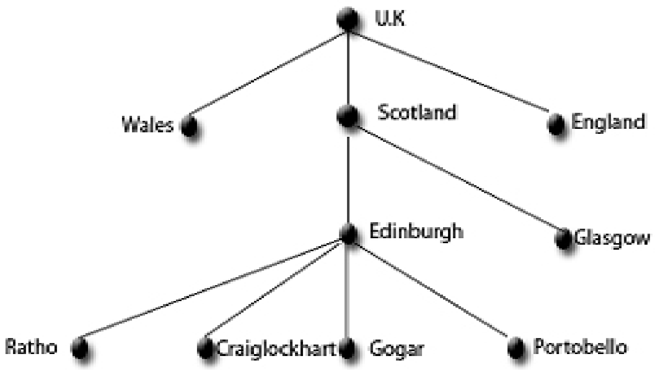


Figure 6.8: A graphical illustration of spatial hierarchy of the Seamless Administrative Boundaries of Europe data sets demonstrating the political administrative district hierarchies.

Table 6.5 Justification for the use of spatial activity maps/plots

	Pros	Cons
1.	It helps in classification of result sets based on spatial and aspatial aspects.	Requires domain specific ontologies to support such classification.
2.	Usage of symbols to associate activity to a geographic location brings in visual distinction on a map.	Each web document may refer to different contextual information making it difficult to associate it with a specific symbols.
3.	Requires appropriate levels of granularity for both contextual and spatial data.	Non-trivial to authenticate the contextual information on the web page referring to specific geographic location, which is termed *information fallacy.* (Wood *et al.,* 1989; Walton *et al.,* 1995)

Limitations in the size of visual displays prompt a need for abstract representations to reduce information clutter. The SPIRIT engine displays 100 documents, ordered by a combined ranking mechanism - which require considerable screen real estate given the constrained display size of 800 x 600 pixels. These constraints prompt an investigation of ways to aggregate information based on three themes - concepts, geography and time. Fredrikson *et al.* (1999) state that data characteristics typically conform to predominantly these three themes and depending on the task and the application domain, aggregations of differing granularity can be performed.

Conceptual aggregation: Conceptual classification of the results obtained from the SPIRIT engine can serve to aggregate objects thematically. A similar approach has been demonstrated by Kartoo (www.kartoo.com), Mooter (www.mooter.com), Touchgraph (www.touchgraph.com), SPIRE (Beth *et al.*, 1998 and Miller *et al.*, 1998). Kartoo (www.kartoo.com) demonstrates a similar classification of documents based on concepts/contexts within each of the documents, facilitating ease of identification of relevant documents of interest for the user. Montello *et al.* (2003) argues that "*spatial information systems benefit from application of knowledge of human conceptualization of space and place to system design*". A similar classification may serve to address user's needs. For an information request such as "*restaurants in Scotland*". SPIRIT returns a set of documents ranked based on combined relevance (van Kreveld *et al.*, 2004) and the results are visualized on the map. Conceptual aggregation could be achieved based on the type of restaurants such as Oriental, Chinese, Malaysian, Mexican, Thai, Italian, Indian etc. or based on the characteristics of the objects searched. However, it may be argued that the conceptual aggregation cannot fully reflect the theme of the document, as it is evident that the semantic or contextual heterogeneity within the document or result set hinders the notion of generalization.

Geographical aggregation: Yip & Zhao (1996) present a computational approach for hierarchical data analysis termed as spatial aggregation, where neighbourhood relationship plays a vital role in spatial aggregation. A similar classification based on neighbourhood relationship (adjacency) and hierarchy at the street, ward, county and country level as shown in Figure 6.8 and 6.9 details the scale of representations of the SABE dataset used for SPIRIT. It should be noted that the geographic extent of the footprint may lie within the specified geographic extent of the user's query or may well be spread across higher levels. For example: for information request such as "*castles in Scotland, United Kingdom*", the geographic extent of the footprints may well lie within the geographic extent of the user's query or may well be spread across other levels as shown in Figure 6.2. Based on the neighbourhood characteristics, footprints therefore be aggregated.

Temporal aggregation: Another characteristic of the data obtained from the SPIRIT is temporal. Documents contain information related to the occurrence of a specific event at a specific instance of time at a geographic location. Herzog (2003) points out that there is growing interest in using GIS tools to try to uncover exactly where the news is happening. News sources are extremely rich in information relating to events at a specific instance of time and at a specific geographic location. A further aggregation based on the instance of time discussed here is achievable. Understanding the complexities, these aspects can be investigated in future work.

Mapping spatial hierarchies onto a conceptual and temporal frame of reference is a non-trivial operation. A simplistic approach is adopted to identify terms closely associated with some activity or resource available at a specific location helping us to define mappings and operations with in the same spatial hierarchical levels or between different levels. In an effort to define different levels of abstractions both on geographic and conceptual aspects, the datasets are abstracted to the mathematical structure, a simple force directed graph is constructed.

This graph comprises of nodes and edges, where the nodes represent the geographic footprint and the relationship between these nodes are expressed based on the spatial relevance and concepts within each of the web documents. The edges connect to nodes according to the concept identified within the web documents. A number of graph drawing techniques which can be used to automatically visualize the data are used to generate the graph. Subsequently these graphs are mapped onto some abstract space so as to facilitate ease of understanding to the users.

6.4.1 Symbolic integration

Sometimes, contextual information available in the web documents associated with a geographic footprint does not reflect the kind of activity at that place of interest. However, data mining of this information might be best done by providing visualizations based on conceptual relationships between documents with the same geographic footprint.

Each geographic footprint is associated with a web document, which in turn refers to some contextual information describing a set of possible activities that could be carried out at that specific location. One or numerous activities could be carried out at each geographic footprint. Thus, to some extent context available in the web documents helps us in defining a spatial activity at that specific location.

6.4.2 Clutter Reduction

To improve the aesthetic appeal of the graph visualization, different algorithms are applied, of which Eades (1984) Spring Embedder is a well known graph drawing technique and has been widely adopted for visualizing geographic data such as network topologies and the internet. In addition to these, different algorithms are applied to improve visualization of data sets amenable to such visualizations (Fruchterman-Rheingold (1991), Kamada-Kawai (Kamada, 1988 and Kamada & Kawai, 1989).

A mock-up spatial activity map is created with the assumption that each document describes a specific activity, which can be associated with a symbology. Figure 6.6 illustrates a textual listing of the documents according to their combined relevance and their corresponding footprints represented as activity (symbol) plotted on a decision grid (for 5x5Kms) with Y-axis referring to the spatial hierarchy X- axis illustrating the distance of the footprint from the centroid of the query footprint. Futhermore, users are also presented with a map based perspective of spatial activities over the defined region of interest. Experimental studies were carried out to understand user's perception of this visualisation using mock-ups (Chapter 8).

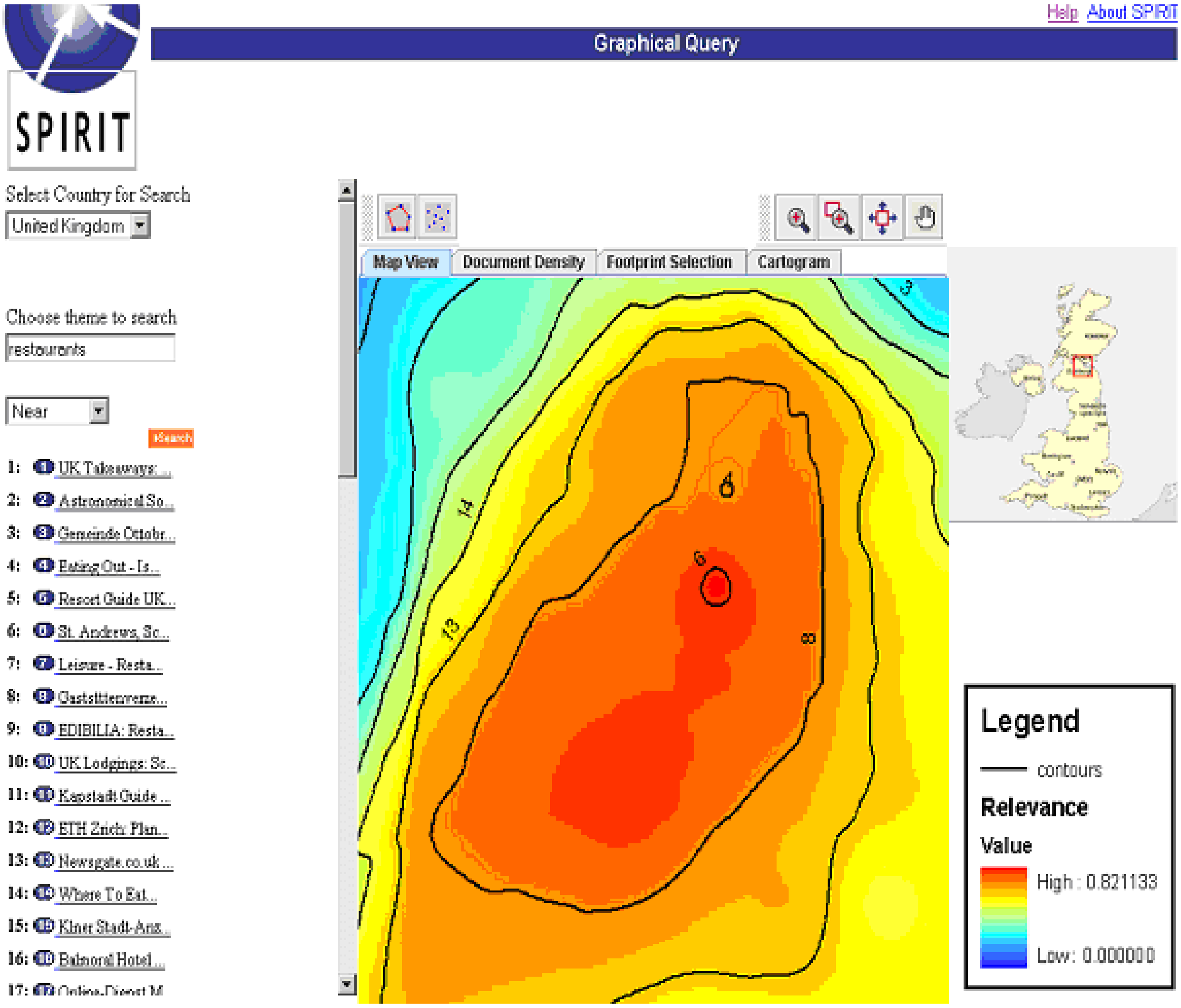
6.5 Isoline representations

A mock-up visualization is constructed using real data sets and ranking of the documents using a query such as "*hotels or restaurants near Edinburgh Castle, Edinburgh, Scotland, U.K*" as shown in Fig 6.10. With footprints in each document sharing similar locations and their occurrence in multiple documents, isolines can be used to demonstrate the spatial variability of the result sets where isolines are defined by footprints sharing the same value. Isolines were first proposed with the aim of characterizing the points with the same relevance. The choice of appropriate intervals based on the relevance score of each of the footprint can allow a visual perception of the locations of relevant documents. The scale of the map governs the purpose it serves. For example, a small scale map with constant intervals can turn out to be disadvantageous if territories with large differences in character are represented on one map sheet.

6.5.1 Method

For a query "*hotels or restaurants near Edinburgh Castle, Edinburgh, Scotland, U.K*", the result set obtained from the SPIRIT engine comprises of geographic footprints with combined ranking. The result set are plotted on the map of Scotland, U.K and isolines are constructed from the footprints sharing the same value of combined relevance.

Subsequent colour coding mechanisms are adopted for visual distinction to the regions of high intensity. A supporting legend is provided to decode the information presented in the mock-up. The efficacy of such an visualization approach is evaluated and discussed in the experimental studies in Chapter 8.



Select Country for Search: United Kingdom

Choose theme to search: restourents

Near

Figure 6.9: An illustration of isoline representation: a mock-up visualization using real datasets obtained from the SPIRIT engine.

Table 6.6 Justification for the use of isoline representation

	Pros	Cons
1.	Isoline maps symbolize changes in (geographic relevance which are colour coded) through the use of isolines linking footprints with equal relevance.	Map users face difficulty in comprehending the region between these isolines, which requires visual interpolation and can be prone to estimation errors.
2.	Isolines are often successfully employed in identification and location of areas of maximum and minimum values (geographic relevance of documents).	They require considerable skills from users to interpret the result presented and require constant attention to decipher it.
3.	Isolines imply a continuous gradient of geographic relevance within a geographic region or a hot spot. A particular geographic location may or may not have a similar geographic relevance in comparison with any other adjacent geographic location or region.	This technique is often not useful in investigation of interrelationships that exist between different regions.

6.6 Summary

This chapter presented potential visualization methods with adequate analytical capabilities to glean through the information retrieved from the web. Visualizing spatially aware information retrieved from the internet is an important but non-trivial problem, with varying granularity and complex relationships existing between individual footprints and the documents. This chapter has presented different visualization approaches that support analysis and interactive visualization of the geospatial data. Potential approaches demonstrated here include density surface representation, footprint based filtering, cartogram techniques, spatial activity plots (mock-ups constructed using real data) and isoline representation (mock-up constructed using real data). Possible cases have been identified based on the choices users embrace to deliver a functional view of their requirements demanded for the SPIRIT system. Provision of interactivity tools to support and increase the level of explorability should not only enhance users ability to find information but also to understand the information presented.

Table 6.7 summarises the strengths and weaknesses of the various representations adopted. For example, density surface representations facilitate an overview of document distribution across user selected regions, while the footprint based filtering details the individual document footprint coverage as well as the spatial extent of the footprint. Cartograms presents a collective summary of the frequency of the footprints in a specific geographic unit i.e. an administrative unit. Both density surface representations and cartograms summarize the result sets, but the former can reflect document distribution across user defined regions of interest while the latter presents a summary of the frequency of the documents within an administrative unit. Unlike the other two visualization approaches, footprint based filtering allows filtering of documents by extent and number.

Table 6.7 Comparison of Strengths and Weaknesses of various representations adopted

No	Density Surface Representation	Footprint based filtering	Cartogram	Spatial activity Maps	Isoline
Pros					
1.	The high density areas in data can be readily seen. Thus, allowing ease of exploration of document distribution across space	Overviews of document and footprint coverage/spatial extent	Provides a collective summary of the frequency of the footprints in a specific geographic unit.	It helps in classification of result sets based on spatial and aspatial aspects.	Isoline maps symbolize changes in (geographic relevance which are colour coded) through the use of isolines linking footprints with equal relevance.
2.	Summarizes large amounts of data.	Allows the filtering of documents by extent and number of footprints	Theme based e.g. The population density of a specific geographic region.	Usage of symbols to associate activity to a geographic location brings in visual distinction on a map.	Isolines are often successfully employed in identification and location of areas of maximum and minimum values (geographic relevance of documents).
3.	Gives quantified density measurements for all areas in the region in a single graphic view.		Creates more space where there are more documents.	Requires appropriate levels of granularity for both contextual and spatial data.	Isolines imply a continuous gradient of geographic relevance within a geographic region or a hot spot. A particular geographic location may or may not have a similar geographic relevance in comparison with any other adjacent geographic location or region.
4.	Can infer a density volume at any location of the study area, not just a footprint centroid.				

Table 6.7 Comparison of Strengths and Weaknesses of various representations adopted

No	Density Surface Representation	Footprint based filtering	Cartogram	Spatial activity Maps	Isoline
5.	Can create a surface representing imprecise regions and derive boundaries of imprecise regions, where a query is made representing such a region.				
Cons					
1.	Requires parameters such as frequency and bandwidth to be set correctly.	Complexity in understanding and readability due to overlapping spatial extents with in a document.	Fails to present a detailed perspective associated to footprints in a specific region.	Requires domain specific ontologies to support such classification.	Map users face difficulty in comprehending the region between these isolines, which requires visual interpolation and can be prone to estimation errors.
2.	Difficult to compare analyses from different times and regions.	Represent documents whose true footprint is either a point or a polygon as a rectangle, which implies a high degree of generalization.	Two variables can be used to demonstrate its representation.	Each web document may refer to different contextual information making it difficult to associate it with a specific symbol.	They require considerable skills from users to interpret the result presented and require constant attention to decipher it.
3.	Doesn’t use administrative boundaries, making it difficult to respond to these high density areas.		Involves spatial transformation while maintaining limited degree of geographic accuracy.	Non-trivial to authenticate the contextual information on the web page referring to specific geographic location, which is termed *information fallacy.*(Wood *et al.,* 1989; Walton *et al.,* 1995)	This technique is often not useful in investigation of interrelationships that exist between different regions.

Table 6.7 Comparison of Strengths and Weaknesses of various representations adopted

No	Density Surface Representation	Footprint based filtering	Cartogram	Spatial activity Maps	Isoline
4.	The density surface is calculated over a map view area rather than the political or administrative boundaries.				
5.	Density surface representation doesnot take into account combined/geographic relevance of the documents.				

The density surface representation can also help us to derive boundaries of imprecise regions by creating a surface representing imprecise regions as a document which may also include footprints referring to places other than described in the user's query (candidate region). This approach requires parameters such as frequency and bandwidth to be set and unlike cartogram it doesn't use administrative boundaries, making it difficult to respond to the high density areas i.e. the density surface is calculated over a map view area rather than by using political or administrative boundaries. Although cartograms present a thematically-based summary (for example with respect to museums) with respect to administrative units, they mask detailed information about variation within these administrative units. Footprint based filtering is complex, and may have problems in terms of understanding and readability due to cluttered and overlapping spatial extents within a document. With all methods, the footprint extent adopted here depends on the accuracy of the geo-coding process.

Unlike density surface representations, isoline maps represent combined relevance (colour coded) within a geographic region. They help in identification of peaks and valleys, areas of maximum and minimum combined relevance but are not useful in investigation of the interrelationships that exist between these different regions.

Spatial activity plots or maps helps in conceptual or spatial classification of result sets, depending on the availability of domain specific ontologies populated within the geographic ontology. Each web document might refer to different contextual information for a specific geographic location, which in turn increases the complexity of visualizing this information. A possible solution is the usage of generalised symbology that reflects the contextual information referring to that specific geographic location in a web page.

The practical usability of various visualizations demonstrated here need to be assessed by developing a comprehensive user study. Subsequent chapters illustrate such a framework and the calculation of empirical measures from these user studies to compare the effectiveness of various visualizations adopted to represent spatially aware information retrieved from the internet.

Chapter 7

Evaluation Methodology

reinforcing theoretical foundations through empirical evidence

7.1 Underlying theories

Usability is a general quality of appropriateness to a purpose, ease of use and learnability any system serves to address (Preece *et al.*, 2002). Measures of appropriateness of a specific representation can be delivered by devising a model to assess with real users. These measures need to be unobtrusive in limiting the influence of the researcher. Several theories are in place, which lay emphasis on system performance: precision and recall being the key measures.

Borlund (2003) presented a model for evaluating interactive information retrieval (IIR) systems that not only aims to "*facilitate the evaluation of IIR systems as realistically as possible with reference to actual information search and retrieval process, though still in a relatively controlled evaluation environment*" but also "*to calculate the IIR system performance taking into account the non-binary nature of the assigned relevance assessments*".

The model presented by Borlund (2003) discusses the creation of a simulated work task situation that facilitates to capture the user's perceptions and situational relevance or subjective relevance (illustrated in Table 7.1).

It has been strongly argued by Borlund (2003) that "*cognitive and relevance revolutions require realism with reference to the formation of information needs, and relevance assessment processes*". These processes are subjected to an individual's perception, knowledge and adopted strategy. The 'interactive revolution' is governed by system feedback and spontaneous reaction (sensory responses) of the users.

Table 7.1: Borlund's model for evaluating interactive information retrieval (Borlund, 2003)

No.	Components of the experimental settings	Activities
1.	Functional, valid and realistic settings.	• focus group. • perceptions and feedback. • measure to record multi-dimensional and dynamic relevance.
2.	Simulated work task/use-case scenario	• selection of information source • problem definition. • objectives & strategies. • permute the order of the search jobs
3.	Empirical measures	relative relevance assessment • Situational relevance • Intellectual topicality • Algorithmic relevance

Winckler *et al.* (2004, p. 167) outlined a set of guidelines that were employed in the construction of simulated work task/use-case scenarios, emphasizing on task abstraction levels aimed to:

- *Produce unambiguous representation of user goals and intents as well as the logical decomposition of sub-tasks required to accomplish that goal.*
- *Define the rendering functions (how the application will be presented in the tasks) that must be checked during the evaluation.*
- *Associate rendering functions to interaction mechanisms enabling the visualization.*
- *Extract a set of possible scenarios.*

Furthermore, Winckler *et al.* (2004) proposed a task abstraction method detailing the level and scope of each task and how the experiments need to be organized to draw meaningful inferences of the same.

Table 7.2 Task abstraction level and scope (Winckler *et al.*, 2004 p167)

Task level	Scope
Goal	User goals and intents
Abstract	Generic tasks
Interaction	User and system interactivity
Visual presentation	Rendering functions performed by the application

These theories can be adopted to build an evaluation framework for spatially aware information retrieval systems, which focused on two main issues:

1. Considering the system as a whole, evaluating the spatially aware information seeking model and the success of the retrieval (precision and recall) by a measure of perceived relevance.
2. Evaluation of various visualization methods adopted for representing the spatially aware information retrieved and the interactions involved in accomplishing the tasks.

Much of the work presented here focuses on the perceived relevance of the geographic information retrieved and on drawing usability metrics for various visualization methods adopted for representing spatially aware information retrieved. Preliminary evaluation studies regarding the first issue have been carried out by Bucher *et al.* (2005). More results will be reported in Section 8.4. The main part of the experiments reported in Chapter 8, however, focus on the second evaluation issue.

A comprehensive task structure is drawn with respect to information, learning and interaction spaces for the SPIRIT system, facilitating us to map user tasks to the functionalities of each of the components or visualization variables. The workflow of the structured interface and the graphical interface (advanced interface) provides an outline of interactions between various components of the SPIRIT system as a whole. However, this limits the ability to explore the usability of various functionalities and interactions involved within each of these components.

7.2 Materials and Testing Environment

The experiments aimed to address web-based users. Personal computers with Pentium IV configuration and Java plug-in enabled web browsers (Internet Explorer and FireFox/Mozilla) have been used. Figure 7.1 shows the evaluation framework (explained in detail in Section 7.4), wherein a series of preliminary pilot studies are carried out as dry runs prior to commencement of the actual tests. These pilot tests helped to improve and consolidate the questionnaire and streamline the logistics of the experiments. Furthermore, in an effort to help users during the self-learning phase, tutorials have been designed so that users can learn through the functionalities of the SPIRIT systems as a whole. In addition, to comprehend an individual's feedback with respect to the design, functionality and learnability of the SPIRIT prototype system, the WINCAM 2000 software was used to document their spatially aware information seeking process in video recorded sessions.

7.3 Focus groups

Valuing multiple perspectives that come into play during the evaluation phase, participants with varied academic background, varied experience and exposure from different science disciplines were included. The participants shared the characteristic of using geographic data regularly. Due to the nature of the studies, individuals invited to take part in the studies were not expected to have varied experience and exposure to the usage of visualization tools but rather basic geographic domain knowledge. For instance, the academic background of the participants selected varies from biology to geography. In particular, the number of participants available for carrying out the studies is crucial to draw significance of the various experiments and questions addressed. Hence, the evaluation procedure was originally designed to obtain around 30 participants in each of two phases, which amount to 60 participants in total. Eventually, a total of 50 (26 + 24) participants actually participated in the experiments *(i.e. 10 persons less than originally planned).*

The focus groups can be broadly classified into three categories:

- Students (first year geography students)
- Teaching assistants and PhD students (with considerable level of geographic knowledge).
- Experts (academic research staff in GIScience).

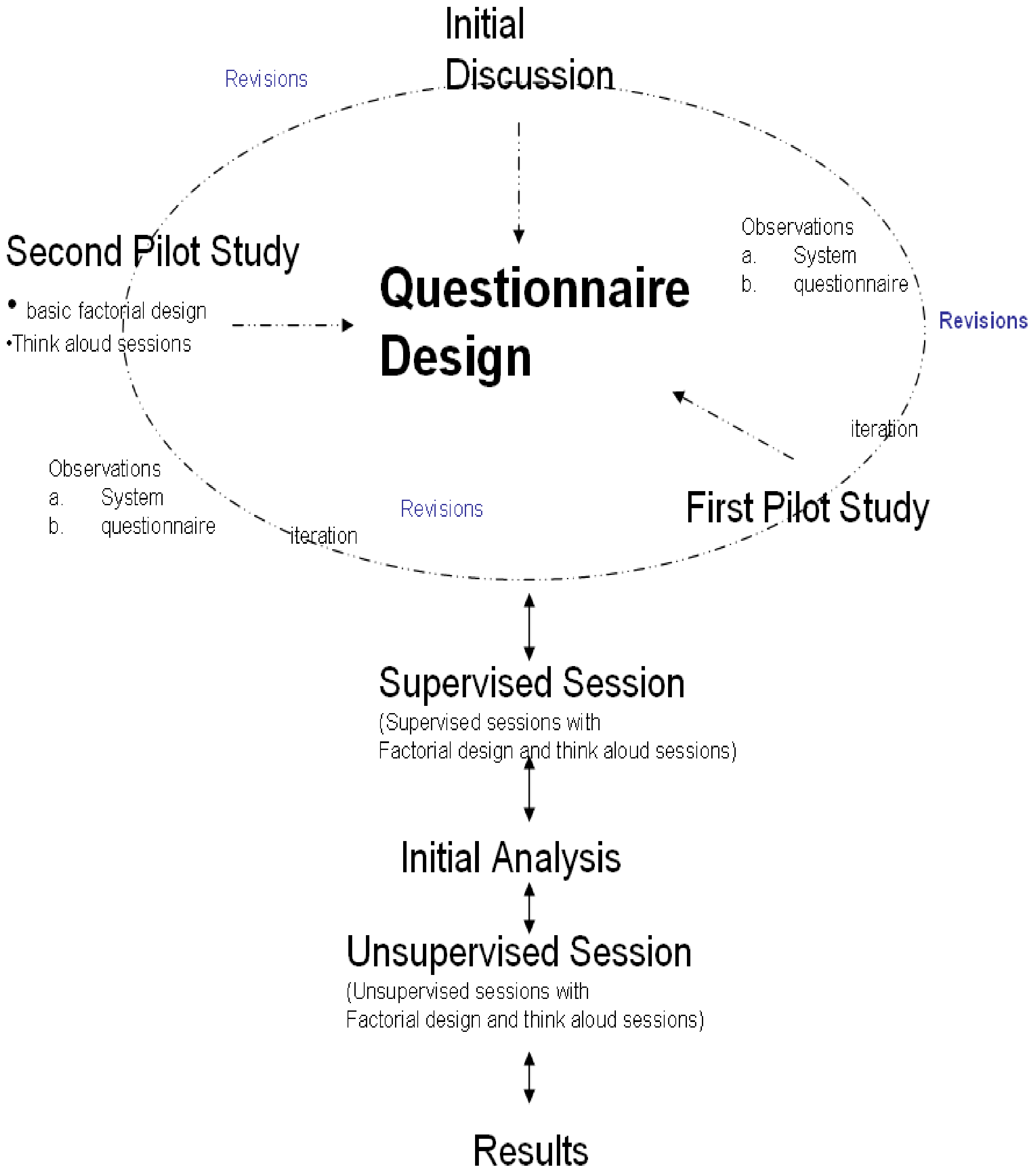


Figure 7.1: Evaluation Framework

7.4 Procedure

An experimental design is touted to be the most prominent of the research designs against which all other designs are validated. A well-constructed design is probably the strongest with respect to its internal validity, for it is at the core of all the cause and effect inferences. The objectives of the experiment design in this thesis are to draw usability metrics for visualization methods and techniques adopted for spatially aware information retrieved from the internet. The scope of the design is to assess system capabilities identified in the requirements analysis (*e.g. geographical query expansion, place name identification, disambiguation and the use of spatial relationships; cf. Table 3.22; 3.23*) and the visualization techniques adopted to represent spatially aware information retrieved.

Figure 7.1 outlines the evaluation procedure to draw meaningful inferences on the usability of various visualization techniques adopted. In the following subsections 7.4.1 and 7.4.2 the elements of the procedure are discussed.

7.4.1 Questionnaire design

To document users' qualitative and quantitative feedback, questionnaires are formulated through a number of iterations. An initial discussion of the questionnaire has been carried out involving expert system designers, leading to the revision of the questionnaire. Two pilot studies were carried out to enhance the questionnaire. Table 7.3 describes the participants involved in the pilot study. Observations are drawn with respect to system functionality, performance and comprehension of the questions posed.

Table 7.3 Description of the participants involved in the design of questionnaire & (I,II) Pilot studies

User	Occupation	Research Interests & Expertise
1.	Ph.D. Student (I)	GIScience, Spatial Cognition, Wayfinding
2.	Ph.D. Student (I)	GIScience, LBS, Map Generalization
3.	Ph.D. Student(I)	GIScience, Map Generalization
4.	Research Associate (II)	GIScience, Map Generalization
5.	Ph.D.Student(II)	GIScience, Spatial Cognition
6.	Ph.D. Student(II)	GIScience, IceSheet Modelling
7.	Ph.D.Student(II)	GIScience

I-indicates first phase of pilot studies; II-indicates second phase of pilot studies.

The pilot study was carried out in 2 iterations. During the second iteration, factorial design and think aloud sessions were incorporated. The factorial design serves to address the evaluation of multiple variables. Factorial experiments are adopted when two or more factors are tested, each with discrete possible values whose experimental units take on all the possible combinations of these levels. Such an experiment allows studying the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable. A promising feature of this approach is that it allows the effect of several factors and even interactions between them to be determined with the same number of trials as are necessary to determine any one of the effects by itself with the same degree of accuracy. (Box *et al.*, 2005) Appendix-H details the factorial design adopted to impart a uniform learning effect, that is, in order to even out the learning effect that may occur as test subjects are getting increasingly acquainted with the SPIRIT prototype over time.

For objectives such as assessing various factors such as *geographic prepositions, place name disambiguation, geographical query expansion etc*. in total 7 participants were involved in two pilot studies. The pilot studies were carried out using 7 real-world task scenarios, each reflecting the key factors identified to be assessed using simple and advanced interface (as shown in Table 7.5). The experiments were designed for 60 subjects, divided into 4 groups (I-IV) each of which were assigned a real-world task scenario, in random order so as to accumulate around 15 respondents for each real-world task situation based on the type of interface expressing the user's interest (i.e. Graphical interface vs. Textual interface/Structured interface). To impart uniform learnability of visualization and interactions, a cyclic design was incorporated as shown in Table: H.1 (Appendix-H), enforcing users to interact. Six possible combinations (3! = 6) involving the three visualization methods density surface, footprint based filtering and cartogram techniques were used.

In addition to these real work task scenarios presented, two additional design concepts (*Spatial Activity Maps/Plots - partially implemented & Isoline Representations - mock-ups with real data*) have been included for all the users to draw qualitative feedback.

Revisions were made based on the observations and insights drawn from the design iterations. The number of real-world task scenarios and the combinations of geographic prepositions had been reduced to four (4). Table 7.4 illustrates the selected set in **bold** and highlighted with a light grey background. Scenario III has been assigned to all the groups for the structured query interface, while among the rest of the groups one scenario (II, IV or VI) was assigned. Which means that the users were given two real world task scenarios, one for the structured query interface and another one for the unstructured (graphical) query interface. Appendix-J details how the questionnaire was designed and presented to the users. Once again, note that all the groups had to carry out scenario III during their experiments.

Table 7.4 Real work task scenarios envisioned during the experiment design phase.

Scenario	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**
Query	Museums in London	**Cottages in "west of Scotland"**	**Hotels near Portree**	**Mountaineering in Scotland**	Restaurants in Edinburgh	**Biking south of London**	Fishing west of Wales.
Geographic preposition	in	**in**	**near**	**in**	in	**south of**	west of
Interface	Unstructured	**Unstructured**	**Structured**	**Unstructured**	Structured	**Unstructured**	Unstructured
User groups	I,III	**II,IV**	**I,II,III,IV**	**IV,II**	III,I	**IV,II**	III,I

The WINCAM2000 tool has been used to track users' interactions and record their responses during the think aloud sessions (video recordings). Participants were randomly picked using a system that generated random numbers between 1 and 60.

Subsequent to the design iterations, 2 participants considered experts in the field of GIScience were chosen to carry out initial tests to assess the feasibility and to analyse the initial outcomes of the revisions. Tutorials were provided detailing functionalities and snapshots of the interface and possible interactions. *Pilot test subjects were requested to criticise and identify critical questions necessary for better validation of visualization methods.* Additionally, video recording sessions were carried out to document their interactions and observations during each session. The following are the findings of these initial tests:

- *Requirements of the Java Virtual Machine patch update made the UNIX based testing environment inconvenient.*
- *Strange behaviours of the interface were recorded on Macintosh platforms and also browsers other than Internet Explorer.*
- *It is necessary to provide colour printed tutorials, to help better understand the colour schema adopted.*
- *Restructuring of the on-line help by introducing system functionalities first and bug descriptions in the latter half was shown as necessary.*
- *Redundancy in information needs to be removed, as conflicting messages were conveyed in the selection of a region (drop down) through graphical selection on the map. It has to be changed to "Select the Country".*

- *Language and usage of terms have to be made simple and easily understandable to the subjects.*
- *The provision of zoom-out to the previous level would prove to be extremely useful.*
- *The geographic extent needs to be clearly outlined in the tutorial.*
- *Topics within the 'coverage' of the document collection and data limits (spatial extent and granularity of geographic data) of the SPIRIT final prototype need to be chosen.*

7.4.2 Evaluation Phase

The experiments were carried out in two phases: Unsupervised and supervised sessions. During the unsupervised sessions, participants accomplished their objectives on their own, whereas in supervised sessions, users were given a set of training or warm-up exercises similar to those they would encounter during the test and that involved the same data set as in the main evaluation. Queries addressing the key system functionalities were addressed during the training sessions.

The rigorous training sessions for simple and advanced interfaces were carried out using real-world task situations such as "*Checkpoint Charlie in Berlin, Germany*" and "*Fishing in Scotland, United Kingdom*". The training sessions lasted for around 30 minutes. The queries were chosen to demonstrate the spatial extent of the results.

The strategy is aimed to understand the impact of learnability of key visualization methods presented to the users to represent spatially aware information and to draw subsequent inferences based on the exposure and training of the system as a whole. The experiments in the two phases were carried out as follows:

- **Pretest:** During this stage qualitative information relating to the user's background, geographic knowledge, search expertise and exposure to various geo-web based services were gathered. Empirical measures were captured relating to the users' response.
- **Self-learning:** Referring to the SPIRIT tutorial, users were instructed to carry out real-world task situations for a time span of 15 minutes. The users were free to choose any of the interfaces. Supporting instructions were provided to search for either "*hotels in Scotland*" or "*hotels near Glasgow*".
- **Real-world tasks:** Two different real-world task scenarios have been assigned to the users. Each for the simple and the advanced interface respectively.

Prior to the supervised sessions, an initial assessment of the results was carried out to gather an initial overview and understanding of the results. With the experimental pattern held constant, the supervised sessions were carried out in a similar fashion as the unsupervised sessions.

7.5 Defining user tasks

Understanding elementary tasks involved in the spatially aware information seeking process is a fundamental step towards assessing the usability of each of the visualization methods. This approach of identification of user tasks has profound impact on the way each of the visualization methods are perceived by the users. Chapter 3 has identified these elementary tasks through an exhaustive series of experiments outlining the process of spatially aware information seeking but hasn't associated them with the visualization methods and techniques adopted. Table 7.6 presents a comprehensive task structure of the SPIRIT system based on the task typology proposed by Crampton (Table 2.4 in Chapter 2).

Table 7.5: Comprehensive task structure of the SPIRIT system, grouped by components and functionality. Subsequent mapping on to the information and interaction typologyies proposed by Crampton (2002).

Phases/Task	**Interaction task (Crampton, 2002)**	**Functionality**	**Example**
Pre-retrieval			
Query formulation			
Textual	Define	Describe the place name and concepts which are required.	Enter a query "castles near Edinburgh".
Graphical	Locate	Describe a region of interest (polygon) and concepts which are required.	Drawing a polygon region for Edinburgh.
	Define	Describe specific terms and how they are used.	Define query term e.g. "castles".
	Relate	Associate concepts with the region of interest specified as a polygon.	Spatial relations such as "near" or "in" or "north of".
	Associate	Associate place name to the region of interest.	Draw a polygon for a given query (e.g: castles in the north of Scotland).
	Rescale	Scale to the required level of detail to define the region of interest.	Identifying chalets near lake Windermere, England, users then zoom in to the required level of detail to express their region of interest graphically.
	Pan	Navigate across the region of interest.	Identifying resorts along the lakes in the United Kingdom, a user browses or navigates through the map for lakes at the required level of detail to identify lakes in England (Lake Windermere) and Scotland (Loch Lomond near Glasgow and Loch Ness near Iverness)
	Scope	Show the extent of geographic information space.	A query such as "Hotels around Grampian Highlands in Scotland" or "Hotels close to Ben Nevis"
Retrieval Phase			
Query Expansion	Expand	Supplementing terms to meet the user's requirements either conceptually or geographically	Cities by the river Thames can be expanded to Cities by the river Thames, close to the Cotswolds

Table 7.5: Comprehensive task structure of the SPIRIT system, grouped by components and functionality. Subsequent mapping on to the information and interaction typologyies proposed by Crampton (2002).

Phases/Task	Interaction task (Crampton, 2002)	Functionality	Example
	Diverge	Changing the course of action either by a geographic region or a conceptual term	A query for "Chalets in Grampian Highlands near Ben Nevis, Scotland" *can diverge to* another geographic region of interest such as "Chalets in Cambrian Mountains, Wales"
	Assign/ Associate	Supplementing a term with a specific geographic location, obtained on initial query.	A query such as "Chalets in Grampian Highlands" can be supplemented with additional geographic references such as "near Ben Nevis, Scotland" changing the initial query to "Chalets in Grampian Highlands near Ben Nevis, Scotland".
	Discriminate	Excluding terms or geographic regions by visual comparison	Find "coal bearing areas in United Kingdom", would return: " Scottish field in the Central Lowlands, near Newcastle, Nottingham, Manchester, Cardiff". Users can then glean through the results for a specific region such as England region of the whole set of results obtained.
Query Disambiguation	Select/Emphasize	Choosing a specific place name within a list of possible matches.	"London, U.K" "London, Ontario, Canada"
	Switch	Change around, as to a new order or sequence; conceptual or geographic	Choosing "London, Ontario, Canada" instead of "London, U.K".
	Correlate	Correspond to a geographic location or concept based on the initial systems response.	"London, U.K" vs. "London, Ontario, Canada"
Query Refinement	Emphasize/ Generalize / Reveal	Cater to a popular pattern emerging from the results specific to a geographic location / concept	For "Cities by the River Thames", we may obtain a number of results indicating Stratford-upon-Avon, Newcastle-on-Tyne. Users shift their emphasis on these geographic locations, which are close to river Thames.

Table 7.5: Comprehensive task structure of the SPIRIT system, grouped by components and functionality. Subsequent mapping on to the information and interaction typologyies proposed by Crampton (2002).

Phases/Task	Interaction task (Crampton, 2002)	Functionality	Example
	Switch	Change over a pattern based on concept or geographic location.	For "Cities by the River Thames", results obtained refer to Stratford-upon-Avon, Newcastle-on-Tyne. Users first select Stratford-upon-Avon and find the results not so interesting and shift their focus on to results about Newcastle-on-Tyne
	Associate	Make logical deduction based on the observed pattern with a concept or geographic location. *Three types of associative attributes and expressions are identified (Chu & Zhang, 1997): 1) simple associative attributes: attributes of relations in the current query. 2) extended associative attributes: attributes of relations introduced into the query by joins with existing relations; and 3) statistical associative information: aggregate functions that are related to the main entity in the query.*	For a query such as "Airports in Europe that can land AirBus /Boeing 787". The outcome of such a query is expected to list all the airports in europe with a runway that can accomodate Boeing 787 or Airbus which is a feature specific query associated to the airport.
	Distinguish	Make comparison of different patterns at different geographic locations.	For "Cities by the River Thames", results obtained refer to Stratford-upon-Avon, Newcastle-on-Tyne. Users compare the results obtained from these two regions.
	Discriminate	Recognize or perceive the difference in the pattern and choose based on geographic location or concept.	For "Cities by the river Thames", results obtained refer to Stratford-upon-Avon, Newcastle-on-Tyne. Users discriminate the results based on the resources, pricing and proximity to river Thames.
Post-retrieval			
Result Presentation			
Textual Listing	Rank/order	A relevance measure imparted by the system and enforced on the user based on a set of formulated measures.	A simple textual listing of results for "Chalets near Lake Windermere, England".

Table 7.5: Comprehensive task structure of the SPIRIT system, grouped by components and functionality. Subsequent mapping on to the information and interaction typologyies proposed by Crampton (2002).

Phases/Task	Interaction task (Crampton, 2002)	Functionality	Example
	Categorize/Classify/Cluster	Group the results either conceptually or geographically.	For a query "Chalets near Lake Windermere, England", they could be grouped based on facilities.
	Associate/ Correlate / Outline	Correspond to a specific pattern and its association with a geographic location.	Associate "B&B" to "Chalets near Lake Windermere, England".
	Browsable summaries	Snippets to draw an overview	Descriptive information available for each result for a query "Chalets near Lake Windermere, England".
Map based representation	Associate, cluster, locate (proximity)	Correspond and group the results geographically.	Numerous web pages referring to same geographic location, "Chalets in Grampian Highlands".
	Categorize, cluster, distinguish (similarity)	Grouping or assigning based on specific characteristics	Grouping similar geographic footprints using a stack or graph
	Reveal (continuity)	Occurrence of a pattern across a geographic region.	Identify wine growing regions in europe where annual rainfall is between 432 mm and 520 mm per square meter.
	Correlate (composition)	Correspond to a specific pattern and its association with a geographic location	Identify wine growing regions in europe where gross per capita income is above 40,000 Euros
	Tabulate, plot, structure, trace	Structure results sharing a same location	Numerous web pages referring to the same geographic location, "Chalets in Grampian Highlands" are stacked or represented as a graph.
	Label, symbolize, portray	A descriptive marker for unique identification	Symbolization based on combined relevance using stacks
	Scale, pan, zoom, overview, browse	Show relative extent based on user's selection.	Zooming to the required level of detail to select the region northeast of Scotland for checking results of the query "Hotels near Portree, Scotland".

Table 7.5: Comprehensive task structure of the SPIRIT system, grouped by components and functionality. Subsequent mapping on to the information and interaction typologyies proposed by Crampton (2002).

Phases/Task	**Interaction task (Crampton, 2002)**	**Functionality**	**Example**
	Generalize, brushing & linking.	Speak in generalities, visual amplification by association.	On selection of a specific stack on the map, the corresponding link/results are highlighted, indicating the result set and also correspondingly highlight documents which share similar geographic location
Density Surface Representation	Identify, locate, associate, cluster	Select regions, stipulate the position of an object, place or event in relation to a specified object, place or events; make logical cause connection; grouping based on specific characteristics.	Select high-density regions and low-density regions spanned across a geographic region.
	Scale, zoom, pan, overview, scope.	Navigate based on user's selection.	Navigate to the peak density regions on the map.
	Highlight, brushing & linking	Visual amplification by association.	Highlighting geographic footprints by using dark colour dots on the density surface representation.
	Differentiate/ compare		Compare between regions of high document density as shown in Figure 6.1. The region around Glasgow has higher document density in comparison to the region of Edinburgh.
Footprint based filtering	Identify, locate, associate, cluster.	Select regions and stipulate the position of an object, place or event in relation to a specified object, place or event; make logical cause connection; grouping based on specific characteristics.	Geographic footprint for McDonald's in Zurich would be limited to Zurich, while the website of McDonald's would have multiple geographic locations spanning across Switzerland.
	Scale, scope, zoom, pan, overview, browse	Navigate based on user's selection	Zooming to the required level of detail for gleaning through geographic footprints for the region of Portree in the northeast of Scotland

Table 7.5: Comprehensive task structure of the SPIRIT system, grouped by components and functionality. Subsequent mapping on to the information and interaction typologyies proposed by Crampton (2002).

Phases/Task	Interaction task (Crampton, 2002)	Functionality	Example
	Highlighting (including), filtering (excluding), correlate, multiple views.	Selective inclusion and refinement based on a set of criteria defined by the user.	Interactively select the geographic footprints based on the geographic scope of the documents and the number of documents.
Cartogram	Identify, locate, associate, cluster.	Select regions, stipulate the position of an object, place or event in relation to a specified objects, places or events; make logical cause connection; grouping based on specific characteristics.	Regions with high document density have larger size in comparison to regions with lower document density.
	Scale, scope, zoom, pan, overview, browse	Navigate based on user's selection.	Zooming to the required level of detail for Portree, in the northeast of Scotland.
	Highlighting (including), filtering (excluding), correlate	Visual amplification by association.	Corresponding selection of circle to highlight the results from the document collection ranked by their combined relevance.

7.6 Empirical Criteria

Mere statistical system performance cannot deliver perceived value indicators that measure the relevance of the information retrieved or presented to the users. Explicit criteria are presented to evaluate the effectiveness of task performance in the user's perception (which means context-specific criteria). Indicators of agreement and disagreement to a set of questions were established by using a 5 point Likert scale (ordinal data), which is often called a summative scale. Sometimes data from Likert scales can be reduced to the nominal level by combining all agreements and disagreements into two categories of "accept" and "reject".

Different criteria are adopted at each level to assess the usability of specific visualization methods and techniques with respect to the tasks they serve. Similar qualitative criteria are adopted for exploratory studies by using control questionnaires to understand the parameters when running through real-world situations.

Table 7.6: Empirical criteria adopted for comparing four (4) different visual representations (Map, Density, Footprint, Cartogram)

	Questionnaire	Criteria/ Interaction tasks
1.	This visualization was easy to use.	Ease of use, interactivity, explorability.

Table 7.6: Empirical criteria adopted for comparing four (4) different visual representations (Map, Density, Footprint, Cartogram)

	Questionnaire	Criteria/ Interaction tasks
2.	It was easy to understand the information presented by this interface.	Understandability, information presentation.
3.	The use of legends and annotation made it easier to understand the visualization	Supporting information for understanding or interpreting the information presented.
4.	The use of colours was useful and appropriate.	Distinction, discrimination, visual relevance.
5.	The visualization provided a good way of looking at the results.	Multiple-view capability, understandability
6.	Zooming was a useful tool in looking at the results.	Filtering (excluding) & highlighting (including), sorting or re-expression.
7.	It was easy to find documents which shared the same location.	Findability, selection.
8.	It was easy to get confused when using this interface.	Counter check for ease of use and understandability.
9.	This visualization made it easy to find the most relevant documents for my task.	Filtering and highlighting.
10.	Identify the regions with highest document density.	Explorability and interactivity
11.	It was easy to find the regions of the documents.	Explorability, interacitivity, highlight.

7.7 Summary

This chapter presented a framework for reinforcing theoretical foundations through empirical evidence. The underlying theories help us design an evaluation framework that allows to establish empirical measures for assessing the practical usage of various visualization developed to represent spatially aware information retrieved from the internet. The experiments were designed to accommodate both the representative and communicative component according to Raper *et al.* (2002), and test the interaction tasks proposed by Crampton (2002). In an effort to optimise the design of experiments, several of iterations were carried out for the design and identification of the questionnaire used in the experiments, and to determine the relevant tasks to assess the different visualizations techniques adopted.

The experiments were carried out in three phases for both supervised and unsupervised sessions: pre-test, self-learning, and real-world tasks. Each of these was serving different purposes and addressed different objectives. In total four real-world task scenarios had been adopted and each of the visualizations were presented in a random cyclic order so as to eliminate bias induced by a pre-defined order of representation. Furthermore, to document and analyse individual interactions a few users were randomly selected for think aloud sessions. Finally, the empirical measures were drawn to assess the user-defined tasks. The subsequent chapter presents the results and insights drawn from the experimental studies.

Chapter 8

Results

“As one explores phenomena or ideas at the frontiers of scientific knowledge it is the unexpected that provides the clues to guide further work”

- Lewis Wolpert 1929 : The Unnatural Nature of Science (1995)

This chapter presents the results of the studies and discusses critical findings of the empirical studies carried out based on the framework presented in Chapter 7. We start off in Section 8.1 with some preliminaries about the participants and the experiment sessions that were held, including missing values, participant profiles, time taken to complete tasks, and self-learning exercise. By far the most part of the results will be presented in Section 8.2, which provides the results of the experiments with the visualisation techniques implemented in the SPIRIT prototype system. This is followed by Section 8.3, where the results of experiments with mock-up visualisations are reported. Additional observations of think aloud sessions are summarised in Section 8.4, followed by initial findings in Section 8.5, and a summary in Section 8.6.

8.1 Preliminaries

8.1.1 Missing values

The responses gathered from the experimental studies are discussed to draw statistically significant measures of different visualizations adopted. In total, 50 participants carried out the experiments in two phases (unsupervised-26 and supervised-24). Participants responded by selecting an integer value along a Likert Scale of 1-5 to rate their confidence values, which refer to the level of agreement (Strongly Disagree-1 to 5-Strongly Agree). In the sample of 50 participants there were 51 missing data values of 1800 total (50 participants x 9 questions x 4 visualizations) (97.17 % complete of 1800 total). Furthermore, a single question (about zooming) had around 19 missing data values, which amounts to around 38 % of the 2.83 % missing data. This question was dropped for comparison between different visualizations. A detailed insight into the data samples collected reveals that the missing data values are spread over 8 question types and over 50 participants, and all four visualizations. For analysis purposes these were considered to occur randomly. Robertson (2001, p. 134) discusses various approaches to this problem of missing data, one of which is “*replacing missing values with the mean. Most statistical software packages have an option that allows missing values to be replaced with the mean of the remaining values*”. Excluding the questionnaire with maximum missing data values in total 1. 78 % (98.22 % complete data) were replaced by the mean response of each question type.

8.1.2 Participant profiles and sessions

Valuing varied perspectives, the participants involved in these studies were drawn from varied academic backgrounds and exposure to geographic information. Support for this is drawn from statistics regarding the use of paper maps, geo-web services and search engine usage. 50 % of the users involved in the studies had familiarity with the use of paper maps, while 72 % of them used search engines such as Google, Yahoo etc. daily. Furthermore, the statistics reveal that 36 % of them used geo-web services e.g. Google maps, www.map24.com, map.search.ch etc. In addition, statistics reveal that there has been no significant difference in responses between male (64 %) and female (36 %) participants for these studies. Table 8.1 details the statistics of the respondents involved in the SPIRIT experimental studies.

Table 8.1: Participants involved in SPIRIT experimental studies (Refer to CD:Descriptive Statistics)

Groups					
Group-I	Group-II	Group-III	Group-IV	Total	
15	11	10	14	50	
Sex					
Male	Female				
32	18				
Scenario Combination					
Simple interface	**Simple interface**	**Advanced Interface**	**Advanced Interface**		
Hotels near Portree	Mountaineering in Scotland	Biking in Scotland	Cottages in Western Scotland		
26	24	25	25		
Factorial Design					
ABCD	ACBD	ACDB	ADBC	ABDC	ADCB
8	11	7	8	8	8
Qualification					
Diploma student	Academic teaching staff	Librarian	PhD Student	Other(Please Specify)	Total
30	2	0	9	9	50

The experiments with the SPIRIT prototype system were organized in the order given in Table 8.1 (refer to Appendix: J - experiment design) drawing responses to a set of questions addressing a range of issues from participants' background, self-learning and two scenarios, one each for the simple interface and one for the advanced visualization interface (see Section 8.2). Finally, responses to a couple of design mock-ups were sought (Spatial Activity Plot and Isoline representations; see Section 8.3).

8.1.3 Time taken to complete tasks

The process of spatially aware information seeking being non-linear in nature and drawn from the experience of the early stage experiments, experiment design and pilot studies, 15 minutes were allocated for the completion of each section as it is non-trivial to document the amount of time spent for the specific tasks. While the Phase-II users have been requested to document the amount of time taken for completion of each section, the time taken by the participants for each episode on average was 15 minutes.

8.1.4 Self-learning exercise

In an effort to mobilise a broader understanding of the system functionalities, participants were involved in a self-learning exercise using any interface. Thus, learning by exploring. Participants were given a brief set of instructions detailing possible simple queries such as "hotels in Scotland" and "hotels near Glasgow". Users were instructed to initiate the self-learning exercise using the simple interface and then move on to learning advanced interface functionalities. Statistics reveal that 94 % of the participants either agree or strongly agreed that it had been easier to get started with the system and initiate an information request given the exercise (Table 8.2, response: 1). The success of the search interface was determined by the ease of usage and understanding; 88 % of the participants either agree or strongly agree (Table 8.2, response: 2).

Table 8.2: Self learning phase descriptive statistics for simple structured interface for a query "hotels in Scotland" and "hotels near Glasgow"

	N	Missing values	Range	Min	Max	Mean	Std. Deviation	Skewness
1. It was easy to get started with the system and make my query	50	0	3.0	2.0	5.0	4.46	0.67	-1.29
2. It was obvious what to type in the different boxes of the simple structured interface	50	0	3.0	2.0	5.0	4.38	0.75	-1.06
3. The list of spatial relationships provided were sufficient	50	0	3.0	2.0	5.0	4.04	0.78	-0.60
4.Did you find the interface interesting?	49	1	1.0	2.0	3.0	2.449	0.50	0.21
5. It was easy to find the locations of documents listed to the right of the map.	49	1	2.0	1.0	3.0	2.53	0.61	-0.958
6. The background mapping was sufficiently detailed.	50	0	3.0	2.0	5.0	3.28	0.96	-0.182
7. The basic map functionality (pan, zoom, etc.) was easy to use.	50	0	3.0	2.0	5.0	4.10	0.99	-0.885

SPIRIT in essence is a system to support the spatially aware information seeking process, which is acquiring information related to a specific geographic location. Early stage studies have revealed that often these geography related queries are expressed in the form of the triplets *"concept", "spatial relation", "geographic location"*. 80 % of the participants either agree or strongly agree that there has been a comprehensive list of geographic prepositions, either topological, proximal or directional that the system supports. A few notably illustrated the possibility of neighbourhood search and the combination of cardinal directions with distance, for example, *maximum 10 km north of Glasgow* (Table 8.2, response: 3).

44 % of the participants found the interface quite interesting, while 54 % found it "a little interesting" (Table 8.2, response: 4). This can be explained due to the simple nature of the starting interface. 58 % of the participants expressed that it was easy to find the document footprints on the map and their relations with the ranked result sets on the left, while 34 % expressed that it was "a

little easier" (Table 8.2, response: 5). Maps being the simplest form of representation, granularity and context serve to address the information request. 52 % of the participants either agree or strongly agree that the background information provided on the map was adequate enough to serve their purposes (Table 8.2, response: 6) and 76 % of the participants either agree or strongly agree that the map functionalities have been adequate enough to explore the required resolution or granularity (Table 8.2, response: 7).

8.2 Visualization Methods of SPIRIT Prototype

Prior to result analysis using repeated measure ANOVA discussed in the sub-sections below, an overall descriptive statistic has been carried out to draw holistic insights into data samples. Table 7.7 details the empirical measures such as: ease of use/usefulness, information presentation and understandability, role of supporting aides, colour usage and appropriateness, visual distinction (relevance), discrimination, viewpoint or relativity, granularity and context (zooming), locate/associate/compare and relevance as a measure to identify documents.

8.2.1 Repeated measures ANOVA

A detailed analysis has helped us to identify which of these empirical measures has yielded significantly different responses. Supporting insights were drawn with simple main effects analysis which is carried out and compared with the response for each empirical measure using General Linear Model Repeated Measures. Dykes (2005, p.657) states that "*the repeated measures ANOVA is a powerful way for accounting of user's attention, motivation, learning (Keppel, 1982) because it takes into account the fact that there is more than one observation per subject, which can be used to supply information on the main effect of each factor. ANOVA tests the null hypothesis of no differences between population variances (or mean squares) of the performance variable measured for each task*". Table 8.3 details the advantages and disadvantages of adopting Repeated Measures ANOVA.

Table: 8.3: Advantages and Disadvantages of repeated measures (Davis, 2002 p.2)

Advantages	Disadvantages
1. Individual patterns of change can be obtained using this method.	1. Analysis is complicated by the dependence among repeated observations made on the same experimental unit.
2. Economizes on subjects.	2. The investigator often cannot control the circumstances for obtaining measurements, so that the data may be unbalanced or partially incomplete.
3. The subjects can control the outcome variable can be measured under both control and experimental conditions for each subject.	3. It is restricted to the setting in which the response variable is normally distributed and to which the data is balanced and complete.
4. By excluding between subject sources of variability from the experimental error, repeated measurement designs often provide more efficient estimators of relevant parameters than cross-sectional designs with the same number and pattern of measurements.	4. The methodology is not fully developed for continuous, normally distributed outcomes but it has drawn significant interests in active areas of research.
5. Reliable data can be collected in which the same subjects are followed repeatedly rather than in a cross-sectional study.	5. The practical application for methods of repeated categorical outcome also lags behind the normal-theory methods due to the lack of readily accessible software.

A repeated measure ANOVA with a significant level of 0.05 was applied using SPSS's General Linear Model Analysis function. In addition to this, GLM Profile Plots (Profile or Interaction Plots) were plotted for comparing marginal means. Each of the visualization line profiles indicate the estimated marginal means of the dependent variable (adjusted for any covariates) at one level of a factor. This helps to show whether the estimated marginal means are increasing or decreasing for different visualization profiles. Figure 8.1 (§ref. p.188) shows the profiles of different visualization types and the empirical measures. While the parallel lines indicate a shift in the user's response. Likewise, Figure 8.2 demonstrates that there are significant interactions between empirical measures and the different visualization types.

Figure 8.1 (§ref. p.188) illustrates that there are significant interactions between different visualization profiles (4) for information presentation & understanding, colour appropriacy and visual distinction of relevance, viewpoint and relativity, and identification of relevant documents. There is a slight shift in the mean response between maps, cartograms and footprint based visualization in its usefulness. While, the mean response for density surface representation is fairly higher because of its ease of use and information presentation and understanding. Inferring that the density surface representation provides an abstract representation of the spread of documents over a selected window region described by the users. Each of these visualization methods provides a good way of looking at the results, which is distinct and serves different purposes, inferences drawn from Figure 8.1 (§ref. p.188) & 8.2 (§ref. p.189) show significant differences in empirical measures for viewpoint and relativity.

8.2.2 Simple Main Effects

The results from the repeated measures ANOVA procedure estimates all of the main and average effects and interactions (generated using SPSS version 14.0). The results are organized by experimental tasks introduced in Chapter 7 (Table: 7.7) and corresponding visualization methods adopted to accomplish them. Table: 8.4 displays a pair-wise comparison between experimental tasks and visualization methods adopted to accomplish them. Significant differences ($p<0.05$) have been identified for the 4 different visualization methods. In some cases. This significant result, indicates that the sphericity assumption has been violated. Sphericity assumption takes into account the variances of the variables are equal, which is not true in this case.

Table 8.4 (§ref. p.189) presents a pair-wise comparison for each experimental task and visualization methods using the general linear model repeated measures. Key findings of the statistics presented in Table 8.4 (§ref. p.189) reveal the following:

1. It is to be noted that, for the task: Ease of use (Usefulness) - significant differences have been identified between Density-Footprint and Density-Cartogram representations. Which is due to their functionality and purpose each of them serves to address. Density surface representation present an overview of the spread of the documents over a geographic region while the Footprint based representations present the scope/geographic extent of individual footprints within and between documents. Whereas for Cartogram they present the number of documents present in each adminstrative unit.
2. While for the task: Information presentation and understandability, there have been significant differences in the users response between Map-Footprint based representations, Density-Footprint, Density-Cartograms. This is due to the level of interactivity and explorability each of these interface provide to the end user.
3. For the task: legends and annotations as supporting aides- there has no significant differences in the users response, which is due to the lack of appropriate legends available to the users in the interface. Users had to dig through the HELP to learn about specifics of each colour schemes adopted for different visualization.
4. The usage of appropriate colour to demonstrate distinctiveness based on relevance has no significant difference in the users response, which could be due to the use of less divergent colour scheme. Having used a more divergent colour scheme from Colour Brewer

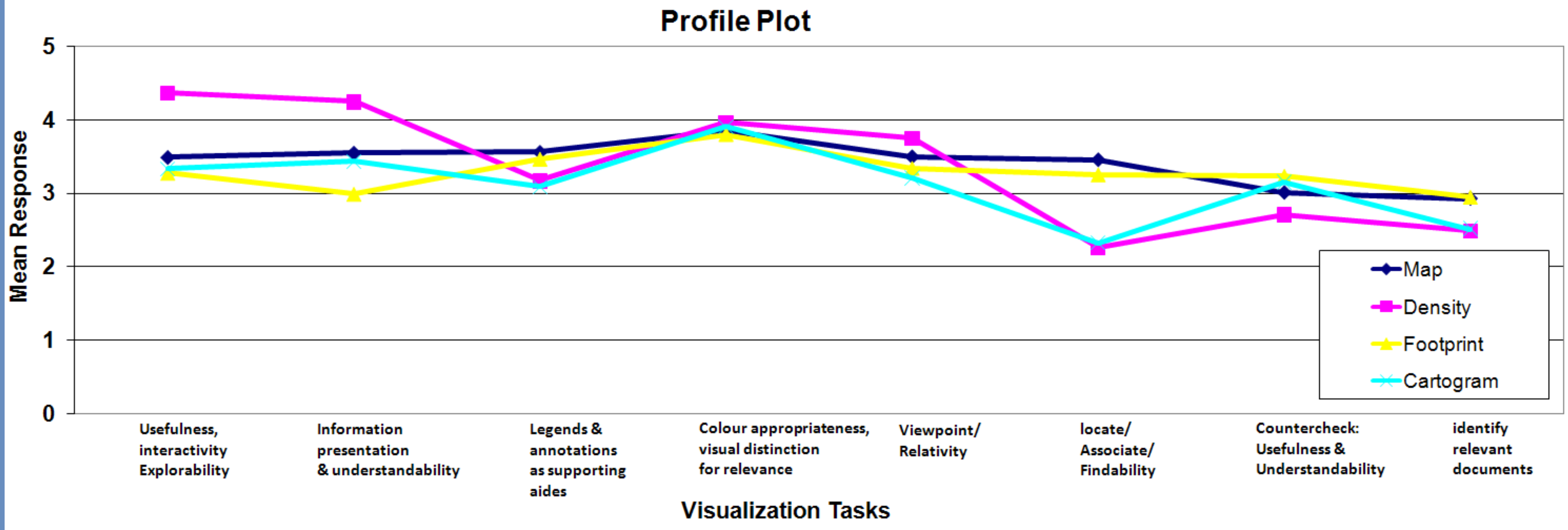


Figure 8.1: Interaction plot depicting the visualization profiles against question type and user response.(GLM mean response).

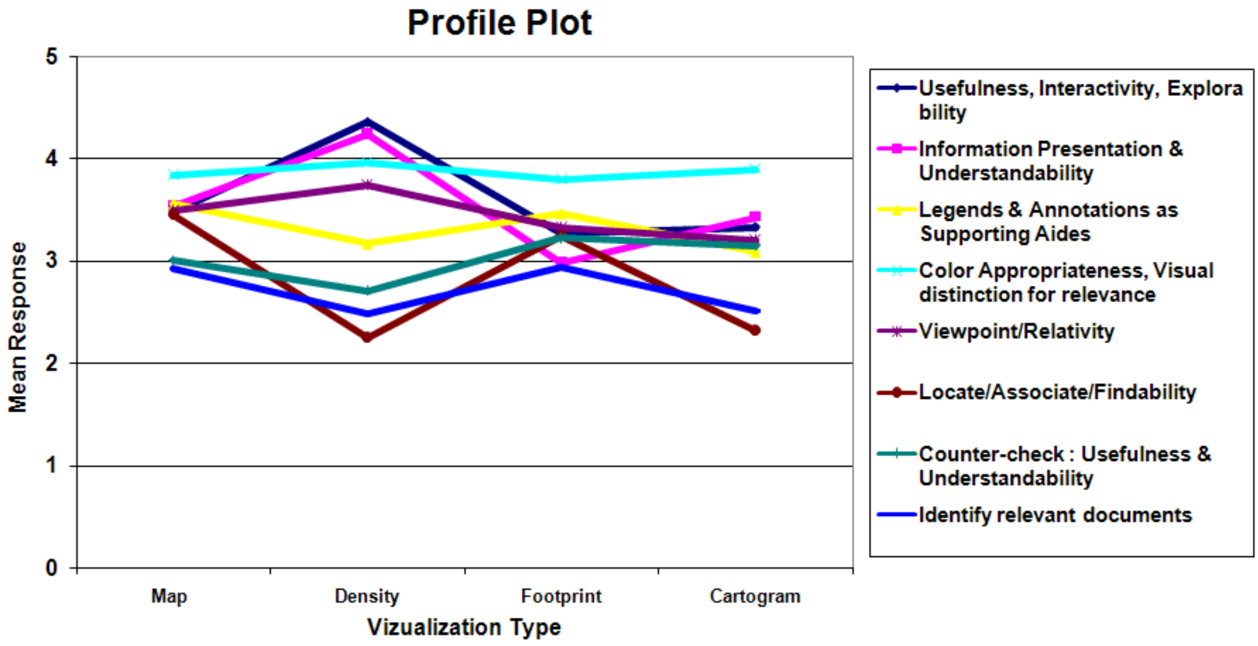


Figure 8.2: Interaction plot depicting the question type against visualization methods and user response (GLM mean response) for each empirical measure.

Table 8.4: Pair-wise comparison between experimental task * visualization methods using General Linear Model Repeated Measures.

(Experimental task Questions)	**(Visualization method) (I) display**	**(J) display**	**Mean Difference (I-J)**	**Std. Error**	**Sig.(a)**	**95% Confidence Interval for Difference (a)**	**95% Confidence Interval for Difference (a)**
						Lower bound	**Upper bound**
Ease of use (Usefulness)	Map	Density	-0.667	0.319	0.055	-1.35	0.017
		Footprint	0.333	0.374	0.388	-0.468	1.135
		Cartogram	0.333	0.422	0.442	-0.571	1.238
	Density	Map	0.667	0.319	0.055	-0.017	1.350
		Footprint	1.00(*)	0.402	0.026	0.137	1.863
		Cartogram	1.00(*)	0.324	0.008	0.306	1.694
	Footprint	Map	-0.333	0.374	0.388	-1.135	0.468
		Density	-1.00(*)	0.402	0.026	-1.863	-0.137
		Cartogram	0.000	0.378	1.00	-0.811	0.811
	Cartogram	Map	-0.333	0.422	0.442	-1.238	0.571
		Density	-1.00(*)	0.324	0.008	-1.694	-0.306
		Footprint	0.000	0.378	1.00	-0.811	0.811

Table 8.4: Pair-wise comparison between experimental task * visualization methods using General Linear Model Repeated Measures.

(Experimental task Questions)	(Visualization method) (I) display	(J) display	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference (a)	95% Confidence Interval for Difference (a)
						Lower bound	Upper bound
Information pesentation and understandability	Map	Density	-0.467	0.256	0.089	-1.015	0.082
		Footprint	0.733(*)	0.330	0.044	0.025	1.442
		Cartogram	0.133	0.256	0.610	-0.415	0.682
	Density	Map	0.467	0.256	0.089	-0.082	1.015
		Footprint	1.200(*)	0.380	0.007	0.384	2.016
		Cartogram	0.600(*)	0.273	0.045	0.015	1.185
	Footprint	Map	-0.733(*)	0.330	0.044	-1.442	-0.025
		Density	-1.200(*)	0.380	0.007	-2.016	-0.384
		Cartogram	-0.600	0.363	0.120	-1.378	0.178
	Cartogram	Map	-0.133	0.256	0.610	-0.682	0.415
		Density	-0.600(*)	0.273	0.045	-1.185	-0.15
		Footprint	0.600	0.363	0.120	-0.178	1.378
Legends and Annotations as supporting aides	Map	Density	0.467	0.336	0.187	-0.254	1.188
		Footprint	0.133	0.215	0.546	-0.328	0.595
		Cartogram	0.533	0.291	0.088	-0.090	1.157
	Density	Map	-0.467	0.336	0.187	-1.188	0.254
		Footprint	-0.333	0.374	0.388	-1.135	0.468
		Cartogram	0.400	0.289	0.189	-0.221	1.021
	Cartogram	Map	-0.533	0.291	0.088	-1.157	0.090
		Density	-0.067	0.228	0.774	-0.556	0.423
		Footprint	-0.400	0.289	0.189	-1.021	0.221

Table 8.4: Pair-wise comparison between experimental task * visualization methods using General Linear Model Repeated Measures.

(Experimental task Questions)	(Visualization method) (I) display	(J) display	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference (a)	95% Confidence Interval for Difference (a)
						Lower bound	Upper bound
Colour Appropriateness, visual distinction for relevance	Map	Density	-0.133	0.133	0.334	-0.419	0.153
		Footprint	-0.067	0.118	0.582	-0.320	0.187
		Cartogram	0.000	0.098	1.00	-0.209	0.209
	Density	Map	0.133	0.133	0.334	-0.153	0.419
		Footprint	0.067	0.118	0.582	-0.187	0.320
		Cartogram	0.133	0.091	0.164	-0.062	0.328
	Footprint	Map	0.067	0.118	0.582	-0.187	0.320
		Density	-0.067	0.118	0.582	-0.320	0.187
		Cartogram	0.067	0.118	0.582	-0.187	0.320
	Cartogram	Map	0.00	0-098	1.00	-0.209	0.209
		Density	-0.133	0.091	0.164	-0.328	0.062
		Footprint	-0.067	0.118	0.582	-0.320	0.187
Viewpoint or Relativity	Map	Density	-0.200	0.175	0.271	-0.574	0.174
		Footprint	0.133	0.274	0.634	-0.454	0.720
		Cartogram	0.400	0.321	0.233	-0.288	1.088
	Density	Map	0.200	0.175	0.271	-0.174	0.574
		Footprint	0.333	0.333	0.334	-0.382	1.048
		Cartogram	0.600	0.306	0.070	-0.055	1.255
	Footprint	Map	-0.133	0.274	0.634	-0.720	0.454
		Density	-0.333	0.333	0.334	-1.048	0.382
		Cartogram	0.267	0.371	0.484	-0.529	1.063
	Cartogram	Map	-0.400	0.321	0.233	-1.088	0.288
		Density	-0.600	0.306	0.070	-1.255	0.055
		Footprint	-0.267	0.371	0.484	-1.063	0.529

Table 8.4: Pair-wise comparison between experimental task * visualization methods using General Linear Model Repeated Measures.

(Experimental task Questions)	(Visualization method) (I) display	(J) display	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference (a)	95% Confidence Interval for Difference (a)
						Lower bound	Upper bound
Zooming (Granularity & Context)	Map	Density	0.600	0.349	0.108	-0.149	1.349
		Footprint	0.667	0.454	0.164	-0.308	1.641
		Cartogram	0.933(*)	0.396	0.034	0.084	1.783
	Density	Map	-0.600	0.349	0.108	-1.349	0.149
		Footprint	0.067	0.300	0.827	-0.577	0.711
		Cartogram	0.267	0.228	0.262	-0.223	0.756
	Cartogram	Map	-0.933(*)	0.396	0.034	-1.783	-0.084
		Density	-0.333	0.187	0.096	-0.734	0.067
		Footprint	-0.267	0.228	0.262	-0.756	0.223
Locate or Associate or Findability	Map	Density	1.00(*)	0.365	0.016	0.217	1.783
		Footprint	0.200	0.341	0.567	-0.531	00.931
		Cartogram	0.867(*)	0.274	0.007	0.280	1.454
	Density	Map	-1.00(*)	0.365	0.016	-1.783	-0.217
		Footprint	-0.800	0.500	0.132	-1.871	0.271
		Cartogram	-0.133	0.363	0.719	-0-913	0.646
	Footprint	Map	-0.200	0.341	0.567	-0.931	0.531
		Density	0.800	0.500	0.132	-0.271	1.871
		Cartogram	0.667(*)	0.270	0.027	0.087	1.246
	Cartogram	Map	0.867(*)	0.274	0.007	-1.454	-0.280
		Density	0.133	0.363	0.719	-0.646	0.913
		Footprint	-0.667(*)	0.270	0.027	-1,246	-0.087
Counter check for usefulness and understandability	Map	Density	0.067	0.371	0.860	-0.729	0.863
		Footprint	-0.133	0.291	0.653	-0.757	0.490
		Cartogram	-0.067	0.345	0.849	-0.806	0.672

Table 8.4: Pair-wise comparison between experimental task * visualization methods using General Linear Model Repeated Measures.

(Experimental task Questions)	(Visualization method) (I) display	(J) display	Mean Difference (I-J)	Std. Error	Sig.(a)	95% Confidence Interval for Difference (a)	95% Confidence Interval for Difference (a)
						Lower bound	Upper bound
	Density	Map	-0.67	0.371	0.860	-0.863	0.729
		Footprint	-0.200	0.262	0.458	-0.762	0.362
		Cartogram	-0.133	0.307	0.670	-0.791	0.524
	Footprint	Map	0.133	0.291	0.653	-0.490	0.757
		Density	0.200	0.262	0.458	-0.362	0.762
		Cartogram	0.067	0.358	0.855	-0.701	0.835
	Cartogram	Map	0.067	0.345	0.849	-0.672	0.806
		Density	0.133	0.307	0.670	-0.524	0.791
		Footprint	-0.067	0.358	0.855	-0.835	0.701
Relevant task accomplishment	Map	Density	0.333	0.333	0.334	-0.382	1.048
		Footprint	0.333	0.232	0.173	-0.165	0.832
		Cartogram	0.400	0.423	0.36	-0.508	1.308
	Density	Map	-0.333	0.333	0.334	-1.048	0.382
		Footprint	0.00	0.324	1.00	-0.694	0.694
		Cartogram	0.067	0.300	0.827	-0.577	0.711
	Footprint	Map	-0.333	0.232	0.173	-0.832	0.165
		Density	0.00	0.324	1.00	-0.694	0.694
		Cartogram	0.067	0.384	0.865	-0.757	0.890
	Cartogram	Map	-0.400	0.423	0.361	-1.308	0.508
		Density	-0.067	0.300	0.827	-0.711	0.577
		Footprint	-0.067	0.384	0.865	-0.890	0.757

Based on estimated marginal means

** The mean difference is significant at the 0.05 level.*

a Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

might have thrown in significant difference in users response.

5. There has been no significant difference in users response for viewpoint or relativity throwing multiple perspectives serving different purposes.
6. Zooming (Granularity and Context) statistics reveal that there have been significant differences in the users responses for Map-Cartogram, which is due to the constant changing granularity and context as the user moves through various zoom levels using Map based representations, while Cartogram provide only an overview of the frequency of the document footprints within an administrative unit.

7. Significant differences(p<0.05) have been identified for the tasks to locate or associate or find footprints between and within individual documents for Map-Density, Map-Cartogram and Footprint-Cartogram, based representations. This is primarily due to the level of interactivity each of these visualizations facilitate to the users. Map and Footprint based representations provide fairly higher level of interactivity in comparison to Density and Cartogram based representations.
8. Finally, there has been no significant differences (p<0.05) for counter-check for usefulness and understandability and for relevant task accomplishment.

8.2.2.1 Within subjects

a) Ease of Use

One of the prominent measures for any visualization is the ease of use that indicates numerous other parameters such as ease of interaction, understanding and explorability of visualization methods presented to the users.

The results signify that of the four different visualizations (Map, Density, Footprint and Cartogram), respondents found that the Density surface representation is easier to use. 86% of the respondents either strongly agree or agree that density surface representations are easier to use in comparison to other visualizations as shown in Table 8.5. Supports for these inferences are drawn from the higher mean and significance in comparison to other visualizations (Map, Cartogram and Footprint). Table: 8.4 and Appendix- CD -Simple Main effects (refer to enclosed CD- Simplemaineffects.pdf file) illustrate the mean and significance of response against various visualizations.

Table 8.5: Empirical criteria-ease of use vs different visualizations (General Linear Model- Repeated Measures- WithinSubjects)

Question The visualization was easy to use ?	Mean	Std.Deviation	N
Graphical Map based representation (2 %- Strongly Disagree, 14%-Disagree, 26%- Neutral, 50% - Agree)	3.48	0.90891	50
Density surface representation (2%-Disagree, 12%- Neutral,34%-Strongly Agree, 52%- Agree)	4.36	0.77618	50
Footprint based representation (4%- Strongly Disagree, 30%-Disagree, 16%- Neutral, 44%- Agree)	3.2654	1.06475	50
Cartogram representation (6%- Strongly Disagree, 16%-Disagree, 26%- Neutral, 40%-Agree, 10%-Strongly Agree)	3.3266	1.05729	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

b) Information presentation and understandability

Participants responded that it was easier to understand information presented in density surface representation in comparison to other visualizations (Map, Cartogram and Footprint). Statistics reveal that 84% of the respondents either strongly agree or agree that information presented using density surface representations is effective as the representations provides a summary of the density of documents for specific geographic regions in the map view (Detailed statistics are shown in Table 8.6)

Table 8.6: Empirical criteria -information representation and understandability for different visualizations (General Linear Model- Repeated Measures- Within subjects)

Question: It was easy to understand the information presented by this interface ?	Mean	Std.Deviation	N
Graphical Map based representation (2 %- Strongly Disagree, 8%-Disagree, 30%- Neutral, 54% - Agree, 6%- Strongly Agree)	3.354	0.81341	50
Density surface representation (4%-Disagree, 12%- Neutral,44%-Strongly Agree, 40%- Agree)	4.24	0.82214	50
Footprint based representation (6%- Strongly Disagree, 30%-Disagree, 32%- Neutral, 20%- Agree, 10%- Strongly Agree)	2.9796	1.07835	50
Cartogram representation (2%- Strongly Disagree, 20%-Disagree, 18%- Neutral,50%-Agree, 8%-Strongly Agree)	3.4286	0.96890	50
Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question			

c) Legends and Annotations as supporting aides

There has been relatively little significant difference in the user's response with respect to the usage of legends and annotations as supporting aides for inferring information presented in different visualizations. It should be noted that the legend and annotation details have been presented with respect to the map based representations in the HELP/Tutorial provided to the users. Based on the observations during the experiments participants when confronted to decipher information within a specific visualization had spent an ample amount of time going through the HELP/Tutorial handbook provided. While users in the supervised sessions are trained with respect to the specifics of each of visualizations. It should be noted that there has been no legend and annotation provided for different visualizations namely Density, Footprint and Cartogram representations. However, its impact on the understandability and ease of use can be helpful for map based representations by analysing the results for supervised sessions (as shown in Table 8.7). In total 56% of the participants either strongly agree or agree that legends and annotations have been useful as supporting aids for understanding the information presented in Map based representations.

Table 8.7: Empirical criteria- Legends and Annotation as supporting aides vs Different visualization. (General Linear Model- Repeated Measures- Within Subjects)

Question The use of legends and annotations made it easier to understand the visualization ?	Mean	Std.Deviation	N
Graphical Map based representation (10%-Disagree, 32%- Neutral 48%- Agree, 8% - Strongly Agree)	3.5510	0.78406	50
Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question			

Table 8.7: Empirical criteria- Legends and Annotation as supporting aides vs Different visualization. (General Linear Model- Repeated Measures- Within Subjects)

Question The use of legends and annotations made it easier to understand the visualization ?	Mean	Std.Deviation	N
Density surface representation (8%- Strongly Disagree, 12%-Disagree, 42%- Neutral,28%- Agree, 8%- Strongly Agree)	3.1632	1.01714	50
Footprint based representation (2%- Strongly Disagree, 14%-Disagree, 26%- Neutral 46%- Agree)	3.4584	0.90257	50
Cartogram representation (6%- Strongly Disagree, 20%-Disagree, 32%- Neutral 36%- Agree, 2%- Strongly Agree)	3.0832	0.94401	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

d) Colour usage and Appropriateness

Colour scheming has been a medium for visual depictions to show relevance. The documents with higher relevance are visualized using bluish grey, while documents with lower relevance are visualized using a lighter grey colour. Similarly, visual distinction of the footprints or documents highlighted by the users is achieved by colour coding it as red. Of the four different visualizations, density surface representation reveals relatively higher response value with respect to other visualizations. Demonstrating the regions with higher density or peak colour coded with bluish grey and valley colour coded with light grey. 78% of the participants either strongly agree or agree that the colour usage was appropriate in Density Surface representation. These bands are constructed using the maximum and minimum density of the documents over a specific geographic region. Table 8.8 details the significant responses for different visualizations.

Table 8.8: Empirical criteria - Colour usage and appropriateness for different visualizations (General Linear Model- Repeated Measures - Within subjects)

Question: The use of colours was useful and appropriate ?	Mean	Std.Deviation	N
Graphical Map based representation(8%-Disagree, 14%- Neutral,64%- agree, 14% - Strongly Agree)	3.8400	0.76559	50
Density surface representation (4%-Disagree, 16%-Neutral, 60% Agree, 20% Strongly Agree)	3.9600	0.72731	50
Footprint based representation (2%- Strongly Disagree, 4%-Disagree, 24%- Neutral,50%- Agree, 18%-Strongly Agree)	3.796	0.8566	50
Cartogram representation (4%-Disagree, 16%-Neutral, 64% Agree, 14%- Strongly Agree)	3.898	0.67748	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

e) Viewpoint/Relativity

From abstractness to analysing the coverage & extent of result sets obtained from the SPIRIT system, these varied perspectives serve to impart specific purposes. Respondents expressed the density surface representations facilitate a relatively better approach in visualizing the results in comparison to other visualizations (Map, Footprint and Cartograms). The density surface representation allows users to obtain an overview of the spread of the documents across a geographic region of interest, while the Footprint based representation helps users to identify geographic extents of each of the footprints & individual documents by interactively filtering and highlighting. On the other hand, cartograms are spatial transformations of the geographic regions to depict the frequency of the documents in that specific region arranged by topology. In total 68% of participants either strongly agree or agree that density surface representation facilitates a good approach to look at the results. While there have been varying responses for different visualizations as illustrated in Table 8.9.

Table 8.9: Empirical criteria - viewpoint or relativity for different visualizations (General Linear Model- Repeated Measures - Withinsubjects)

Question: This visualization provided a good way of looking at the results ?	Mean	Std.Deviation	N
Graphical Map based representation.(2%- Strongly Disagree, 10%-Disagree, 30%- Neutral,50%- Agree, 6% - Strongly Agree)	3.498	0.83599	50
Density surface representation (2%- Strongly Disagree, 8%-Disagree, 22%- Neutral, 50%- Agree, 18% - Strongly Agree)	3.7400	0.92162	50
Footprint based representation (4%- Strongly Disagree, 18%-Disagree, 28%- Neutral 34%- Agree, 12% - Strongly Agree)	3.332	1.03674	50
Cartogram representation (2%- Strongly Disagree, 22%-Disagree, 32%- Neutral, 38%- Agree, 4% - Strongly Agree)	3.2040	0.90305	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

f) Locate/Associate/Findability/Compare

Of the many tasks that the visualization supports one essential task is the ability to identify footprints and documents with higher relevance and their associations with geographic locations. Map based representations fare relatively well (48% either agree or strongly agree) in comparison to different visualization as shown in Table 8.10. A relatively closer response is obtained for Footprint based representations as they facilitate user's with the ability to interactively filter and highlight geographic extents of footprints within the documents or individual documents. It is natural that the Density and Cartogram representations serve a different purpose such as spread of the documents density over a geographic space or documents frequency for a political administrative region

Table 8.10: Empirical criteria - locate/associate/findability for different visualizations (General Linear Model- Repeated Measures - Withinsubjects)

Question: It was easy to find documents which shared the same location ?	Mean	Std.Deviation	N
Graphical Map based representation.(2%- Strongly Disagree, 16%-Disagree, 32%- neutral, 32% - Agree, 16%- Strongly Agree)	3.449	1.01139	50
Density surface representation (22%- Strongly Disagree, 42%-Disagree, 20%- Neutral, 10%- Agree, 2%- Strongly Agree)	2.25	0.97938	50
Footprint based representation (4%- Strongly Disagree, 24%-Disagree, 22%- Neutral, 40%- Agree, 8% - Strongly Agree)	3.2448	1.04062	50
Cartogram representation (14%- Strongly Disagree, 52%-Disagree, 14%- Neutral, 12%- Agree, 2%- Strongly Agree)	2.3192	0.92816	50
Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question			

g) Counter check - Ease of Use and Understandability

In an effort to validate user responses for being user friendly some analysis was done. Statistics reveal that 40% of the participants either agree or strongly agree that the footprint based representation is relatively difficult to understand followed by cartograms, as they (cartograms) do not facilitate users with the details with respect to the individual documents. Table 8.11 presents statistics of different visualizations.

Table 8.11: Empirical criteria - counter check for ease of use and understandability for different visualizations (General Linear Model- Repeated Measures - Withinsubjects)

Question: It was easy to get confused when using this interface ?	Mean	Std.Deviation	N
Graphical Map based representation.(34%-Disagree, 36%- Neutral, 26%- Agree, 4%- Strongly agree)	3.00	0.88063	50
Density surface representation (6%- Strongly Disagree, 42%-Disagree, 34%- Neutral, 12%- Agree, 6%- Strongly Agree)	2.70	0.97416	50
Footprint based representation(26%- Disagree, 32%- Neutral, 32%- Agree, 8% - Strongly Agree)	3.2244	0.93165	50
Cartogram representation (22%-Disagree, 48%- Neutral, 20%- Agree, 8%- Strongly Agree)	3.1428	0.85714	50
Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question			

h) Relevance/Rank/Findability

One of the primary objectives of the spatially aware information seeking process is to identify the most relevant information with respect to a geographic location. Footprint based representation fairs relatively well in comparison with other visualizations (Map, Cartogram and Density) as is shown in Table 8.12. There is a significant difference between Footprint and Map based representations, in that they facilitate users to identify individual documents as well as footprints. In total 30% of the participants either agree or strongly agree, while 36% disagree in using Footprint representation, to easily identify the most relevant documents to accomplish their tasks. The mean response values of the map and footprint based representations are relatively higher than those of Density and Cartogram representation since the former helps users to identify, rank and explore the geographic footprints and extents of individual documents, While the latter provides an abstraction or a summary of density or frequency of documents with respect to a geographic region or a spatial unit.

Table 8.12: Empirical criteria - relevant task accomplishment for different visualizations (General Linear Model- Repeated Measures - Withinsubjects)

Question: This visualization made it easy to find the most relevant documents for my task ?	Mean	Std.Deviation	N
Graphical Map based representation.(8%- Strongly Disagree, 34%- Disagree, 20%- Neutral, 30%- Agree, 6%- Strongly Agree)	2.9184	1.10355	50
Density surface representation (22%- Strongly Disagree, 32%- Disagree, 24%- Neutral, 20%- Agree, 2%- Strongly Agree)	2.480	1.11098	50
Footprint based representation (2%- Strongly Disagree, 36%- Disagree, 28%- Neutral, 26%- Agree, 4%- Strongly Agree)	2.9376	0.93473	50
Cartogram representation (14%- Strongly Disagree, 40%- Disagree, 20%- Neutral, 18%- Agree, 2%- Strongly Agree)	2.5106	1.00757	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

Each of the visualizations presented to the users serves different purposes. 86% of the participants either strongly agree or agree that they were able to identify regions with highest document density using Density surface representation. When using footprint based representations 70%,62% and 64% of the participants either strongly agree or agree that they had been successful in identifying the geographic extent of the 1st, 2nd and 3rd ranked documents from the result set obtained from the SPIRIT engine. This usually is not the case, with cartogram representations as they demonstrate the frequency of the documents in a political administrative district or boundary, only 32%, 30% and 28% of the participants either strongly agree or agree that they have been successful in identifying the regions using cartograms. Table 8.13 presents detailed descriptive statistics of different visualizations to locate the geographic scope/extent of the documents using different visualization.

8.2.2.2 Between Subjects

i) Unsupervised and Supervised

The experiments were organized in two phases, One being unsupervised and the other supervised sessions, where in the former case participants were provided with a tutorial to train them through the different visualization methods; while in the latter case participants had undergone a training session for about 30 minutes detailing the functionalities and features of different visualization approaches used. Also, it should be noted that in the self-learning exercise users were asked to explore and gain a preliminary understanding of the SPIRIT prototype system, whereas in the unsupervised sessions users were provided with a TUTO-

Table 8.13: Descriptive statistics of different visualizations to locate geographic scope of the documents

	Density surface representation	Cartogram representation - It Was easy to find the regions of the following documents ?			Footprint based representation - It was easy to find the regions of the following documents ?		
	Identify the regions with highest document density	Ist ranked document	2nd ranked document	3rd ranked document	1st ranked document	2nd ranked document	3rd ranked document
N Valid Missing	48 2	47 3	47 3	47 3	47 3	47 3	47 3
Mean	4.2917	2.6809	2.5957	2.4681	3.9574	3.6383	3.7872
Std.Error of Mean	0.11894	0.22782	0.21648	0.20818	0.16367	0.18854	0.16361
Median	4.000	2.000	2.00	2.00	4.00	4.00	4.00
Mode	5.0	1.00	1.00(a)	1.00(a)	5.00	4.00	4.00
Std.Deviation	0.82406	1.56186	1.48411	1.42724	1.12206	1.29255	1.12165
Skewness	-1.314	0.381	0.493	0.606	-1.070	-0.73	-0.814
Range	3.00	4.0	4.00	4.00	4.00	4.00	4.0
Min	2.00	1.00	1.00	1.00	1.00	1.00	1.0
Max	5.00	5.00	5.00	5.00	5.00	5.00	5.0

a Multiple modes exist. The smallest value is shown. Missing values are neglected. The above table presents descriptive statistics of Density, Cartogram and Footprint based representations. Coded using a Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question.

RIAL to help them learn the functionalities of the SPIRIT system.

Figure: 8.3 & Fig: 8.4 demonstrates that there has been significant interactions between different visualizations during both the phases. It should be noted that the density surface representation has a significantly high mean response in comparison with other visual representations in its usefulness, information presentation and cognition. These figures illustrate a slight shift in the users' response and reveal little significant difference associated to the ability to comprehend of specific visualizations.

a) Ease of Use

Investigating the impact of learning and significant shifts in the user response, footprint and cartogram based representations have little significant differences in the mean responses (mean diff>0.5) as shown in Table 8.14. The footprint based representation involve a higher level of interactivity and facilitate the users to selectively filter documents based on their geographic extents. As the notion of geographic extent requires considerable user understanding, it is evident that there have been slight significant shifts in the user's responses after the supervised training session. Similarly using cartogram based representations involve a higher degree of understanding to correlate spatial transformation with the ranked results and their association with highlighted geographic regions on the map.

Table 8.14: Mean responses for ease of use using GLM (Between Subjects)

Question: The visualization was easy to use ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.500	0.94868	26
	2. Supervised	3.54	0.93153	24
	Total	3.52	0.9311	50
Density surface representation	1. Unsupervised	4.30	0.8375	26
	2. Supervised	4.41	0.717	24
	Total	4.36	0.77618	50
Footprint based representation	1. Unsupervised	3.0104	1.05962	26
	2. Supervised	3.5417	1.02062	24
	Total	3.2654	1.06475	50
Cartogram	1. Unsupervised	3.0385	1.14824	26
	2. Supervised	3.6388	0.8677	24
	Total	3.326	1.05729	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

b) Information Presentation and Understandability

Table 8.15 details that there has been a significant shift in the user's response for the footprint based representation (mean difference >0.74), followed by cartogram based representation. The exploratory nature of footprint based representations demonstrate the geographic extents of the footprints between and within documents, and the correlation and relevance involved to the features of selective filtering. This constant shift in the user's response can be attributed to the supervised training session detailing key functionalities involving footprint based representation. Cartogram based representations accommodate the frequency of the result sets within a specific political or administrative boundary.

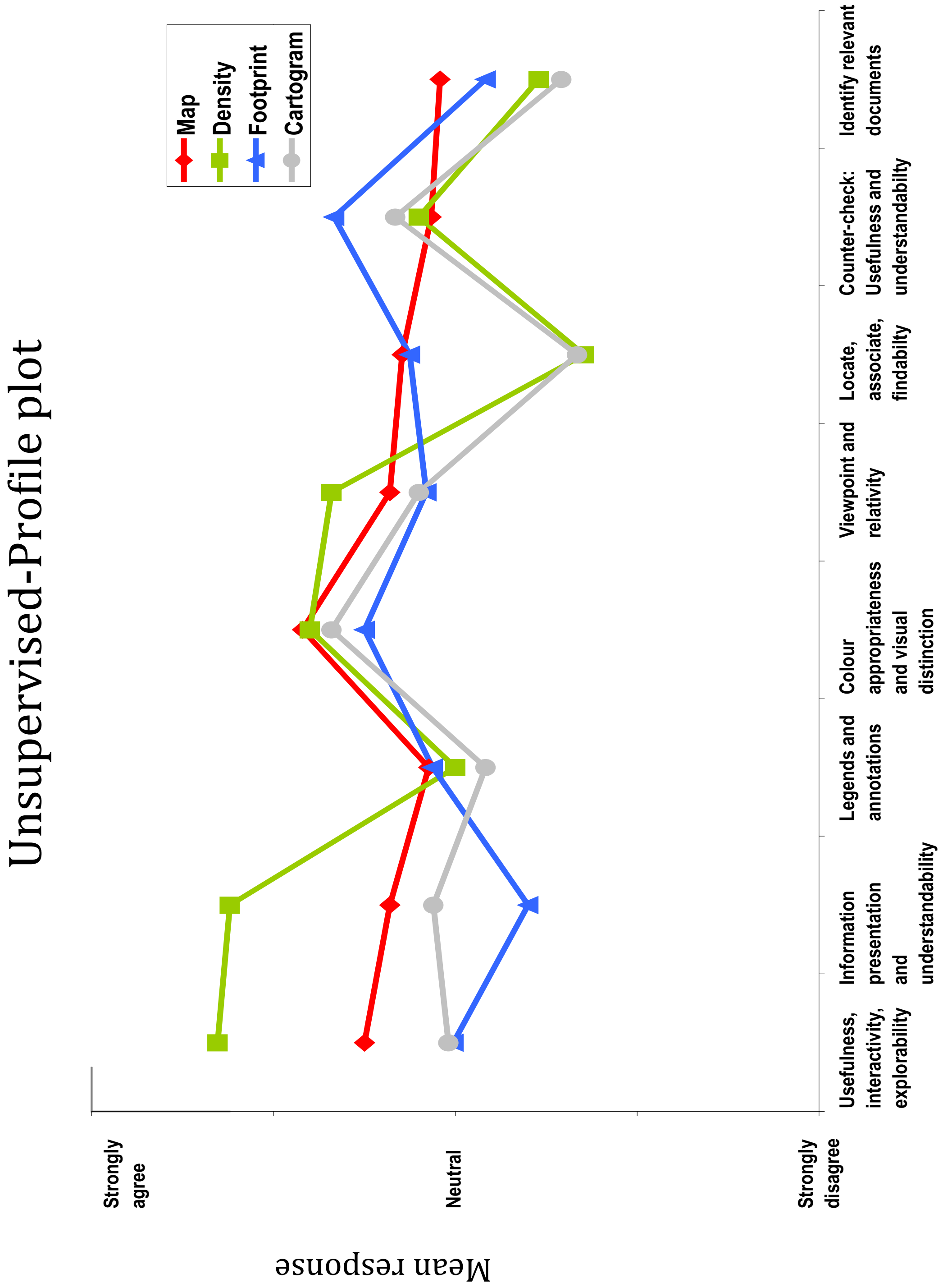


Figure 8.3: Interaction plot for mean response between visualization tasks for unsupervised session.

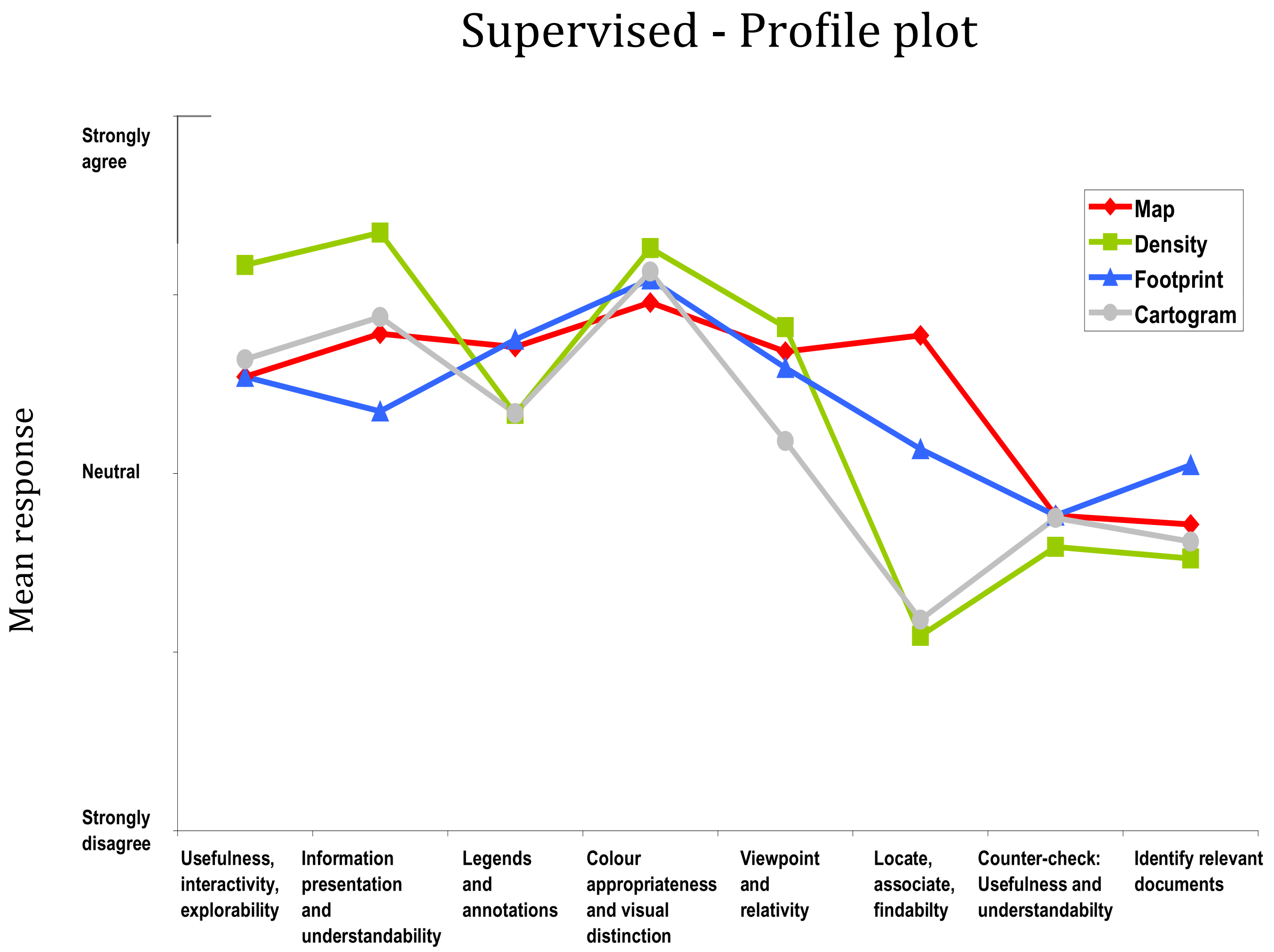


Figure 8.4: Interaction plot for mean response between visualization tasks for supervised session

Table 8.15: Mean responses for information presentation and usage with GLM (Between Subjects)

Question: It was easy to understand the information presented by this interface ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.36	0.70	25
	2. Supervised	3.7826	0.735	23
	Total	3.56	0.74	48
Density surface representation	1. Unsupervised	4.24	0.879	25
	2. Supervised	4.34	0.714	23
	Total	4.29	0.79	48
Footprint based representation	1. Unsupervised	2.60	1.0	25
	2. Supervised	3.34	1.070	23
	Total	2.95	1.090	48
Cartogram	1. Unsupervised	3.12	1.053	25
	2. Supervised	3.82	0.716	23
	Total	3.458	0.966	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

c) Legends and Annotation as supporting aides

A shift in the mean response (mean difference >0.5) of the users has been documented in the footprint and cartogram based representation for supporting aides such as legends and annotations that have been useful in inferring key functionalities of various representations. Table 8.16 details the mean responses for legends as supporting aides using GLM.

Table 8.16: Mean responses for legends as supporting aides using GLM (Between Subjects)

Question: The use of legends and annotatins made it easier to understand the visualization ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.4167	0.77553	24
	2. Supervised	3.7083	0.80645	24
	Total	3.5625	0.79643	48
Density surface representation	1. Unsupervised	3.00	1.02151	24
	2. Supervised	3.333	1.0495	24
	Total	3.1667	1.03827	48
Footprint based representation	1. Unsupervised	3.125	0.99181	24

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

Table 8.16: Mean responses for legends as supporting aides using GLM (Between Subjects)

Question: The use of legends and annotatins made it easier to understand the visualization ?	Test Group	Mean	Std. Deviation	N
	2. Supervised	3.75	0.794	24
	Total	3.4375	0.9432	48
Cartogram	1. Unsupervised	2.8333	1.0495	24
	2. Supervised	3.3367	0.81524	24
	Total	3.085	0.96381	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

d) Colour usage and appropriateness

Colour has been a surrogate of relevance to draw in visual distinctions between related documents and footprints. A significant shift in the mean response (mean diff > 0.5) is documented in the density surface representation (mean diff 0.46 <0.5), footprint and cartogram based representation. Bluish grey signifies documents or footprints with relatively high significance, while documents with light grey colour signify the least significance based on combined relevance. Table 8.17 details the mean responses for colour usage and appropriateness for different visualizations using GLM.

Table 8.17: Mean responses for colour usage and appropriateness for different visualization using GLM (Between Subjects)

Question: The use of colours was useful and appropriate ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.840	0.687	25
	2. Supervised	3.9565	0.76742	23
	Total	3.8958	0.72169	48
Density surface representation	1. Unsupervised	3.80	0.8165	25
	2. Supervised	4.2609	0.54082	23
	Total	4.0208	0.72902	48
Footprint based representation	1. Unsupervised	3.56	0.91652	25
	2. Supervised	4.087	0.73318	23
	Total	3.8125	0.86679	48
Cartogram	1. Unsupervised	3.68	0.74833	25
	2. Supervised	4.1304	0.54808	23
	Total	3.8958	0.69158	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

e) Viewpoint/Relativity

It was noted that there was no significant difference (mean diff >0.5) in the mean responses between different phases (as shown in Table 8.18) for view point and relativity.

Table 8.18: Mean responses for viewpoint or relativity for different visualization using GLM (Between Subjects)

Question: This visualization provided a good way of looking at the results ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.36	0.90738	25
	2. Supervised	3.6818	0.77989	22
	Total	3.51	0.856	47
Density surface representation	1. Unsupervised	3.68	0.90	25
	2. Supervised	3.818	1.00647	22
	Total	3.7447	0.94335	47
Footprint based representation	1. Unsupervised	3.16	0.98658	25
	2. Supervised	3.5909	1.09801	22
	Total	3.3617	1.05141	47
Cartogram	1. Unsupervised	3.20	0.95743	25
	2. Supervised	3.1818	0.90692	22
	Total	3.1915	0.92403	47

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

f) Countercheck-ease of use and understandability

It was noted that there was a significant difference (mean diff >0.5) in the users' response to the visualizations have been effective and also to them giving a comprehensive understanding in being user friendly. (as shown in Table 8.19)

Table 8.19: Mean responses for counterchecks for ease of use and understandability for different visualization using GLM (Between Subjects)

Question: It was easy to get confused when using this interface ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.133	1.12546	15
	2. Supervised	2.7647	1.09141	17
	Total	2.0375	1.10534	32
Density surface representation	1. Unsupervised	3.2	1.20712	15
	2. Supervised	2.5882	0.87026	17
	Total	2.8750	1.07012	32
Footprint based representation	1. Unsupervised	3.667	1.11270	15
	2. Supervised	2.7647	0.90342	17
	Total	3.1875	0.90342	17

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

Table 8.19: Mean responses for counterchecks for ease of use and understandability for different visualization using GLM (Between Subjects)

Question: It was easy to get confused when using this interface ?	Test Group	Mean	Std. Deviation	N
Cartogram	1. Unsupervised	3.333	1.04654	15
	2. Supervised	2.7059	0.58787	17
	Total	3.0	0.87988	32

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

g) Locate/Associate/Findability

There was no significant difference (mean diff >0.5) in the users' response between the two phases (as shown in Table 8.20) for finding documents that share the same location, which are either represented as stacks or illustrated as graphs.

Table 8.20: Mean responses for findability of documents sharing the same location using GLM (Between Subjects)

Question: It was easy to find documents which shared the same location ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.2917	0.80645	24
	2. Supervised	3.7727	1.06600	22
	Total	3.5217	0.96007	46
Density surface representation	1. Unsupervised	2.2917	0.85867	24
	2. Supervised	2.0909	1.18088	22
	Total	2.1957	0.98024	46
Footprint based representation	1. Unsupervised	3.250	0.9890	24
	2. Supervised	3.1364	1.12527	22
	Total	3.1957	1.046	46
Cartogram	1. Unsupervised	2.333	0.96309	24
	2. Supervised	2.1818	0.90692	22
	Total	2.2609	0.92939	46

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

h) Identify relevant documents

There was no significant shift (mean diff >0.5) in users' response with respect to the identification of relevant documents (ref. Table 8.21) in different phases of supervises and unsupervised sessions.

Table 8.21: Mean responses for effectiveness of visualization to aid users in identifying most relevant documents using GLM (Between Subjects)

Question: This visualization made it easy to find the most relevant documents for my task ?	Test Group	Mean	Std. Deviation	N
Graphical map based representation	1. Unsupervised	3.0833	1.01795	24
	2. Supervised	2.7143	1.18924	21
	Total	2.911	1.10417	45
Density surface representation	1. Unsupervised	2.5417	1.14129	24
	2. Supervised	2.5238	1.167	21
	Total	2.533	1.14018	45
Footprint based representation	1. Unsupervised	2.8333	0.86811	24
	2. Supervised	3.047	1.07127	21
	Total	2.933	0.96595	45
Cartogram	1. Unsupervised	2.4167	1.21285	24
	2. Supervised	2.619	0.864	21
	Total	2.511	1.05792	45

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

ii) Spatial qualifiers

To investigate the effects of different scenarios presented to the users, which impart varying levels of granularity, context and scope of results obtained from the SPIRIT system. Between subjects Repeated Measures ANOVA is carried out after a detailed descriptive statistic reveals a significant difference for different visualization tasks.

a) Simple Interface

After analysing the users' response as to whether there was a significant difference in their responses based on the scenarios presented to them two scenarios were constructed serving different regions with varying granularity, context and indexed documents.

Figure 8.5 (shown in p.210) illustrates the interaction plot between mean user's response and visualization tasks for the two different scenarios. Statistics reveal that there was been a significant shift in the users response.

b) Advanced Visualization

In contrast to the simple structured interface, the advanced visualization interface facilitated users to illustrate or describe a region of interest by drawing a polygon on the map. An interaction plot was constructed for various visualization profiles for each of the tasks and mean responses of the users. The approach adopted for expressing the region of interest is by drawing a polygon on the map, which implied that not all users exactly define the same region of interest. Thus, the region of interest, context and results obtained for each of them might vary. However, our objective was to investigate what impact it had on the users' response using different visualization methods, which remain constant irrespective of the scenario presented to them.

Figure 8.6 & 8.7 (shown in p. 212 & 213) present an interaction plot for mean response between visualization tasks comparing map, density, footprint and cartogram representation for queries such as " cottages in western Scotland" and "biking south of London". Statistics reveal that there has been a relatively higher interaction for the scenarios "cottages in western Scotland" in comparison with "biking south of London", where there has been a significant shift in the users' response. Further, investigation is carried out by using GLM repeated measures for each of the criteria.

i) Ease of use

With varying levels of granularity, context and results obtained for each of the scenarios, the impact of usefulness of the visualization for various scenarios was investigated. Statistics reveal significant shifts for various visualizations. Map (mean diff = 0.4), Density (mean diff = 0.4), Footprint (mean diff = 0.3), Cartogram (mean diff = 0.2). Table 8.22 details mean response for usefulness of different visualizations.

Table 8.22: Mean response for usefulness of different visualizations (Spatial Qualifiers).

Question: the visualisation was easy to use ?	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.32	0.90	25
	2.0 Cottages in western Scotland.	3.72	0.936	25
	Total	3.52	0.9311	50
2. Density surface representation	1.0 Biking south of London.	4.12	0.92736	25
	2.0 Cottages in western Scotland.	4.60	0.50	25
	Total	4.36	0.77618	50
3. Footprint based representation	1.0 Biking south of London.	3.1308	1.1299	25
	2.0 Cottages in western Scotland.	3.40	1.00	25
	Total	3.265	1.06475	50
4.Cartogram	1.0 Biking south of London.	3.2532	1.12743	25
	2.0 Cottages in western Scotland.	3.40	1.00	25
	Total	3.326	1.05729	50

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

ii) Information presentation and understandability

Statistical analysis reveals that there has been negligible difference in the mean responses for the information presented and understanding of the visualization (as shown in Table 8.23)

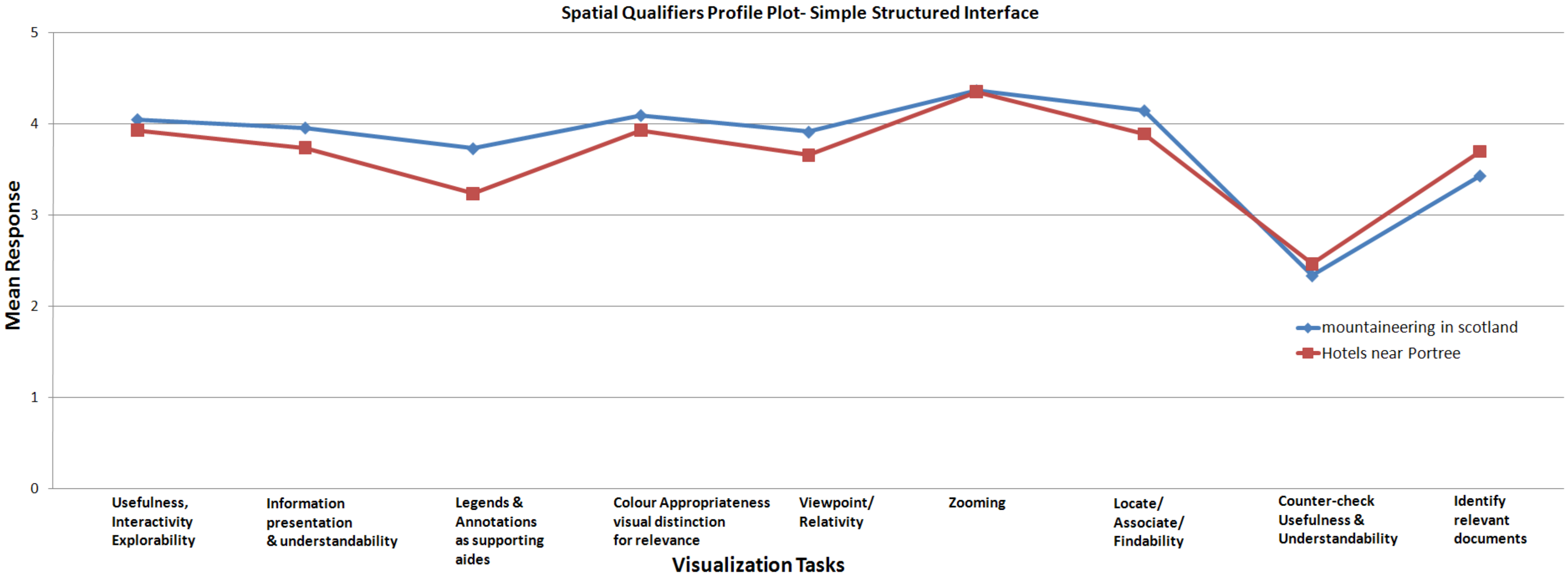


Figure 8.5: Interaction plot for mean response between visualization tasks for two different scenarios (Mountaineering in Scotland and Hotels near Portee)

Table 8.23: Mean response for information presentation and understandability of different visualizations (Spatial Qualifiers).

Question: It was easy to understand the information presented by this interface ?	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.6957	0.7029	23
	2.0 Cottages in western Scotland.	3.44	0.7681	25
	Total	3.56	0.74108	48
2. Density surface representation	1.0 Biking south of London.	4.2609	0.91539	23
	2.0 Cottages in western Scotland.	4.32	0.69041	25
	Total	4.29	0.79783	48
3. Footprint based representation	1.0 Biking south of London.	2.9565	1.10693	23
	2.0 Cottages in western Scotland.	2.96	1.09848	25
	Total	2.9583	1.0907	48
4.Cartogram	1.0 Biking south of London.	3.4358	1.03687	23
	2.0 Cottages in western Scotland.	3.480	0.91833	25
	Total	3.4583	0.9664	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

iii) Legends and Annotations as supporting aides

It is noted that a negligible difference (Significance) is drawn between the responses of the users for legends and annotations as supporting aides illustrated in Table 8.24.

Table 8.24: Mean response for legends and annotations as supporting aides for different visualizations (Spatial Qualifiers).

Question: The use of legends and annotations made it easier to understand the visualization ?	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.4783	0.84582	23
	2.0 Cottages in western Scotland.	3.64	0.757	25
	Total	3.562	0.79643	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

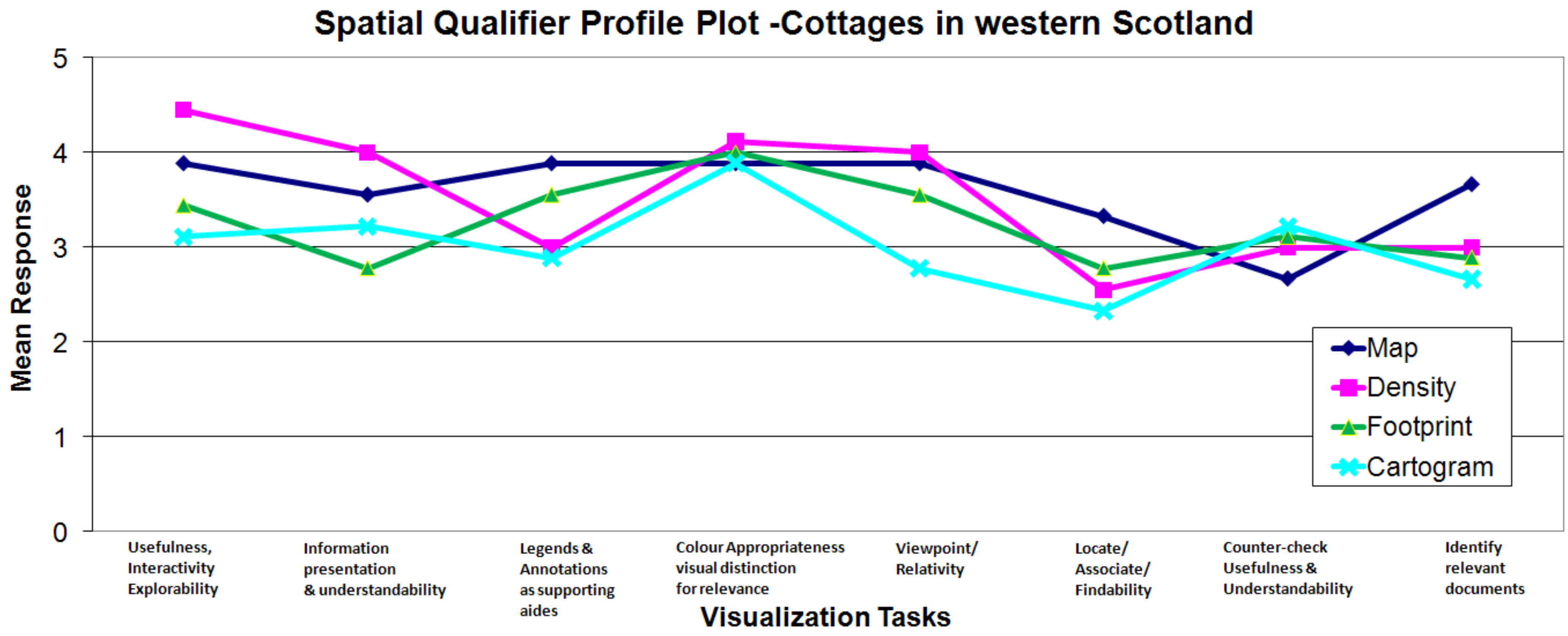


Figure 8.6 Interaction plot for mean response between visualization tasks comparing map, density, footprint and cartogram representations for a query cottages in “western Scotland”.

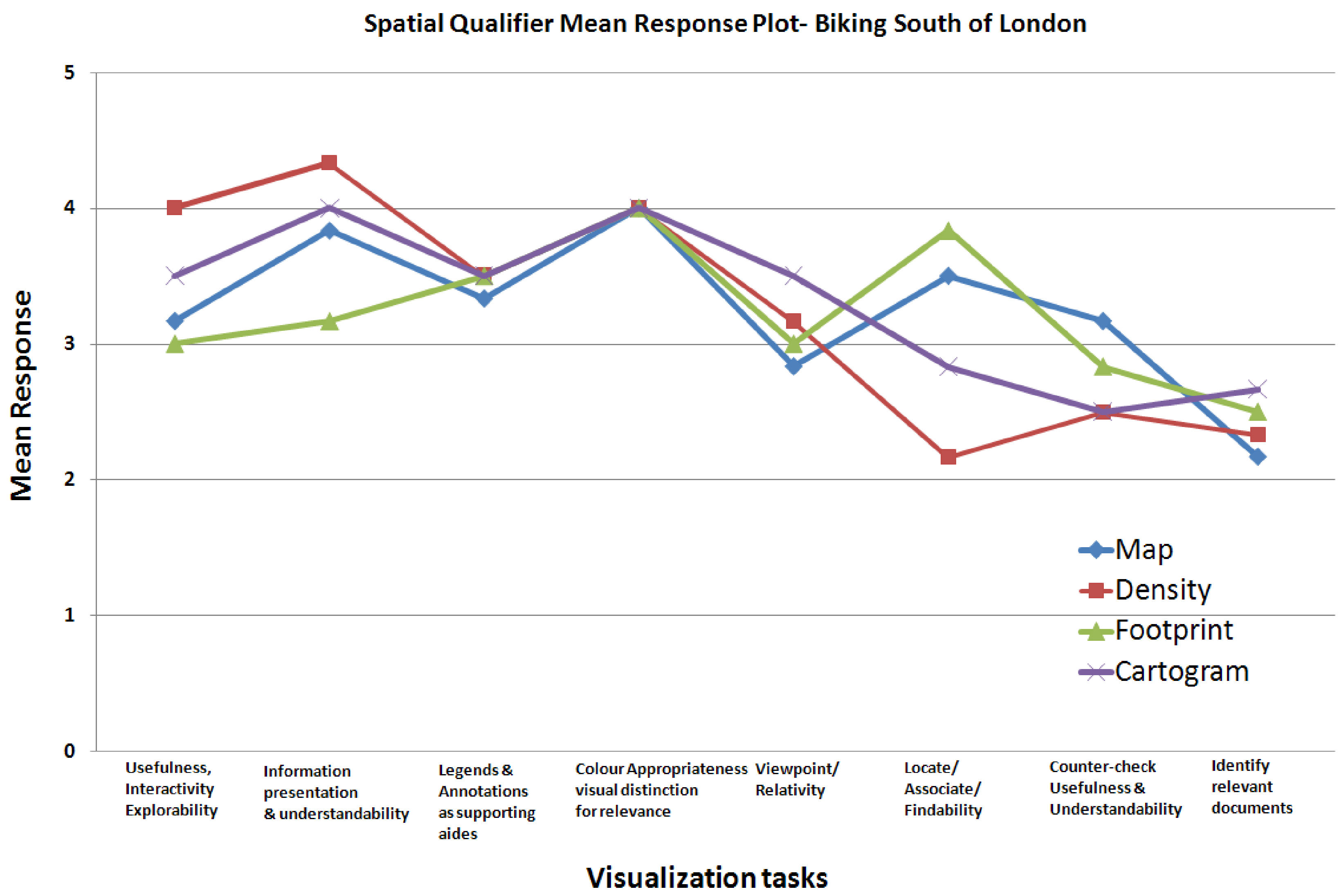


Figure 8.7 Interaction plot for mean response between visualization tasks comparing map, density, footprint and cartogram representations for a query "biking south of London"

Table 8.24: Mean response for legends and annotations as supporting aides for different visualizations (Spatial Qualifiers).

Question: The use of legends and annotations made it easier to understand the visualization ?	Advanced interface	Mean	Std. Deviation	N
2. Density surface representation	1.0 Biking south of London.	3.1739	0.93673	23
	2.0 Cottages in western Scotland.	3.16	1.1431	48
	Total	3.1667	1.03877	48
3. Footprint based representation	1.0 Biking south of London.	3.4783	0.94722	23
	2.0 Cottages in western Scotland.	3.40	0.9573	25
	Total	3.4375	0.9432	48
4.Cartogram	1.0 Biking south of London.	3.1339	0.86849	23
	2.0 Cottages in western Scotland.	3.04	1.05987	25
	Total	3.085	0.96381	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

iv) Colour usage and appropriateness

There was no impact on the user's response with colour usage and its appropriateness to illustrate visual distinction in relevant documents (as shown in Table 8.25).

Table 8.25: Mean response for colour usage and appropriateness for different visualizations (Spatial Qualifiers).

Question: The use of colours was useful and appropriate ?	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.95	0.767	23
	2.0 Cottages in western Scotland.	3.84	0.687	25
	Total	3.89	0.72	48
2. Density surface representation	1.0 Biking south of London.	4.0435	0.63806	23
	2.0 Cottages in western Scotland.	4.00	0.8165	25
	Total	4.020	0.729	48

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

Table 8.25: Mean response for colour usage and appropriateness for different visualizations (Spatial Qualifiers).

Question: The use of colours was useful and appropriate ?	Advanced interface	Mean	Std. Deviation	N
3. Footprint based representation	1.0 Biking south of London.	3.826	0.9367	23
	2.0 Cottages in western Scotland.	3.80	0.8165	25
	Total	3.8125	0.8667	48
4.Cartogram	1.0 Biking south of London.	3.8696	0.919	23
	2.0 Cottages in western Scotland.	3.92	0.40	25
	Total	3.895	0.69158	48
Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question				

v) Viewpoint/Relativity

The footprint based representation shows little significant difference (mean diff = 0.4) in comparison with other visualizations for there has been significant impact in the scenarios presented to the users (ref. Table 8.26)

Table 8.26: Mean response for viewpoint and relativity for different visualizations (Spatial Qualifiers).

Question: The visualization provided a good way of looking at the results ?	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.4783	0.845	23
	2.0 Cottages in western Scotland.	3.54	0.883	24
	Total	3.510	0.856	47
2. Density surface representation	1.0 Biking south of London.	3.6087	1.07615	23
	2.0 Cottages in western Scotland.	3.875	0.79741	24
	Total	3.7447	0.94335	47
3. Footprint based representation	1.0 Biking south of London.	3.1739	1.114	23
	2.0 Cottages in western Scotland.	3.54	0.97709	24
	Total	3.3617	1.05141	47
4.Cartogram	1.0 Biking south of London.	3.2609	1.00983	23
Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question				

Table 8.26: Mean response for viewpoint and relativity for different visualizations (Spatial Qualifiers).

Question: The visualization provided a good way of looking at the results ?	Advanced interface	Mean	Std. Deviation	N
	2.0 Cottages in western Scotland.	3.125	0.8509	24
	Total	3.19	0.924	47

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

vi) Locate/Associate/Findability

The ability to identify documents with higher relevance and filter them out has shown significant difference in user's response for footprint based representation with a mean difference greater than 0.5, while map based representation shows a mean difference of 0.4. This is due to the number of results obtainted from the system based on the scenarios presented to the users (ref. Table 8.27).

Table 8.27: Mean response for findability of documents which shared the same location for different visualizations (Spatial Qualifiers).

Question: It was easy to find documents which shared the same location	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.7273	0.88273	22
	2.0 Cottages in western Scotland.	3.333	1.00722	24
	Total	3.5217	0.96	46
2. Density surface representation	1.0 Biking south of London.	2.136	0.884	22
	2.0 Cottages in western Scotland.	2.25	1.07339	24
	Total	2.1957	0.98	46
3. Footprint based representation	1.0 Biking south of London.	3.454	1.0107	22
	2.0 Cottages in western Scotland.	2.958	1.0417	24
	Total	3.19	1.046	46
4.Cartogram	1.0 Biking south of London.	2.272	1.031	22
	2.0 Cottages in western Scotland.	2.25	0.846	24
	Total	2.26	0.92939	46

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

vii) Counter check- ease of use and understandability

Map based representation demonstrate that it has been fairly understood by the users, in comparison with other visualizations drawing a mean difference in user's response of 0.3 (ref. Table 8.28).

Table 8.28: Mean response for counter check- ease of use and understandability for different visualization (Spatial Qualifiers).

Question:It was easy to get confused when using this interface	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	3.181	1.167	11
	2.0 Cottages in western Scotland.	2.8095	1.07792	21
	Total	2.9375	1.10534	32
2. Density surface representation	1.0 Biking south of London.	2.818	1.078	11
	2.0 Cottages in western Scotland.	2.9048	1.0910	21
	Total	2.875	1.07012	32
3. Footprint based representation	1.0 Biking south of London.	3.0909	1.22103	11
	2.0 Cottages in western Scotland.	3.2381	1.04426	21
	Total	3.238	1.090	32
4.Cartogram	1.0 Biking south of London.	3.00	0.89443	11
	2.0 Cottages in western Scotland.	3.00	0.89443	21
	Total	3.00	0.87988	32

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

viii) Identify relevant documents

Each of the scenarios return varying levels of granularity, and results. Map based representation has shown a significant difference in the users' response with a mean difference of 1.4, while Density and Cartogram based representations exhibit a mean difference in response of 0.4 (ref. Table 8.29). The least significant difference being in Footprint based representation (mean difference of 0.3).

Table 8.29: Mean response for effectiveness of different visualizations in their ability to identify relevant documents(Spatial Qualifiers).

Question: This visualization made it easy to find the most relevant documents for my task ?	Advanced interface	Mean	Std. Deviation	N
1. Graphical map based representation	1.0 Biking south of London.	2.2174	0.795	23
	2.0 Cottages in western Scotland.	3.63	0.90214	22
	Total	2.911	1.104	45
2. Density surface representation	1.0 Biking south of London.	2.304	0.9739	23
	2.0 Cottages in western Scotland.	2.772	1.269	22
	Total	2.533	1.14018	45
3. Footprint based representation	1.0 Biking south of London.	2.7826	0.90235	23
	2.0 Cottages in western Scotland.	3.0909	1.01929	22
	Total	2.933	0.9629	45
4.Cartogram	1.0 Biking south of London.	2.304	0.92612	23
	2.0 Cottages in western Scotland.	2.7273	1.16217	22
	Total	2.511	1.0579	45

Coded using Likert Scale data where the scale is set as follows (Strongly Agree-5, Agree-4, Neutral-3, Disagree-2, Strongly Disagree-1). Missing data values are replaced by the mean of user's response for this question

8.3 Mock-up Visualization

A mock-up is understood to be a graphical representation that approximates to what is foreseen as a possible solution to the challenge of representing spatially aware information retrieved from the internet. Two mock-up visualizations, spatial activity maps/plots (partially implemented) and isoline representations have been presented to the participants. Statistical insights corresponding to these different mock-ups serve to envisage practical challenges in information presentation and comprehension. This approach not only gives in adequate room for enhancement of the visualization methods, but also they give us insights where they fail to meet the requirements of the users.

8.3.1 Spatial Activity Maps/Plots

Spatial activity maps are a form of pictorial representations that illustrates activities based on the availability of the resources at a specific geographic location. Web pages are rich with information describing resources, events and detailing activities that can be performed at a specific geographic location. Activities such as cycling, One such example is http://www.Cardiff.gov.uk/cycling/ a document detailing cycling resources in and around Cardiff region. Activities offer the further possibility of classification of the documents based on the resources or activities based on the theme of the web pages. Subsequent classifications can be achieved and documents can be visualized using the Touchgraph visualization technique. Figure 6.4 demonstrates the organization of activities in reference to the region of interest on a two-dimensional plot, with the distance from the centroid of the expressed region of interest on the x-axis and spatial hierarchy on the y-axis. A decision grid is provided to support users to visualize the proximity in reference to the centroid of the expressed region of interest. Adding to this, the activity resources are plotted on the map at the required resolution as shown in Figure 6.5. Table 8.30 presents descriptive statistics gathered for spatial activity maps/plots presented to the users.

Table 8.30: Descriptive statistics for spatial activity plots or maps.

Criteria	N	Range	Min	Max	Mean	Std.Deviation	Skewness
1.Easy to understand the information presented in the mock-up.	46	4	1.0	5.0	3.913	0.9147	-1.279
2.Supporting aides- use of legends and annotation has made it easier to understand the information presented.	47	4.0	1.0	5.0	3.531	1.080	-0.680
3.Colour usefulness & appropriateness	48	3.0	2.0	5.0	3.875	0.7034	-0.203
4.Viewpoint(Provide a good way of looking at the results).	48	4.0	1.0	5.0	4.0	0.8251	-1.186
5.Usage of symbols to illustrate activities was effective and appropriate.	44	4.0	1.0	5.0	4.318	0.9092	-1.664
6.Activity based contextual classification an effective approach to representing information.	47	4.0	1.0	5.0	3.936	0.894	-0.823
7.Cross check- Usefulness & understandability (it was easy to get confused when using this interface)	48	4.0	1.0	5.0	2.5	0.945	0.867
8.Are decision grids an effective means of identifying documents based on proximity and direction.	47	4.0	1.0	5.0	3.297	0.857	-0.629
9.This mock-up visualization makes it easy to find the most relevant documents for my task.	44	3.0	2.0	5.0	3.772	0.6773	-0.630
10.It is an effective way of abstraction of the results.	43	3.0	2.0	5.0	3.930	0.6688	-0.422

Statistics reveal that 76% of the participants either agree or strongly agree that it was easy to understand the information presented in the mock-ups. Activities were represented using symbols, 76% of the participants either agree or strongly agree that the symbols were effective and appropriate. Activities were in abstract terms & themes described with in the web pages. 68% of the participants either agree or strongly agreed that it had been an effective way to depict the results. Foreseeing classifications based on the kind of activity each of the web page represents, 68% of the participants either agree or strongly agree that the contextual classification based on activities was an effective way of representation. While only 46% of the participants either agree or strongly agree that decision grids of 5x5 km was an effective means of identifying documents/activities based on proximity and direction of a specific candidate region. The candidate region here was the expressed region of interest by the user.

8.3.2 Isoline representation

Isoline representations are a form of representation to demonstrate the spatial variability of the result sets. Spatial averaging through an interpolation process yields a surface, portrayed by isolines, which can be used to mask geography related statistical information. The concept of geographic masking was first introduced by Armstrong (1999) to preserve confidentiality of health related data, in an effort to protect privacy of individuals. Isoline representations can be widely used for representing such information. Leitner and Curtis (2004, p.215) claimed to be "*the first to utilize an empirical perceptual study to identify acceptable design solutions for presenting confidential point data on maps*". With serious limitations for visualizing confidential information, isoline based representation has the potential to be widely adopted for keeping annoymous confidential information or statistical information.

80% of the participants either agree or strongly agree that it was easy to understand the mock-up visualization presented in the Figure 6.6. 68% of the respondents either agree or strongly agree that it was easier to understand the information presented using this mock-up. 80% of the participants either agree or strongly agree that colour usage was useful and appropriate and 54% either agree or strongly agree that it has been an useful way of looking at the results. Only 18% of the participants consented that this visualization could help users find documents with shared location, which is ofcourse obvious because of the interpolated surface representation. While 44% of the respondents said that it was easier to get confused using this interface. Finally, 38% of the population either agree or strongly agree that this visualization makes it easy to find the most relevant documents.

Table 8.31: Descriptive statistics for isoline representation

Criteria	N	Range	Min	Max	Mean	Std.Deviation	Skewness
1. It was easy to understand the information presented in the figure above ?	48	3.0	2.0	5.0	3.895	0.8565	-1.066
2.The use of legends and annotations made it easier to understand the visualization ?	48	4.0	1.0	5.0	3.667	1.1172	-1.0107
3. The use of colours was useful and appropriate ?	48	3.0	2.0	5.0	4.083	0.7672	-0.735
4.The visualization provided a good way of looking at the results ?	47	3.0	2.0	5.0	3.617	0.8736	-0.171
5.It was easy to find documents with the same location ?	41	4.0	1.0	5.0	2.853	0.90997	-0.116
6. It was easy to get confused when using this interface ?	46	4.0	1.0	5.0	2.60	0.74471	0.791
7. This visualization makes it easy to find the most relevant documents for my task ?	44	4.0	1.0	5.0	3.181	0.9710	-0.383

8.3.3 Comparison of mean response for working systems and mock-ups

Statistics reveal that there has been a higher mean response value for mock-ups in comparison with visualizations that users have interacted with. To understand the effects interaction plots were constructed as shown in Figure 8.7 & 8.8 (§ref. p.223 & p.224). These compare the criteria information presentation and understandability, use of legends and annotations as supporting aides, colour usage and appropriateness, and viewpoint/relativity. These plots reveal that there is a stronger effect on interaction with working visualizations, hence relatively lower mean response in comparison

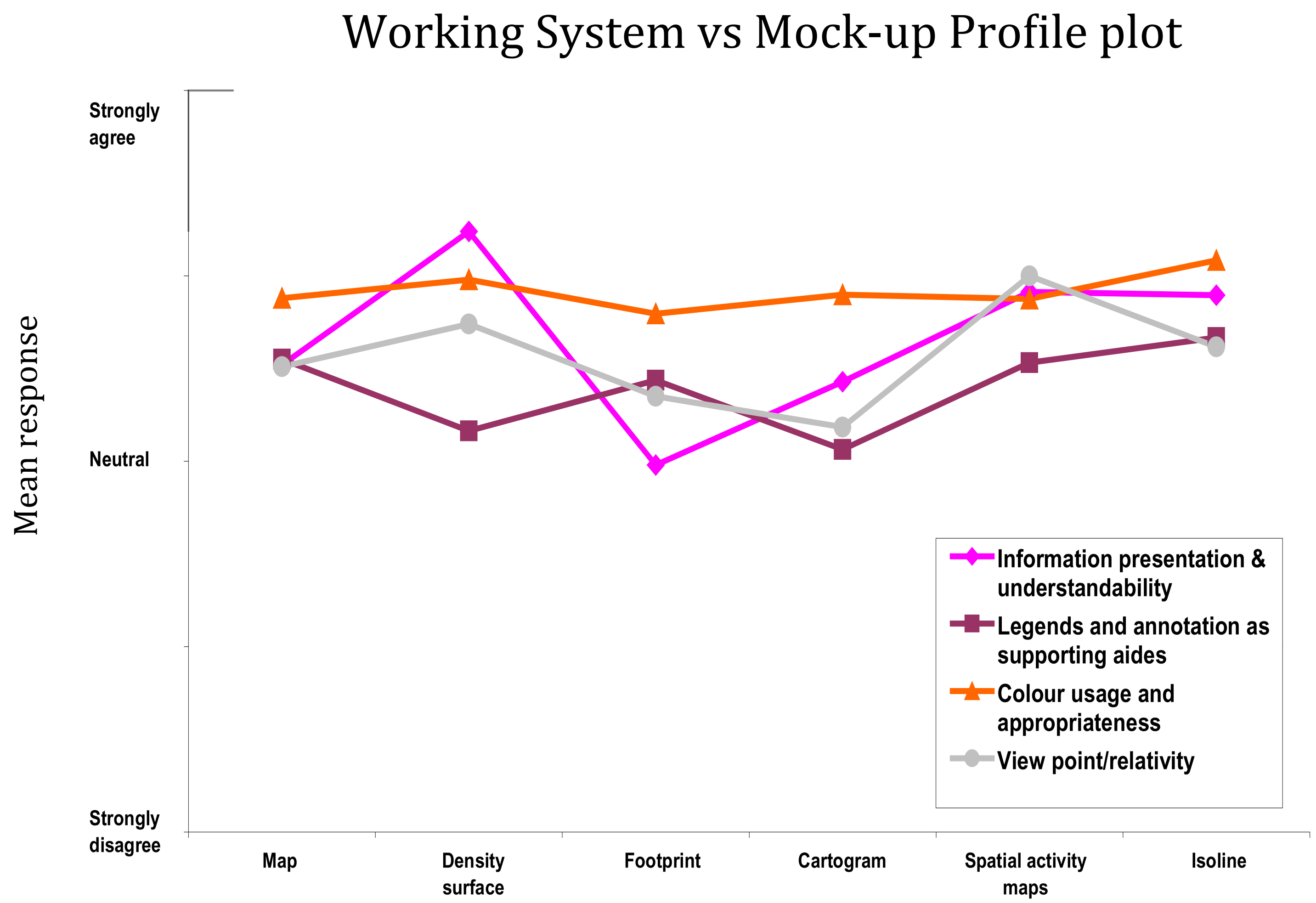


Figure 8.8: Interaction plot comparing working system (map, density, footprint and cartogram based representations) with mock-up visualization (spatial activity map and isoline representation) based on common criteria.

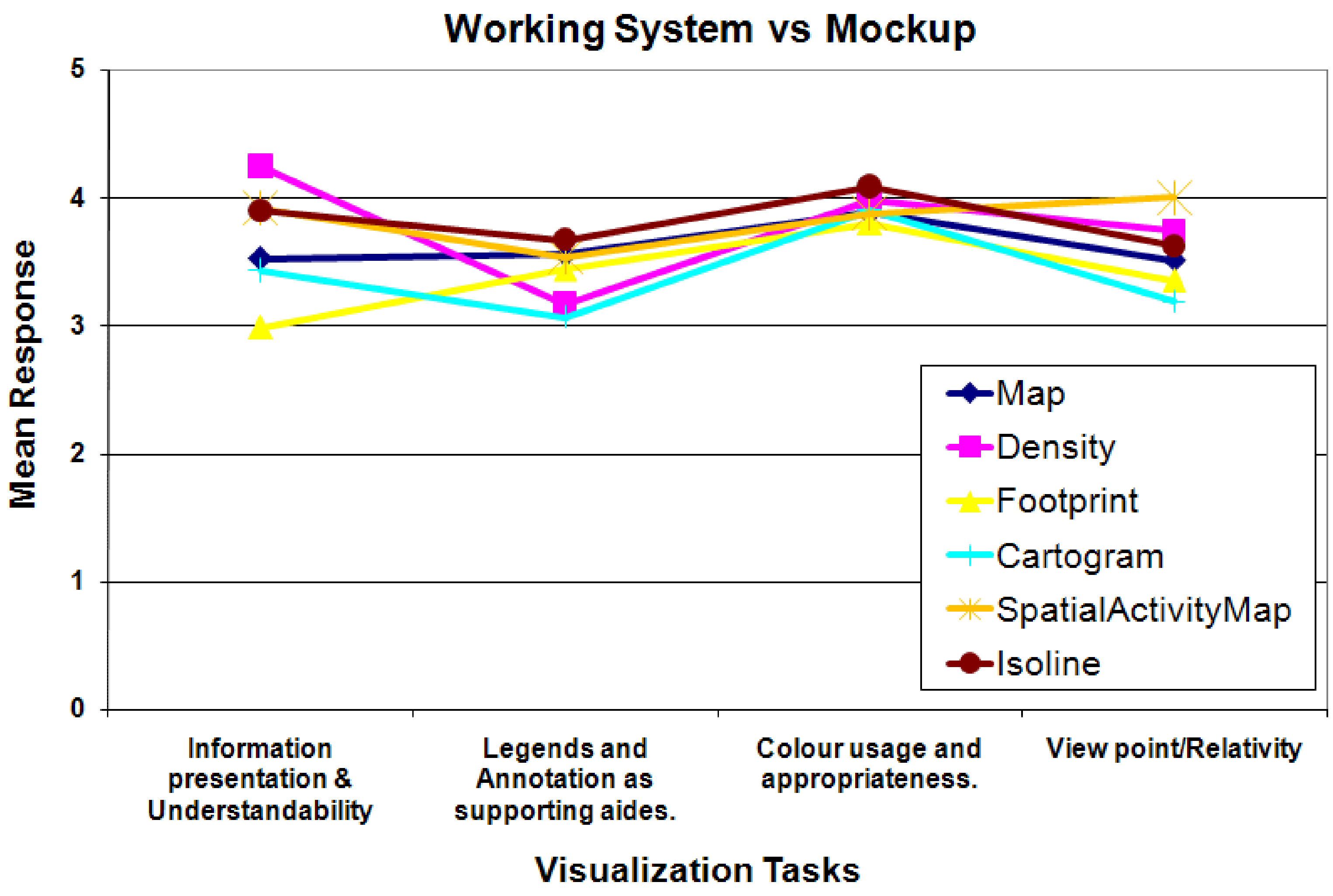


Figure 8.9: Interaction plot comparing working system (map,density,footprint and cartogram) with mock-up visualization (spatial activity map and isoline representation) based on common criteria.

with the mock-ups which are static and do not involve interactions. An inference can be drawn that design concepts when presented to the users have fairly higher mean response values in comparison with working visualization systems. This may be due to numerous factors such as interactivity, cognitive inference of the users, exploration with real data etc.

8.4 Combined relevance vs. perceived relevance

The results obtained from the SPIRIT engine are based on a combined normalized score which takes into account both spatial relevance as well as textual relevance based on a 'Best Match' approach (cf. Chapter 5). For the interim prototype, a document's spatial similarity score is the highest footprint similarity score of the footprints it contains (Arampatzis and van Kreveld, 2004). If a user is looking for "*airports near London*", a document that refers to both "Gatwick" and "Stansted" is scored as referring only to "Gatwick" since that is the nearest airport of the two. Such a document, however, should be scored higher than others that refer only to "Gatwick" as it provides more relevant information (Arampatzis *et al.*, 2004).

The results presented by the SPIRIT prototype are based on the assumption that a query can have only one footprint (a user is interested in only one location), while the data characteristics reveal that the documents have multiple footprints (referring to more than one location). The experimental studies have focused on drawing statistical inference to estimate what percentage of the respondents have been capable of identifying relevant documents (system generated ranking) and identifying documents which the user perceives as more relevant than those generated by the SPIRIT system.

In total 10 questions (cf. Appendix-G, J and Table 8.32), were designed primarily to extract qualitative feedback to draw statistical inference between system generated relevance and perceived relevance of the user.

Table 8.32: Qualitative feedback questionnaire to draw statistical inference between system generated relevance and perceived relevance for the user.

1.	List the top 5 documents found. - *(measure of combined relevance obtained from the SPIRIT system)*
2.	List the names and addresses of 3 places of interest found from the relevant documents for your task - *(measure of combined relevance obtained from the SPIRIT system)*
3.	Identify places where relevant documents were found with respect to your query? - *(perceived relevance measure)*
4.	List the 5 documents, from the results list, which are most relevant with respect to your query? - - *(perceived relevance measure)*
5.	Identify the regions with the highest document density? (Density surface representation) -*(measure of combined relevance obtained from the SPIRIT system)*
6.	It was easy to find the region of the following documents (Footprint based representation): • 1st ranked document • 2nd ranked document • 3rd ranked document *(measure of combined relevance obtained from the SPIRIT system)*
7.	Identify the top five documents with the smallest footprint (geographical extent)? (Footprint based r*epresentation) - (measure of combined relevance obtained from the SPIRIT system)*
8.	Identify the top five documents with largest footprint (geographical extent)? (Footprint based r*epresentation) - (measure of combined relevance obtained from the SPIRIT system)*

Table 8.32: Qualitative feedback questionnaire to draw statistical inference between system generated relevance and perceived relevance for the user.

9.	List the 5 documents which you perceive as most relevant with respect to your query after using footprint based filtering technique? - - *(perceived relevance measure)*
10.	It was easy to find the regions of the following documents (Cartogram based representation): • 1st ranked document • 2nd ranked document • 3rd ranked document *(measure of combined relevance obtained from the SPIRIT system)*

Statistical insights drawn from these qualitative feedbacks reveal that almost all the participants had either listed all the top 5 documents or partly listed top 5 documents obtained from the SPIRIT search engine. In total, 38 participants out of 50 (i.e. 76%) identified geographically relevant documents which do not resemble the top 5 documents presented by the SPIRIT engine. This indicates that the geographic relevance perceived by the users is considerably different from that generated by the SPIRIT engine.

8.5 Observations

During the experimental studies, a few users were selected by randomly generated numbers for participating in think aloud sessions. Think aloud sessions were carried out using WINCAM2000 and user's interactions were recorded for the whole study period of 1.5 hours. In total 15 participants were involved in this exercise during phase-I (Unsupervised sessions). Additional qualitative feedback was gathered from the participants during the experimental studies. A list of key findings during the experimental studies are detailed here:

8.5.1 Background & Expertise

1. 47 participants were chosen who had prior experience with search engines and were aware or have used Google (www.google.com). A few others cited that they are aware of other search engine tools such as vivismo (www.vivisimo.com) or www.search.ch.
2. Participants had varying experience and exposure to a number of geo-web services such as map24 (www.map24.com -11 participants had used it) or Google maps (8 had prior experience). A few others cited that they have used mapping services such as www.mapquest.com, map.search.ch, the web GIS browser for the city of Zurich, go2map.sina.com.cn, swissgeo.ch, www.gis.zh.ch, or the map-based visualization of SBB and ZVV (www.sbb.ch and www.zvv.ch).
3. Six participants documented that they had widely used geo-web services for the following purposes:
    - travel/tourism related information gathering or route planning
    - acquiring geographic information related to a specific place information about topography

4. Participants conceptualized a number of activities that they wish to perform and have queried in the past using geo-web services or search engines. These were:
    - *Hiking*
    - *Climbing*
    - *Travel maps*
    - *Ski-touring*
    - *Route Planning*
    - *Military service*
    - *Outdoor activities*
    - *Jogging.*
    - *Rollerblading*
    - *Cycling or biking*
    - *Walks*
    - *Paragliding or flying.*

8.5.2 Map Interactions

The think aloud sessions also yielded feedbacks concerning map interaction techniques that are summarised here:

1. Annotations at high resolutions can be helpful in orientation, especially when the user zooms in. Additional legends based on the type of roads, railways or motorways would be extremely helpful.
2. Better readability and understanding can be achieved by decreasing the density of results displayed using the map based representation. This can be achieved by allowing users to limit the number of results to be displayed.
3. Suitable colour distinction needs to be achieved for the density surface representation, as it resembles the colour of the sea.
4. The possible reasons for slow system response on zooming and subsequent activation of links could be due to the client side scripting.
5. It is difficult to comprehend short snippets or summaries within each document.
6. Difficulties exist in comprehending the footprint concept. This concept should be better illustrated.
7. Colour coding cartograms according to relevance and the linking with documents in the result sets would present comprehensive understanding of their association with a geographical unit.
8. Additional information relating to the size, colour and orientation of cartograms is necessary to grasp the purpose of adopting such visual representation.
9. Overlaying or overlapping stacks/clustering or cluttering makes it difficult to read.

8.5.3 Respondents' Wishlist

The participants also voiced several suggestions for enhancements of the SPIRIT prototype during the think aloud recordings:

1. Visual system feedback is an essential component for an interactive system; this would considerably increase the efficiency of the users to glean through results sets.

2. Provision of history functionalities, to memorise the user's selections, and interaction to select and remove documents.
3. Provision of title and visual system feedback is necessary for the query disambiguation interface.
4. Provision of "Northing" on the map would be helpful to the users in identification of cardinal directions.
5. Provision of a numerical scale for filtering documents would be helpful
6. Provision of annotation for the cartograms and summary of the document results would be helpful.
7. Provision of topographic information would be extremely helpful to provide an overview of the landscape.
8. Provision of a "point" as reference to a query is necessary for spatial relations such as "near", "north of", "south of", "east of "and "west of".
9. Incorporation of viewsheds could be helpful.
10. Spatial relationships with a combination of cardinal directions within a distance, for example, "max. 10 km north of Glasgow" would be helpful.
11. Additional annotation on selection of footprints describing the relation with the documents and illustrating the extent of the information within the documents would be helpful.
12. Functionality to have multiple search regions, or comparison between the results, provision of statistical information can enhance the visualization.
13. The capability to save the user's selections and groupings as layers would be helpful.
14. Activity maps with respect to a single document or user selection, or reference to a single document region or candidate region to plot activities within or in reference to or based on resources available at a specific geographic region would be helpful.
15. Provision of tool tips and labels for exploratory tools would be helpful.
16. Adaptive zooming or step-wise zoom (in or out) functionality would be helpful to explore through the results.
17. Higher resolutions for density surface representations would be useful.
18. Toggling and grouping (based on interactive selection) of the result layers would enhance readability.
19. Toggling the density surface representation as layers could be helpful in improving the readability of background maps. Similar functionality to toggle layers should be facilitated for the mock-up isoline representation.

8.5.4 Arguments and criticisms

Finally, several arguments and criticisms of the system were brought forward by the participants:

1. Explain why the geographic extents of the footprint have been depicted as rectangle
2. The use of pictogram is a nice idea to demonstrate the activity-based representation. The distance on the X-axis is one dimension and can be understood. However, it is unclear what the Y-axis means. What does the spatial hierarchy mean?

3. Why isn’t it possible to take the geographic limit as an approach for footprint based representation?
4. The test collections comprise of directory pages. How relevant are directory pages within the result set?

8.5.5 Suggested improvement to the existing SPIRIT system

During the experimental studies, a few participants provided us with valuable insights to enhance the existing SPIRIT system. They are:

- provision to formulate multiple queries for a single or multiple geographic regions . Thus enabling users to perform a comparative analysis based on user’s requirements.
- support for clustering of results using the geographic information present can help users in analysing the result set.
- improving the disambiguation history by actively enaging users and provision of recording and training the engine based on the geographic disambiguation.
- support for complex geographic prepositions between concepts and geographic scope.
- enabling of exploratory tools on selection of different visualizations could facilitate zoom-in/out to the required level of details. Increased explorability and interactivity when using density surface representation and cartogram based representation.
- combined ranking information in the visualization has been challenging to perceive to the extent expected. Provision of effective visual feedback based on user’s selection, highlighting and filtering would server to meet the users requirements.
- effective visual feedbacks on selection of specific feature in each of the visualizations.

8.6 Summary

Comprehending the results obtained from the experimental studies, this chapter presented the significance of different empirical measures for different visualization methods adopted or designed. Preliminary analysis of the results using descriptive statistics revealed some missing data and missing value questionnaires that were discarded. In the course of this chapter, we discussed a number of subjects involved and the nature of the responses for each session be it, self-learning, supervised or unsupervised sessions. Valuing varied perspectives, subjects involved in the studies were drawn from varied academic backgrounds. To understand the nature of the results, and significant differences in users response, repeated measures ANOVA has been carried out to perform pair-wise comparisons of different visualizations. Repeated measures ANOVA was adopted because it demonstrated individual patterns of changes and it economised on the subjects. By excluding between-subject sources of variability from the experiment error, repeated measurements designs often provide more efficient estimators of the relevant parameters than cross-sectional designs with the same number and pattern of measurements.

Furthermore, to understand the interactions between different measures, profile plots or interaction plots were drawn. Similar analysis was carried out to investigate the effects of different scenarios presented to the users, which impart varying levels of granularity, context and scope of the results. Statistics reveal that there was relatively higher interaction for the scenario “cottages in western Scotland” in comparison with “ biking south of London”. While in the simple structured interface, there were also significant shifts in the users’ response for varying levels of granularity, context and indexed documents. Furthermore, mock-ups (spatial activity and isoline representations) have been evaluated for their practical understandability. Also, insights have been drawn from the users’ responses by comparing the working prototype representations (map, density, footprint and cartogram) with the mock-up representations (spatial activity and isoline representation).

8. Results

The subsequent chapter runs through the critical findings of this experimental study and tries to draw meaningful inferences from statistical insights and experiments carried out not only in this chapter but also during the earlier course of this research, reported in previous chapters.

Chapter 9

Discussion

"Reasoning-power; or Ratiocination, called by some Dianoetic Reason, is the power of drawing inferences that tend towards the truth, when their premises or the virtual assertions from which they set out are true.".

-An Essay towards Improving Our Reasoning in Security and in Uberty', EP 2:464, 1913

The experiments reported in Chapter 8 have focused on evaluating the effectiveness of several techniques for the visualization of search results in spatially aware information retrieval of web documents, in order to infer their strengths and weaknesses. We will start this chapter by a brief summary of the main characteristics of the visualization methods developed in the SPIRIT prototype (Section 9.1). Following that, we will discuss the results of the experimental evaluation conducted in our study:

- Information seeking process (Section 9.2)
- Cognition and conceptualization of geographic space (Section 9.3)
- Mock-up representations (Section 9.4)
- Impact of granularity and context on visualization of search results (Section 9.5)
- Combined relevance vs. perceived relevance (Section 9.6)
- Learnability of the SPIRIT prototype (Section 9.7)
- Limitations of the experimental design (Section 9.8)

9.1 Visualization methods used

One of the prime challenges for any visualization method is the problem of accommodating as many documents on the geographical representation as can be fitted in the ranked textual list. With varying spatial extents of the individual document footprints within the result sets the task of visually presenting this information in a convincing manner to the users is non-trivial in nature. Although a relevant document may well be contained in the top ten, the ability of the simplest form of representation – a 2-D map of document centroids displayed by point symbols – to convey information about many more footprints is severely limited. Since most documents have identical footprints when using point symbols, zooming at any level of detail will not allow them to be visualized independently and instead two interaction techniques were adopted in the SPIRIT prototype to convey their locations to the user, i.e. brushing and highlighting corresponding documents and (URLs) with corresponding footprints. However, footprints still obscure each other and must be schematically illustrated (e.g. through a different symbol) to better convey information to the user.

Problems encountered by the users when interacting with the map based representation of search results led to the investigation of alternative visualization approaches such as Density surface representations, Footprint based filtering technique and Cartogram based representations. The spatial patterns exhibited by the geographic extents (footprints) of a document and their underlying relations between documents make it non-trivial to derive boundaries. Thus, a density surface representation was created which was particularly aimed at representing such imprecise regions. This approach also equips users with perspectives to explore how the density of the documents varies across space, showing "hot spots" of incidence.

The Footprint based representation focuses on the scope and shared position of footprints within and between documents, building a hypothesis that documents with a few or single narrow geographic extents may describe a single location and be thematically well defined, while documents with multiple and large geographic scopes may typically describe activities that might be carried out in a region or a journey through a region, facilitating more general descriptive information.

Another approach adopted to collectively summarize the frequency of documents in a specific geographic unit (or administrative unit) was demonstrated using cartograms that reflect a theme based representation. In the following sections, we discuss the outcomes of the experimental studies with respect to the usability of the above visual representations implemented in the SPIRIT prototype as well as mock-up visual representations presented to the users.

9.1.1 Outcome of the empirical studies.

This section sets out discussing the critical findings of the experimental results presented in Chapter 8 and summarizes the inferences drawn from the statistical insights. The efficacy of the various representations adopted for visualizing geographic information retrieved from the internet are assessed using the empirical criteria (ref. Table 7.6 Chapter 7 p.181) that records users' acceptance on a Likert Scale. These empirical measures assess the usability of specific visualization methods and techniques adopted with respect to the tasks they serve.

Pre-retrieval

Query retrieval

The concepts of spatial search, regional search and place name search have been at the core of this study, where participants have expressed their information requests either by using natural language or by graphically illustrating the region of interest. The distinction between spatial search, regional search and place name search can be drawn based on the perspectives they serve. From both qualitative and a quantitative perspective, spatial search employs either of these means:

- *Graphical depiction of a geographic location on a map or*
- *Geographic name identification with or without the use of keywords or by using coordinates.*

For users employing either of these means, their expression is converted into geographic coordinates to extract this information.

The place name search process is merely a lookup from a collection in a place name gazetteer from possible matches of geographic place names. Place name based search can be broadly used to identify the kind of resources available at a specific geographic location. It can however, be widely argued that there is considerable overlap between spatial search and regional search. Observing the interactions, we have identified that participants expressed their region of interest either by sketching an area of interest by drawing a polygon, instilling a point on the map using the Graphical Interface. While the simple map based representation was used to express their query in natural language as a triplet ("*concept, spatial relation, @location/place name*"). Given the exercise, participants find that it has been easier to get started with the system and initiate an information request using simple map based representation. Support for such inferences are drawn from strong response values of acceptance (Table 8.2 response 1). For instance, the scenario "Hotels near Portee" , very few users had geographic knowledge of the location "*Portree*", which is a town on an island and thus they opted to express their region of interest by typing in *Portree* in the Simple Structured Interface, which translated *Portree* into geographic coordinates using the Geographic Ontology, which then georeferences the information with respect to the region "*Portree*".

The Graphical Interface provides the users with the possibility to graphically illustrate the region by sketching the region of interest. For instance "*western Scotland*" or "*south of London*" which can be termed as a perceptual region or vernacular region. For each of the scenarios assigned to the users during the experimental studies, the perceptual regions sketched differed from user to user. Support for these inferences are drawn from the think aloud video recorded session that detail how users expressed their information request using the graphical interface.

Geographic prepositions

Users expressed that the SPIRIT prototype supports a comprehensive list of spatial relationships (Table 8.2, response: 3) either topological, proximal or directional such as *near, in (inside), north of, south of, west of, east of, within distance, cross, outside.* A few users cited that it would be useful if SPIRIT could support complex queries with a combination of spatial relationships.

Result presentation

By understanding the elementary tasks involved in the spatially aware information seeking process, the empirical criteria have been adopted to assess the efficacy of various visualization methods. We discuss the critical findings of the experimental studies against each empirical criterion adopted.

Questionnaire	Criterion/interaction task
1.This visualization was easy to use	*usefulness, interactivity, explorability*

- For the empirical criteria usefulness, interactivity and explorability, significant differences between Density-Footprint and Density-Cartogram representations were recorded, which is due to the functionality and purpose each of them serves to address (ref. Table 8.4, Profile plot Fig: 8.1). The Density surface representation presents an overview of the spread of the documents over a geographic region while the Footprint based representation presents the scope/geographic extent of individual footprints within and between documents. Whereas Cartograms depict the number of documents present in each administrative unit.
- Further analysis, to understand the training effects (ref. Fig: 8.3 & 8.4) reveals that there has been a significant shift in the users' response for the Footprint based filtering technique (supervised session) and Map based representation (supervised session). Following the training, concepts such as geographic extent of individual footprints and their underlying relations within and between documents were well understood by the users which reflects the significant shift in their response.

- There have been significant shifts in users' responses for the scenarios "*Cottages in western Scotland*" and *"Biking south of London*" using the graphical interface while no significant difference has been recorded for scenarios using the simple structured interface. This is attributed to the varying levels of granularity, context and scope of the results obtained from the SPIRIT system, as different users define the perceptual region of interest by sketching. This in turn influences the outcome of the results obtained from the SPIRIT system.

2. It was easy to understand the information presented by this interface	*Understandability, information presentation*

- Significant differences in the users' response between Map-Footprint based representations, Density-Footprint and Density-Cartogram representation in Table 8.4 are attributed to level of interactivity, unique perspective each of these representations serves to address. Further analysis to investigate the training effect reveals significant shifts in users' response for Footprint and Map based representations, attributing to their improved understanding of the concepts of geographic footprint following training. Significant shifts in users' response for Footprint based representation attribute to the perspective they present that users can selective prune through the results based on two parameters, namely the size of the footprint and the number of documents.
- For the cartogram based representation the learning effects were less pronounced.
- Statistical analysis reveals that there has been a negligible difference in the mean responses for spatial qualifiers for the simple and graphical interface for all the scenarios presented to the users.

3. The use of legends and annotations made it easier to understand the visualization	*Supporting information for understanding or interpreting the information presented*

- There has been negligible difference in users' response for various representations, which is attributed to the lack of appropriate legends available to users in the interface to interpret the results presented on the map. Users had to dig through the HELP to learn about specifics of each scheme adopted for different visualizations.
- Learning effects were more pronounced for footprint and cartogram based representations in comparison with other representations.

4. The use of colours was useful and appropriate	*Distinction, discrimination, visual relevance*

- The usage of appropriate colour to demonstrate distinctiveness based on relevance (i.e. bluish grey (high relevance) to light grey (low relevance)) has no significant difference in the users' response, which is attributed to the use of a less divergent colour scheme. Having used a more divergent colour scheme from ColourBrewer (Brewer, 1994) might have thrown in significant differences in users' response.
- Significant shifts in users' response are accounted for training effects for Density, Footprint and Cartogram based representations.

5. The visualization provided a good way of looking at the results

Multiple-view capability, understandability

- The density surface representation is most widely accepted representation in comparison with other visual representations, which is attributed to the unique perspective and abstractness with which it presents spatial density of document occurrences over a defined geographic region.
- Learning effects are less pronounced for Density surface and Map based representations in comparison with Footprint and Cartogram representations. Despite the fact that uniform training was provided with similar examples for all the visualizations detailing key functionalities of each of them.

6. Zooming was a useful tool in looking at the results.

Filtering (excluding), highlighting (including), sorting or re-expression

- Significant differences in users' response for Map vs. Cartogram are recorded, which is attributed to the constant changing granularity and context as the user moves through various zoom levels using the map based representation, while the cartogram based representation provides only an overview of the frequency of the document footprints within an administrative unit.
- Density and Cartogram representations provide only up to two zoom levels. Added to that the results are displayed at a coarser granularity for map and footprint based representations, while only three study areas had higher resolution available, namely Edinburgh (UK), Cardiff (UK) and Zurich (CH).

7. It was easy to find documents which shared the same location.

Findability, selection.

- Significant differences have been recorded for the tasks to locate or associate footprints between and within individual documents for Map vs. Density, Map vs.Cartogram and Footprint vs. Cartogram. This is attributed to the level of interactivity Map and Footprint based representations offer to interactively select individual footprints and prune them. While the Density surface representation and Cartogram provide only an abstract overview of document densities over geographic regions.
- Learning effects are more pronounced for Map and Footprint based representations which is attributed to the detailed explanation of key functionalities of both visualizations.
- With varying granularity, context and results obtained from each of the scenarios, the spatial qualifier profile plot reveals significantly higher response for Map and Footprint based representations, attributed to their support for interactive and selective pruning of the results.

8. It was easy to get confused when using this interface.

Counter check for ease of use and understandability.

- Footprint based representation was least understood and confuses users most. This is attributed to their lack of understanding of the concepts such as geographic extent of a footprint, represented as a bounding box. In addition to that users find the overlap of these footprints and highlighting of multiple geographic extents for a document selection most confusing.
- Learning effects are more pronounced for Footprint based representations, which is attributed to detailing of key functionalities that this specific visualization addresses.

9. This visualization made it easy to find the most relevant documents for my task	*Relevance/Rank/Findability*

- One of the primary objectives of the spatially aware information seeking process is to identify the most relevant information with respect to a geographic location. The footprint based representation followed by the Map based representation has been the most widely accepted visualization to identify the most relevant documents for a given task. This is attributed to the inherent characteristics of the visualization that helps users to identify, rank and explore the geographic footprints and spatial extent of individual documents.
- Learning effects are more pronounced for Footprint based representations followed by Map based representations. This is attributed to the fact that the Footprint based representation provides users with functionality to selectively prune the results based on two parameters, namely the spatial extent of the footprints and the number of documents.
- Each of the scenarios returns varying levels of granularity and results. The map based representations has been the most widely accepted representation to find the most relevant documents for the given tasks using the graphical interface.

10. Identify the regions with highest document density	*Explorability and interactivity. highlight*

- Significantly higher response is recorded for Density and Cartogram representations in comparison with Map and Footprint based representation. This is attributed to the inherent functionality of the Density surface representations that provides the spatial density of document occurrences over a defined geographic region in a single graphic view. The peaks and valleys indicate regions of highest and lowest document density. While in the case of Cartogram representations the size of the circles illustrates regions of high and low document concentrations in each administrative unit in the case of Cartogram representations.
- Regions of highest document density can also be identified with clusters and stacks in Map and Footprint based representations. The height of the stacks in Map based representation is indicative of the number of documents that share a similar location.

11. It was easy to find the regions of the documents	*Explorability, interactivity, highlight*

- The Density surface representation is the most widely accepted representation to identify regions of high document density, followed by Footprint based representations to identify regions of the documents (ref. Table 8.13).

9.1.2 Preferred representations

This study reveals that the Density Surface Representation is a widely accepted visual representation in comparison with other visual representations. Such inferences are supported by a stronger response value favouring Density surface representations. This is due to their inherent characteristics that they present a comprehensive summary of the spread of documents over a defined geographic region. Legends and annotations serving as supporting aides have little impact on users in inferring the information presented to them.

The map based representation in its simplest form was widely accepted as it provided users to express their information request by sketching a region of interest, or by clicking a point on the map and the conceptual definitions to lookup. The 2-D map illustrates the structural characteristics of the documents and their corresponding footprints at the same location arranged as a stack. The relatively low acceptance for findability of individual footprints is attributed to the increasing overlap and clustering of footprints at a same location, which affects the readability of the geographical distribution of the footprints either qualitatively or quantitatively. The map based representation, was ideally used as a basic viewpoint to visualize the results. Users opted to switch to other visualizations for further analysis of the results obtained from the SPIRIT engine.

Of the different visual representations, the Footprint-based representation is least understood. Support for this inference is drawn from the statistics in which users state that this is the visualization that gets them most confused. Further analysis of the results reveals that this has improved considerably by providing adequate training to the users. A significant shift in the users' response from unsupervised to supervised training sessions has been recorded with participants responding positively in the latter half (supervised sessions) of the experiments. This is due to the lack of understanding of the concept "*geographic extent of a footprint*" by the user. On the other hand, it has been easier to identify, locate, rank and understand the underlying relationships between footprints and documents using the footprint-based representation. This is due to its ability to selectively prune through the results based on the geographic extent and the number of footprints.

Cartogram based representations was relatively least understood in comparison with Footprint-based filtering representation and Density Surface representations, as it presents an abstract view of document frequency/footprints of the documents within an administrative unit where in the geographic area is not preserved. The degree of abstraction and unfamiliarity with the concept is even higher for cartograms than it is for Density Surface representations. In addition, there is arguably an over-generalization of larger (or elongated) administrative units using cartograms based on graduated circles. Cartograms and Density Surface representations can be widely adopted for geographic masking, ensuring spatial confidentiality. The resulting aggregated views based on the frequency or spread over a region help in regional comparison, which cannot be achieved using map or footprint based representations.

9.2 Information seeking process

Furthermore, experimental studies reveal that the spatially aware information seeking process tends to be more of a non-linear nature, where users tend to move their goal posts based on the information encountered or presented to them by the system. Relevance is a vital aspect of the spatially aware information seeking process. The studies reveal that there are possible mismatches in perceived relevance and system relevance. System relevance methods generate 'objective' measures allowing benchmarking and comparison, in the latter case (perceived relevance), human cognitive systems have developed in such a way that our perceptual mechanisms tend automatically to activate potentially relevant assumptions and our inferential mechanisms tend spontaneously to process observable phenomena, thoughts, memories based on the information presented to them. The process of spatially aware information seeking is initiated through query formulation, with the system facilitating users with two means of expressing their queries either through the use of the Simple Structured Query interface or the Advanced Graphical Visualization interface. In the former users with little knowledge of a geographic region have expressed their information request using natural language in the form of the triplet: "*concept*", "*spatial relation*", "*geographic location*". To address the ambiguities in geographic location, a selective disambiguation approach serves as a tangible option. The latter method facilitates users to graphically express their information request by drawing a polygon and define the region of interest termed as "*candidate region*". Subsequently, the query expansion mechanism helps users to accommodate further expansion or refinement to move their goal posts based on the information encountered during the process.

9.3 Cognition and conceptualization of geographic space

Given the set of scenarios presented in the early stage experimental studies (cf. Chapter 7) and the final experiments (cf. Chapter 8), it reveals that users conceptualize and express their queries based on their experience in navigating through geographic spaces. Users have demonstrated a subjective perception of distance and their information request constantly changed based on their understanding and association of context with respect to a geographic location.

Textual conceptualization of the information request often is referential or relative to a specific geographic location or a place name that can be imprecise or ambiguous in nature. Moreover, the use of spatial concepts and relations also varies according to language and culture. 94 % of the participants strongly agree or agree that the simple structured query interface which facilitates textual conceptualization can be considered as an effective means for information requests and 88% of the participants either agree or strongly agree that the simple structured interface has been easy to use and understand. The simple structured interface has been the preferred choice for participants with adequate place name details.

Graphical conceptualization of the information request is possible only when the users demonstrate considerable awareness of the geographic location on a map. For instance, *Portree* is a small town on an island north west of Scotland, United Kingdom. Not many individuals participating in the experimental studies have been capable of identifying this location using a graphical query interface, due to lack of geographical knowledge of this particular area. Participants expressed their graphical queries either by drawing a polygon across a region on a map and specifying their contextual interest and their spatial association using a geographical preposition or alternatively in reference to a prominent landmark and their spatial association using a geographical preposition.

9.4 Mock-up representations

Two representations – Spatial Activity Maps and Isoline representations – were originally foreseen for implementation, yet ended up being evaluated only in mock-up versions due to temporal restrictions. Both representations received higher response values than the other visualization methods, which however should not be interpreted as a possible outcome of their ease of use. The higher acceptance primarily may be due to the lack of interactivity of the mock-up versions, which limits the influence of external factors such as cognition, interactivity and explorability which may bias the user's perception of a particular representation method.

A fairly higher response (§ref. Table 8.30) for the conceptual illustration of Spatial Activity Maps justifies the necessity to exploit and represent the inherent spatial hierarchy based on which the information can be classified. These pictorial representations demonstrate how symbols and associations of a specific activity can be meaningful in visualizing relevant context. A positive feedback about the envisaged functional requirements demonstrated in the Spatial Activity Maps is the usage of a decision grid which enables the users to visualize the distance from the centroid of the 'region of interest' expressed in their query.

Similar to the Density surface representations, Isoline representations demonstrate the gradient of the spread of documents over the map view. Isoline representations can serve to be useful when specific information needs to be masked or it is required to list all the documents based on a specific similarity measure. Significantly higher response values were attributed to the use of colour gradients to illustrate strong variations in the density of the documents over a geographic region.

The user reactions gathered in the evaluation of mock-up representations have also helped in defining future research directions detailing key functional requirements needed by individual users.

9.5 Impact of level of granularity and context on visualization of search results.

The representative sample of the whole set of documents corresponds to the spatial granularity of the theme based on spatial relevance. The success of the spatially aware information retrieval process indeed relies on the fraction of web pages that contain recognizable geographic context. The SPIRIT system used several gazetteers such as SABE (Seamless Administrative Boundaries of Europe) data, the Ordnance Survey 1:50,000 scale Gazetteer for the U.K., and the Getty Thesaurus of Geographic Names (TGN).

It should also be noted that the SPIRIT prototype had relatively high granularity for three test-regions, namely Edinburgh (United Kingdom), Cardiff (United Kingdom), and Zürich (Switzerland), in comparison to other regions. Figure 4.10 presents a snapshot of the data characteristics where a single web page refers to places which are disparate and irrelevant to the user's request. Grounding of these distinct and varying spatial extents requires uniformity in the underlying spatial granularity for users to explore and respond to the relevant geographic information.

A relevant document on "cottages in western Scotland" describes chalets or cottages in and around the areas of Scotland, namely Loch Lomond, Loch Katrine, Loch Achray, Perthshire etc. In reality (cf. Figure 4.10), the nature of the context has been diverse detailing not only locations outside the region such as in Nepal or the Alps but also significantly diverse contextual information. Often, this has prompted users to re-initiate the query process, or opt out for query expansion, or give up altogether, depending on the user's perception.

In total, four different scenarios reflecting varied resolution have been adopted, each of which presents results based on the specified geographic preposition given/defined in the scenario. Statistical analysis revealed that there has been little impact on the users' responses for granularity and context on the visualization of the search results. The mean response significantly varied for Map based representation, where the users were able to identify relevant documents such as "*cottages in western Scotland*" with higher significance in comparison to "*biking south of London*".

9.6 Combined relevance vs. perceived relevance

Table 8.13 presents descriptive statistics of different visualizations to locate the geographic scope of the documents for the advanced visualization, namely Density surface representation, Cartogram representation and Footprint based representation. Statistics presented in Table 8.13 reveal that participants were successful in identifying regions of higher density using Density surface representations. With the Cartogram based representation it wasn't easy to identify the regions of the top 3 documents listed, as this representation presents the frequency of documents for specific administrative units. For Footprint based representations, there has been a higher mean response indicating positive feedback with respect to explorability and interactivity.

It is interesting to note that the level of interactivity and explorability has considerable influence on the users' perception of spatial relevance. Support for these inferences is drawn from the qualitative feedback obtained for the Footprint based representation where almost all of the 50 participants were successful in either listing partly or fully all the 5 documents with the smallest footprint and the largest footprint, respectively. Interestingly there were different documents listed for each of the options using the same visual representation (Footprint based representation).

This work can be extended to define a quantitative measure and draw insights between system generated Combined Relevance and what users perceive as spatially relevant.

9.7 Learnability of the SPIRIT prototype

Learnability is a key determinant of acceptance of the system by the users. A key finding of the experimental studies reveals that 94% of the participants either agree or strongly agree that it has been easy to get started with the system and initiate an information request (cf. Table 8.2). The success of any search interface has been through the demonstrated ease of usability and understandability, 88% of the participants either agree or strongly agree that the search interface has been easy to use and understand. Among many factors that influence the ease of understanding and learnability are the granularity and context of the underlying map data. 52% of the participants either strongly agree or agree that the map background was adequate enough to serve their purposes.

A prime functional requirement was the support for geographic prepositions and 80% of the participants have responded that the system provided support for interpreting the geographic prepositions and spatial relationships expressed in their queries. Varying results were obtained based on the use of different geographic prepositions within their query. A few users notedly indicated future directions by combining cardinal directions with neighbourhood search functionality, which is an illustration of a complex query.

9.8 Limitations of the experimental design

The experimental design process reported in Chapter 7 was based on iterative incorporation of suggested amendments during the preliminary pilot studies carried out with seven Ph.D. students working in the GIScience domain. The experiments were originally designed for a total of 60 candidates with equal numbers participating in each of four groups. Eventually, only 50 candidates participated in the study, of whom only 1/3 were female. The experiments were designed for primarily three study areas that had high resolution data available, namely Edinburgh, United Kingdom; Cardiff, United Kingdom; and Zürich, Switzerland. The limited scope of the underlying geographic data taken as study areas limited the ability to investigate other geographic prepositions and these visualisations present the information retrieved based on the specified criteria. Furthermore, it was difficult to document the time taken to accomplish each task due to the non-sequential and exploratory nature of the search process. In addition to the above, it was difficult to communicate real world task situations given to non-English speaking users, as they were described in English. Notedly, some users find it difficult to illustrate and express their search process in English during the think aloud sessions. Finally, as a research system, SPIRIT wasn't optimal in terms of efficiency and reliability, which has impaired its usability and probably in turn impaired the evaluation of the results.

9.9 Summary

Comprehending the results obtained from the experimental studies, this Chapter ran through the relevance of the research and supporting empirical studies detailing the necessity to investigate the problems outlined in Chapter 1. It built on discussing critical findings of the thesis and then to the statistical insights obtained from Chapter 8.

To summarize, the statistical insights shed light on subjective response to the use of different visualization methods adopted to serve the underlying characteristics of the data. With Map based representations as a fundamental approach to visualize the results, the other visual representation proved to be helpful to the users when used in conjunction with the Map based representations.

Furthermore, a significant insight is that system generated relevance and perceived relevance significantly vary when exploratory and interactivity techniques are coupled with the visualization.

Chapter 10

Conclusions

"Here's where things get interesting. We're at an inflecting point in the evolution of findability"

- Peter Morville (2005, p.2)

This thesis has elaborated on several aspects of extracting geographic information and contemporary geovisualization. To reiterate, the specific objective of this thesis was to investigate various approaches to representing geographic information that had been extracted from unstructured textual sources such as the internet and finally validate their practical use in relation to the type of representation adopted. Here the weaknesses of current approaches adopted for geovisualization were made explicit and at the same time our work clearly illustrated the benefits of using different approaches such as *map based representation, density based representation, footprint based representation* and *cartogram representation.*

10.1 Insights

This thesis has offered several significant research contributions relevant to the area of geographic information extraction and representation from unstructured textual sources such as *web documents*. For each of the research questions forming this thesis, the key findings are summarised below:

1. To what extent different information seeking models help us in defining the process of spatially aware information seeking ?
    - The "berry-picking model" proposed by Bates (1986) reflects the appropriate nature of the spatially aware information seeking process. Support for these inferences is drawn from the requirements gathering and think aloud sessions.
    - The spatially aware information seeking process can be defined as the ability to perceive (to see or recognize) space and retrieve the implicit knowledge which helps users define the position and orientation within the geographic space.

- The experiments were designed to achieve specific work tasks to achieve "Goal Posts/ Milestones" towards achieving the users' final objective. Analysis of the video sessions (of 1/4 participants) reveals differences in how the users achieved their objectives, reflecting the non-linear and exploratory nature of information seeking
- Different users expressed their needs that are essentially based on their understanding and knowledge about a specific geographic region. For instance a query "biking south of London" has been interpreted differently by different users and the graphical illustration of the query region varied from user to user. Similar insights were drawn for a query "*Hotels near Portree*". A significant number of users wasn't aware about the whereabouts of the place name "*Portree*" which is north-east of Scotland.

2. How does our understanding of the spatially aware information seeking process guide us in designing exploratory representations ?

   - The spatially aware information seeking process is primarily influenced by the user's implicit knowledge about a specific geographic region. The experimental studies reveal that users tend to express their queries for unknown locations such as Portree, Les Diablerets using a simple text query interface, while for known locations the prefer to specify their query region by graphically defining the query region as a polygon. Support for these inferences are drawn from the Thinkaloud sessions, where users when prompted to identify regions such as Portree switched to the simple text query interface, while for queries refering to London, users expressed their query region by drawing a polygon on the map.

3. Which of the exploratory representations assist users to identify geographic information retrieved from unstructured textual sources such as the internet ?

   None of the representations that were tested in this work seem to be capable of covering the full gamut of requirements of exploratory visualization of geographic information retrieval, but are in many ways complementary to the simple map based representations. Each of these visualization techniques serves varying functionalities and key insights drawn from the experimental studies reveal that

   - the density surface representation is the most widely accepted visual representations in comparison with other visual representations. Such inferences are supported by stronger response value favouring the density surface representation (ref. Fig: 8.3 - Usefulness, Interactivity and Explorability ; Information presentation & Understandability). This is due to their inherent characteristic that they present a comprehensive summary of the spread of the documents over a defined geographic region. While in Fig: 8.3 (ref. locate/associate/findability), density surface representation is least favoured to locate/associate/find individual footprints or document footprints.
   - the footprint based filtering representation is least understood and users tend to get most confused (ref. Fig: 8.3- Counter check for usefulness & understandability), while this is attributed to the lack of understanding of the concepts relating to "geographic extent of a footprint". Fig: 8.4 records a significant shift in the users' response when provided with adequate training (supervised session) and it has been easier to identify, locate, rank and understand the underlying relationships between footprints and documents. This is due to the ability to selectively prune through the results based on the geographic extent and the number of footprints.

- a stronger response value favouring map based representations and footprint based representations is attributed to the training effects and understanding of the concepts relating to footprint and document footprints. The map and footprint based filtering representations have been demonstrated as effective tools to explore through the result set based on the geographic extent and the number of footprints.
- for cartogram representations, the degree of abstraction and unfamiliarity was relatively higher than density surface representation.
- for cartogram representations, the learning effect was less pronounced.
- Map and Footprint based representations provided users with a relatively high exploratory visualization in comparison with Density and Cartogram based representation giving the flexibility to selectively prune the result sets based on the spatial extent or geographic scope of the footprints or the individual documents.
- in reference to working visual representations, the mock-up representations received very positive judgements by the test users, despite the fact that by definition mock-up representations are not functional. The lower response values can be attributed to the level of interactivity and explorability they offer.
    - With the association of symbology to illustrate a spatial activity based on the information available within a document listed according to relevance has drawn significantly high response, despite users failing to understand the concept of spatial hierarchy. This is supported by the recorded observations during the think aloud session.
    - For the isoline representations, the stronger response values can be attributed to the effective use of colour to illustrate regions of high geographic relevance in comparison to regions of low geographic relevance, which differentiates it from the density surface representation.

4. What metrics assess the efficacy of the visualization techniques adopted to represent geographic information retrieved from unstructured textual source such as the internet ?
    - The comprehensive task structure of the SPIRIT system was mapped on to the information and interaction typology proposed by Crampton (2002). Table 7.6 details the empirical criteria adopted for comparing four different visual representations (map, density, footprint and cartogram).
    - 11 empirical criteria are adopted to account for users' attention, motivation, learning, explorability and interactivity. These criteria are presented to the users in the form of a simple questionnaire.

Geographic relevance

To reiterate, one of the objectives of the experimental studies was also to investigate perceived relevance as compared to system generated relevance. At the core of any search process, the topic of contentions has been that of defining relevance. Different users had varying responses to what is to be perceived as a geographically relevant document collection presented as a textual list based on combined ranking (spatial + conceptual ranking; van Kreveld *et al.*, 2005). In total 38 participants out of 50 (i.e. 76%) identified geographically relevant documents which did not resemble the top 5 documents presented by the SPIRIT engine. This indicates that geographic relevance perceived by the users is considerably different from that generated by the SPIRIT engine. Moreover, this perceived relevance constantly differed between users when they visually explored using the advanced visualizations (Density, Footprint and Cartogram based representations).

Geographic scope

Figure 4.10 illustrates the nearest approximation of the characteristics of the data, which details numerous geographic references within a single web page. The web page refers to places such as Nepal, Norway, Alps etc. while the users information request has been to identify “hotels inside Scotland, U.K”. These geographic references within a web page may or may not relate to the candidate’s query topic being “hotels inside Scotland, U.K”. The footprint of such a document extends from Nepal to Greenland. Similarly, the concepts within a single web page refer to other concepts than those requested by the user, i.e. hotels. On interaction with the web page (system generated relevance web page), the users’ perception of relevance differs. By its very nature, the representation of these imprecise regions is very challenging.

10.2 Limitations

In this section, we detail the constraints that limited and influenced either directly or indirectly our research.

Evaluation

Since it was a research system, the SPIRIT prototype wasn’t optimized in terms of efficacy, usability and reliability, which impaired the evaluation of the results in several ways.

The experiments investigate the results at a coarser granularity while only three study areas had higher resolution available, namely Edinburgh (UK), Cardiff (UK) and Zurich (CH). Due to restrictions of the test data available (limited web collection and spatially indexed data), specific work tasks had to be assigned to the users during the experimental studies in order to prevent them from facing misbehaviours of the SPIRIT prototype, which is likely to happen if the users were allowed to try unrestricted queries. Otherwise, the query for “Indian restaurants in Les Diablerets” would simply fail because there are neither spatial data nor web documents available in the test database about this particular place and topic.

While the above point can be argued that having homogeneous simulated work tasks would help in uniform assessment of visualization based on the underlying data sets, it is to be noted that each document illustrates a hetrogeneous context (§ref. Fig 4.10) making it difficult for the users to define the relevance of the document in reference to their expressed information request.

Information Fallacy

Users acquiring information from unstructured textual sources such as web documents often encounter contradicting information resulting in misconceptions from incorrect reasoning. This may be due to the (lack of) reliability of the information source or the information contained within the web documents. This can be termed *information fallacy,* which may be created unintentionally, or may be created intentionally in order to impart misconceptions. One such illustration is Wikipedia, where misinformation may propagate although more recently, there have been efforts to improve the accuracy of the information by peer reviewing. In the case of spatially aware information retrieved from the web documents, the effect of information fallacy and its subsequent representation cannot be ignored. Future work should investigate this aspect, to what percentage of the web constitutes this misleading information and ways to eliminate information fallacy in geographic information retrieval.

Encoding Semantic Relationships between Geography and Conceptual terms

Existing gazetteers seldom encode semantic relationships between places and such information is least exploited by existing search engine facilities. The Getty Thesaurus of Geographic Names (TGN) encodes such relationships by recording hierarchical relationships between place names based on administrative and spatial features. TGN emphasizes more the relation between places than their actual location on the earth in terms of map coordinates. The gazetteers provide us with geometrical locational data confined to a simple footprint in the form of a centroid or a minimum bounding rectangle. Efforts should be made to exploit such information and populate the gazetteers with spatial relationships (such as distance, direction) based on their actual location on the earth in terms of map coordinates. A user searching for information about a location or place would often like to have information about them denoted by names that are super ordinate terms, subordinate terms, synonyms or related terms.

Other limitations

- It was difficult to document the time taken to accomplish each task in the experimental studies, due to the non-sequential and exploratory nature of the spatially aware information seeking process.
- It was difficult to communicate real world task situations (scenarios) to non-English speaking test users, as they were described in English.

Open Ends

Besides leading to some essential insights, this thesis still leaves some questions not completely answered . These include:

With diverse contextual information contained within a web page, geographic classification of a web collection or the results of a web search is a non-trivial task. It is also quite challenging to associate a specific activity based on a diverse set of activities illustrated within a specific web page, thus making it difficult to implement spatial activity maps. Future work, drawing upon the geographical classification of web pages can serve to build activity based thematic representations such as spatial activity maps, which were thus far only implemented as mock-ups in this thesis.

Interesting questions are also related to varying the footprint indexing strategies that are adopted, which would contribute to the accuracy of the results returned by the search engine. This can also be expected to have a direct impact on the way the visual representations are constructed. Hence, the visual representations are driven by the underlying footprint strategies adopted, requiring further investigation.

Among other feature enhancements discussed in this Chapter, the success of the spatially aware information seeking process primarily depends on the adoption of geographic query expansion techniques such as interactive feedback, interactive place name expansion (by using a spatial hierarchy) and incorporation of distance and direction based query expansion. This can be achieved by improving geographic named entity recognition techniques, their classification and mapping them onto real world spatial resolution.

Finally, while our empirical research has been able to clearly establish that there exist significant differences between the various techniques used to visually represent the results of spatially aware information seeking, we have still not provided a definitive visualization strategy. Future work will have to extend on our research and develop and implement a visualization strategy that dynamically adapts the representation method used to the search task that the user is trying to fulfil. Hence, future work will have to clearly determine which representation method works best under which circumstances.

10.3 Outlook

In recent years, the interest of the academic research community as well as search engine providers into geographical information retrieval has been soaring, making the work presented in this thesis highly relevant and timely. Yet, since geographical information retrieval as well as the associated methodologies are a very novel and developing field, there are naturally still quite a few areas where further improvement is necessary and possible. This research has laid the ground for future work, hence helping in this research endeavour.

Inclusion of Geographic Information

The surge of interest in GIS and geography sheds light on merging more geoweb services and integration of disparate data sets with a purpose to use the earth itself as an organizing metaphor for digital information. Thus, technology is driving us to integrate and manage geographic data to create a true *Digital model of the Earth*, that looks beyond; where any individual has the capability to integrate spatial data of any structured or unstructured type that goes beyond the base map of the world approach. The challenge here is to see to what extent does technology exponentially scale up/support the visualization function, which is the key to provide organized information to the end users without the problem of information overload in the design of on-line/on-demand mapping services.

The Spatio-temporal Dimension

Of the many characteristics that the information available on the internet describes, time is a key one. Time can be defined as either the date of creation of the web document or information which details the occurrence of a specific event at a geographic location. This can potentially be exploited by the users to answer spatio-temporal queries such as “hotels within 5 minutes walk from Haupt-bahnhof, Zurich”. Future work could investigate spatio-temporal aspects of the geographic information and its visual representation.

News corpuses are often rich with spatio-temporal indications of events occurring at a specific geographic location. These events are often associated with bursts (many occurrences in a short period) of news stories. The events documented in the news corpus either have occurred in the past at a specific instance of time or are occurring at present or will occur in the future (Yang *et al.*, 1999). These spatio-temporal events referring to a geographic location occur for short durations or sometimes without defined occurrence in time, often referred to as “aoristic”, a term used by Ratcliffe (2002). Support for these arguments is drawn from the works being carried out by Yang *et al.* (1999), Ratcliffe (2002), Herzog (2003), and Tezuka and Tanaka (2005). Furthermore, the news corpus is considered to be highly reliable source of information as it documents events at a specific geographic location at a specific instance of time.

Landmark Extraction from Corpus

A continuation of this research should also analyse and develop methods for the extraction of landmarks relating to a specific geographic location from a corpus. This was suggested during the experimental studies when some participants indicated that their queries often associate less significant geographic objects to more significant and outstanding geographic objects. Such objects are often termed landmarks. Theoretical grounding can be drawn from the works of Couclelis *et al.* (1987) in their anchor-point theory that illustrates how humans tend to construct hierarchical structures to relate insignificant geographic objects using the neighbourhood of significant hierarchical objects. Future work should investigate what percentage of the queries actually refer to landmarks and how we can present localised visual representations, orienting users in reference to the chosen or illustrated landmarks in the query.

10.4 Closing remarks

This thesis has been embedded in the SPIRIT project in which six European research groups explored new avenues of finding and extracting geographic information from unstructured textual sources such as web documents. In this project the research consortium also built a prototype system that implements the spatially aware information seeking process using a web document collection of around one terabyte. Making use of this prototype, this thesis brought about qualitative evidence that indicates that the "berry-picking model" is the appropriate model to reflect the nature of the spatially aware search process. Most importantly, experimental studies have been carried out that yielded meaningful insights with respect to the use of different geovisual representations adopted to visualize the geographic information retrieved from the internet and their significant differences. Finally, the study suggests that no single visualization method is capable of covering the entire spectrum of requirements of visualization methods for this purpose, yet the results also indicate that the different methods are complementary to each other.

References

Abdelmoty, A. I., Smart, P. D., Jones, C. B., Fu, G and Finch, D. 2005. A critical evaluation of ontology languages for geographic information retrieval on the internet. Journal of visual languages & computing, 16 (4), 331-358.

Adamic, L. 2001. Network Dynamics: The World Wide Web. PhD Thesis. Stanford University, June, 2001.

Agarwal, P. 2004. Contested Nature of Place: Knowledge mapping for resolving ontological distinctions between geographical concepts. Third International Conference GIScience, Springer-Verlag. Max J. Egenhofer, Christian Freska and Harvey J. Miller (eds.) (Oct. 2004), 1-22.

Agarwal, P. 2005. Ontological considerations in GIScience. International Journal of Geographical Information Science, 19(5), 501-536.

Ageno, A., Ferrés, D., Rodríguez, H. 2005. The Gestalt-IR System at GeoCLEF-2005: Experiments Using a QA-based IR System, Linguistic Analysis, and a Geographical Thesaurus. In: Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF. Sep. 2005, Vienna, Austria.

Agnew, J. A. 1987. Place and politics: the geographical mediation of state and society, Boston and London: Allen and Unwin.

Ahlberg. C. and Shneiderman. B. 1994. Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays. In proceedings of the conference on Human Factors in Computing Systems (CHI'94), pages 313-317. ACM Press.

Ahlqvist, O. and Arnberg, W. 1997. A theory of geographic context, the fifth dimension in geographical information. In: H. Hauska ed., Proceedings of the 6th Scandinavian research conference on Geographical Information Systems, Stockholm, Sweden, SanGIS'97, 67-84.

Allan, J. 1995. Automatic Hypertext Construction. Ph.D. Thesis, Cornell University, January, 1995.

Alani, H. 2003. TGVizTab: An ontology visualization extension for Protégé. In Proceedings of Knowledge Capture (K-Cap'03), Workshop on Visualization Information in Knowledge Engineering, Sanibel Island, Florida.

Alani, H., Jones, C. and Tudhope, D. 2001. Voronoi-Based Region Approximation for Geographical Information Retrieval with Gazetteers. International Journal of Geographical Information Science, 15 (4). 287-306. ISSN 1365-8816.

Allen, B. R. 1995. Timelines as information system interfaces. In proceedings of International Symposium on Digital Libraries. 175-180.

Allen, B. 1996. Information tasks. Towards a user-centered approach to information systems. San Diego: Academic Press.

Allen, B. R. 2004. A Query Interface for an Event Gazetteer. JCDL'04.ACM, June 7-11, 72-73.

Allen, J. F. and Perrault, C. R. 1980. Analyzing Intention in Utterances. Artificial Intelligence. 15:143-178.

Alonso, D. L. 1998. The effects of individual differences in spatial visualization ability on dual-task performance. Doctoral Dissertation, University of Maryland.

Altman, D. 1994. Fuzzy set theoretic approaches for handling imprecision in spatial analysis. International Journal of Geographic Information Sciences. vol. 8. 271-289

Amar, A. R. and Stasko, J. T. 2005. Knowledge Precepts for Design and Evaluation of Information Visualization. IEEE Transactions on Visualization and Computer Graphics. July/August 2005. Vol. 11 No.4. 432-442.

Amitay, E., Har'El, N., Sivan, R., and Soffer, A. 2004. Web-a-Where: Geotagging web content. SIGIR'04. July 25-29. Sheffield, South Yorkshire, U.K., 273-280.

Amitay E. 2001. What Lays in the Layout: Using anchor-paragraph arrangements to extract descriptions of Web documents. PhD Thesis, Macquarie University, February, 2001.

Amrhein, C. G. and Reynolds, H. 1996. Using spatial statistics to assess aggregation effects. Geographical Systems, 2, 83-101.

Amrhein, C. G. and Reynolds, H. 1997. Using the Getis statistic to explore aggregation effects in Metropolitan Toronto Census data. The Canadian Geographer, 41(2), 137-149.

Amrhein, C. G. 1993. Searching for the elusive aggregation effect: Evidence from statistical simulations. Environment and Planning A, 27, 105-119.

Ana Paula Dutra de Aguiar, Gilberto Câmara, Antônio Miguel Vieira Monteiro, Ricardo Cartaxo Modesto de Souza. 2003. Modelling Spatial Relations by Generalized Proximity Matrices. V Simpósio Brasileiro de GeoInformática - Campos do Jordão, SP, Brasil.

Ando, R. K., Boguraev, K. B., Byrd, J. R. and Neff, S. M. 2000. Multi-document summarization by visualizing topical content. ANLP-NAACL. Workshop on Automatic Summarization.

Andrews, K. and Heidegger, H. 1998. "Information Slices: Visualizing and exploring large hierarchy using cascading, semi-circular discs". In: Proc. of IEEE Information Visualization (Late Breaking Hot Topics Procs), 1998, pages 9-12, Raleigh Durham, North Carolina, October 1998. IEEE Computer Society. available online at http://www.iicm.edu/liberation/iicm_papers/ivis98.pdf viewed on 18th March 2003.

Andrienko, G. L. and Andrienko, N. 1998. Intelligent visualization and dynamic manipulation: Two complementary instruments to support data exploration with GIS. In proceedings advanced visual interface '98. (New York: ACM Press).

Andrienko, G. L. and Andrienko, N. 1999. Interactive maps for visual data exploration. International Journal of Geographical Information Science, Vol. 13. No. 4. 355-374.

Andrienko. N. and Andrienko. G. L. 2006. Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach. Springer-Verlag. Berlin. Heidelberg.

Angluin, D. and Smith, C. H. 1983. Inductive Inference: Theory and Methods. ACM Comput. Surv. 15, 3 (Sep. 1983). 237-269. DOI=http://doi.acm.org/10.1145/356914.356918

Annett, J. and Stanton, N. (eds.) 2000. Task Analysis. London. Taylor & Francis.

Annett, J. 2003. Hierarchical Task Analysis. In Holnagel, E., Handbook of Cognitive Task Design, Chapter 2, 17-35. Mawhah NJ: Lawrence Erlbaum.

Anselin,. L. 1995. Local indicators of spatial association- LISA. Geographical Analysis. 27. 93-115.

Antes, J. R. and K. Chang. 1990. An empirical analysis of the design principles for quantitative and qualitative area symbols. Cartography and Geographic Information Science 17(4): 271-77.

Arampatzis, A. and van Kreveld, M. 2004. Practical Similarity Ranking. SPIRIT Technical Report

- D18 5302. Available online at http://www.geo-spirit.org

Arampatzis, A., van Kreveld, M. and Reinbacher, I. 2004. Abstract Multi-attribute Similarity Ranking. SPIRIT Technical Report - D17: 5301. Available online at http://www.geo-spirit.org

Arampatzis, A., van Kreveld, M. J., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P. D., Joho, H., Sanderson, M., Benkert, M and Wolff, A. 2004. Web-based delineation of imprecise regions (abstract). In workshop on Geographic Information Retrieval (with SIGIR).

Arampatzis, A., van Kreveld, M. J., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P. D., Joho, H. and Sanderson, M. 2005. Web-based delineation of imprecise regions. UU-CS (Int. rep. 2005-036), UU WINFI informatica en Informatiekunde.

Arampatzis, A., van Kreveld, M., Reinbacher, I., Jones, C. B., Vaid, S., Clough, P. D., Joho, H. and Sanderson, M. 2006. Web-based delineation of imprecise regions. In the Journal of Computers, Environment and Urban Systems (CEUS).

Aretz, A. J. and Wickens, C. D. 1992. The mental rotation of map displays. Human Performance. 5, 303-328.

Armstrong, M. P., Rushton, G. and Zimmerman, D. L., 1999. Geographically masking health data to preserve confidentiality. Statistics in Medicine. 18. 497-525.

Armstrong, M. P. 2002. Geographic information technologies and their potentially erosive effects on personal privacy. Studies in the Social Sciences. 27. 19-28.

Arnheim. R. 1997. Visual thinking. University of California Press. Berkeley 1969. renewed 1997.

Aronoff, S. 1989. Geographic Information Systems: A Management Perspective. WDL Publication, Ottawa.

Arpinar, I, B., Seth, A. and Ramakrishnan, C. 2004. Geospatial Ontology Development and Semantic Analytics. Handbook of Geographic Information Science, In J. P. Wilson and A. S. Fotheringham eds. Blackwell Publishing.

Artandi, S. 1973. Information concepts and their utility. Journal of the American Society for Information Science, 24(4): 242-245.

Assa, J., Cohen-Or, D., and Milo, T. 1997. Displaying data in multidimensional relevance space with 2D visualization maps. In Proceedings of the 8th Conference on Visualization '97 (Phoenix, Arizona, United States, October 18 - 24, 1997). R. Yagel and H. Hagen, Eds. IEEE Visualization. IEEE Computer Society Press, Los Alamitos, CA, 127-ff.

Attneave, F. 1959. Applications of Information Theory to Psychology: A Summary of Basic Concepts, Methods and Results. Holt-Dryden, New York.

Axelrod, A. E. 2003. On building a high performance gazetteer database. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 63-68, Alberta, Canada: ACL.

Axhausen, K. W. 2000. Activity based modelling: Research directions and possibilities. Draft paper in D. Simmonds and J. J. Bates (2000) New look at Multi-modal Modelling report from the department of Environment, Transport and the Regions, London, Cambridge and Oxford.

Bade, R., Schlechtweg, S., and Miksch, S. 2004. Connecting time-oriented data and information to a coherent interactive visualization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM Press, New York, NY, 105-112. DOI= http://doi.acm.org/10.1145/985692.985706.

Baik, S., Bala.J., and Hadjarian, A., 2001. Inferring and Visualizing Classification Rules. Fairfax, VA: George Mason University: School of Information Technology and Engineering.

Balley, S., Bucher, B., Petrelli, D., Ruas, A., van kreveld, M., Sanderson, M., Sester, M., 2002. User Requirements Specification, SPIRIT deliverable Report D3-7101.

Balabanovic, M. 1998. Learning to Surf: Multiagent Systems for Adaptive Web Page Recommendation. PhD Thesis. Stanford University Department of Computer Science. March'98.

Baldonado, M. Q. W. 1997. An Interactive, Structure-Mediated Approach to Exploring Information in a Heterogenous, Distributed Environment. Ph.D. Dissertation, Stanford University, December, 1997.

Baldonado, M., Woodruff, A., Kuchinsky, A., 2000. Guidelines for using multiple views in information visualization, Proc. ACM Advanced Visual Interfaces `00, May. http://cite-seer.ist.psu.edu/wang-baldonado00guidelines.html

Bamber, D. 1979. State trace analysis: A method of testing simple theories of causation. Journal of mathematical psychology. 19. 137-181.

Barclay, T., Gray, J., and Slutz, D. 1999. Microsoft TerraServer: A Spatial Data Warehouse. Micorsoft Research: Technical Report MS-TR-99-29.

Bar-Ilan, J. 1998/9. Search Engine Results over Time- A case study on search engine stability. International Journal of Scientometrics, Informetrics and Bibliometrics. Vol. 2/2: 1-1. Available online at http://cybermetrics.cindoc.csic.es/pruesbas/v2i1p1.htm.

Barker, J. C. 2005. The major vernacular regions of Tennessee. Masters Thesis. Marshall University. December.

Barry, C. L. 1994. User-defined Relevance Criteria: An Exploratory Study. Journal of the American Society for Information Science, 45(3): 149-159.

Bartz, B. S. 1970. Experimental use of the search task in an analysis of type legibility in cartography. Cartographic Journal. 7(2): 103-112.

Bates, M. J. 1986. An exploratory paradigm for online information retrieval. In: Brookes, B.C. (ed.) Intelligent Information Systems for the Information Society. Amsterdam: North-Holland.

Bates, M. J. 1989. The design of browsing and berry-picking techniques for the on-line search interface. Online Review, 13(5): 407-431.

Battele, J. 2005. The Search: How google and its rivals rewrote the rule of business and transformed our culture. Nicholas Brealey Publishing, United Kingdom.

Batty, M. 2002. Michael Batty's discussion of the city as clusters of spatial events in Michael Batty Thinking About Cities as Spatial Events, Environment and Planning B: Planning and Design, 129(1), Jan, 1-2.

Bavoil. L., Callahan. S. P., Scheidegger. C. E., Vo. H. T., Crossno P. J., Silva C. T., Freire J, 2005. VisTrails: Enabling Interactive Multiple-View Visualizations, vis, p. 18, IEEE Visualization 2005 - (VIS'05).

Beaudoin, L., Parent, M.-A. and Vroomen, L. C. 1996. CHEOPS: A Compact Explorer for Complex Hierarchies. Proc. of IEEE Visualization '96 (October 27 - November 1, San Francisco, USA), Computer Society Press, 87-92.

Beaulieu, M., Robertson, S. and Rasmussen, E. 1996. Evaluating interactive systems in TREC. Journal of the American Society for Information Science, 47(1), 85-94.

Beaulieu, M. 1997. Experiments with interfaces to support query expansion. Journal of Documentation, 53(1), 8-19.

Beale. R., McNab. R. J. and Witten. I. H., 1997. Visualizing sequences of queries: a new tool for information retrieval. First international conference on information visualization(IV'97). iv, 57.

Becker. R. A. and Cleveland. W. S. 1987. Brushing Scatterplots. Technometrics. 29. 127-42. Also in Dynamic Graphics for Statistics. Eds. W.S.Cleveland and M.E. McGill. Belmont, CA: Wadsworth. 201-224.

Bederson, B. B. and Hollan. J. D. 1994. Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics. Proc. UIST, 17-26.

Bederson, B. B., Meyer., J., and Good. L., 2000. Jazz: An extensible Zoomable user interface graphics toolkit in java. UIST 2000. ACM Symposium on user interface software and technology. CHI Letters, 2(2), 171-180.

Beeferman, D. and Berger, A. 2000. Agglomerative clustering of a search engine query log. In Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Boston, Massachusetts, United States, August 20-23, 2000) KDD'00. ACM Press, New York, NY, 407-416. DOI=http://doi.acm.org/10.1145/ 347090.347176.

Beitzel, S. M. 2006. On Understanding and Classifying Web Queries. PhD Thesis, Illinois Institute of Technology. May 2006.

Beitzel, St., Jensen, E., Chowdhury, A. and Grossman, D. 2004. Hourly Analysis of a very large topically categorized web query log. Proceedings of the 2004 ACM Conference on Research and Development in Information Retrieval (ACM-SIGIR), Sheffield, UK, July'04.

Belzer, J. 1973. Information theory as a measure of content. Journal of the American Society for Information Science, 24(4): 300-304.

Belkin, N. J., and Robertson, S. E. 1976. Information science and the phenomenon of information. Journal of the American Society for Information Science, 27: 197-204, July-August.

Belkin, N. J., Oddy, R. N. and Brooks, H. M. 1982. ASK for Information Retrieval: Part I. Background and history. Journal of Documentation, 38(2), 61-71.

Belkin, N. J. Cool, C., Stein, A. and Thiel, U. 1995. Cases, Scripts and information seeking strategies: On the design of interactive information retrieval systems. Expert Systems with Applications. 1995. Vol.9 (3), 379-396.

Bellardo, T. 1985. An investigation of online searcher traits and their relationship to search outcome. Journal of the American Society for Information Science, 36(4), 241-250.

Benelli, G., Caporali, M., Rizzo, A. and Rubegni, E. 2001. Design concepts for learning spatial relationships. In Proceedings of the 19th Annual International Conference on Computer Documentation (Sante Fe, New Mexico, USA, October 21-24, 2001). SIG-DOC'01. ACM Press, New York, NY, 22-30. DOI=http://doi.acm.org/10.1145/501516.501522.

Benjamin, B. and Shneiderman, B. 2000. The Craft of Information Visualization: Reading and Reflections. The Morgan Kaufmann Series in Interactive Technologies.

Berberich, K. L. 2004. Time-aware and trend-based authority ranking. Masters thesis, Universität des Saarlandes, Germany, November 2004.

Berendt, B., Barkowsky, T., Freska, C. and Kelter, S. 1998. Spatial representation with aspect maps. In C. Freksa, C. Habel,C. & K.F. Wender (eds.). Spatial Cognition- An inter-disciplinary approach to representing and processing spatial knowledge. 313-336. Berlin:Springer.

Bergman, L. D., Rogowitz, B. E and Treinish, A. L. 1995. A rule-based tool for assisting colour map selection. IBM Research. IBM Thomas J. Watson Research Center, Visualization, 1995. Visualization'95. Proceedings of IEEE Conference, 118-125. 444, Available online at http://www.research.ibm.com/dx/proceedings/pravda/

Bergman, M. K. 2001. The Deep web: Surfacing Hidden Value. The Journal of Electronic Publishing. August 2001. Volume 7, Issue 1. Available online at http:// www.press.umich.edu/jep/07-01/bergman.html

Berners-Lee, T. and Fischetti, M. 1999. Weaving the web: the original design and the ultimate destiny of the world wide web by its inventor Tim Berners-Lee with Mark Fischetti. San Francisco: Harper.

Berners-Lee, T., Hedler, J. and Miller, E. 2001. The Semantic Web. Scientific American, 35-49.

Berry, M. W. and Browne, M. 1999. Understanding search engines: Mathematical modeling and text retrieval. SIAM, ISBN:0898714370.

Bertin, J. and Berg, W. J. 1983. Semiology of Graphics: Diagrams, Networks, Maps. The University of Wisconsin Press Ltd.

Bertoline, G. R. 1988. The implications of cognitive neuroscience research in spatial abilities and graphics instruction. Proceedings of the International Conference on Engineering Design Graphics. Vienna (1), 28-34.

Bier. E. A., Stone, M. C., Pier, K., Buxton, W., and DeRose, T., 1993. Tool glass and magic lenses: the see-through interface. Proceedings SIGGRAPH'93. Anaheim, CA, 73-80.

Bilhaut, F., Charnois, T., Enjalbert, P., and Mathet, Y. 2003. Geographic reference analysis for geographic document querying. In Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References - Volume 1 Human Language Technology Conference. Association for Computational Linguistics, Morristown, NJ, 55-62.

Bilhaut, F., et al. 2003. Geographic reference analysis for geographic document querying. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 55-62, Alberta, Canada: ACL.

Bivand, R. 1998. A review of spatial statistical techniques for location studies. Prepared in connection with CEPR symposium on New Issues in Trade and Location (2277), Lund, Sweden, 28-30 August, 1998.

Björk, S. 2000. Hierarchical flip zooming: enabling parallel exploration of hierarchical visualizations. In Proceedings of the Working Conference on Advanced Visual interfaces (Palermo, Italy). AVI '00. ACM Press, New York, NY, 232-237. DOI= http:// doi.acm.org/10.1145/345513.345324

Blaut, J. M. 1999. Maps and Spaces. Professional Geographer, 51(4), 510-515.

Blickle, G. 1996. Personality traits, learning strategies, and performance. European Journal of Personality, 10, 337-352.

Blok, A. C. 2005. Dynamic Visualization variables in animation to support monitoring of spatial phenomena. PhD Thesis. Institute of Geo-Information Science and Earth Observation. Available Online at http://igitur-archive.library.uu.nl/dissertations/2005-0520-200915/

Blustein, W. J. 1999. Hypertext Versions of Journal Articles: Computer-aided linking and realistic human-based evaluation. PhD thesis, University of Western Ontario, April 1999.

Board, C. and R. M. Taylor. 1977. Perception and maps: Human factors in map design and interpretation. Transactions of the Institute of British Geographers, New Series 2: 19-36.

Bollmann, J. and Tainz, P. 2001. Ikonizitat, lexicon der kartographie und geomatik, Heidelberg, vol.1, 391-392.

Bollmann, J. 2002. Argument, Diskurs, Kartographische Informationsverarbeitung, Kol-laboratives Arbeiten, Lexikon der Kartographie and Geomatik, Heidelberg, Vol.1, 31-166. vol.2, 26-77.

Bonnel. N., and Chevalier. M., 2006. Evaluation des Interfaces Utilisateur d'Information. 4e atelier Visualization et Extraction de Connaisances - EGC 2006.

Boote, D. N. and Beile, P. 2005. Scholars before Researcher: On The Centrality of Dissertation Literature Review In Research Preparation. Educational Researcher, 34(6), 3-15.

Borgman, C. L. 1996. Why are online catalogs still hard to use? Journal of the American Society for Information Science. 47(7): 493-503.

Borlund, P. and Ingwersen, P. 1997. The development of a method for the evaluation of interactive information retrieval systems. Journal of Documentation, 53(3), 225-250.

Borlund, P. 2000. Evaluation of interactive information retrieval systems. Abo(Turku) Finland: Abo Akademi University Press (Doctoral Dissertation).

Borlund, P. 2003. The IIR evaluation model: a framework for evaluation of interactive in-formation retrieval systems. Information Research, 8(3). paper no. 152. Available at http://informationr.net/ir/8-3/paper152.html.

Borghoff, U. M., Chevalier, P-Y, and Willamowski, J. 1996. Adaptive Refinement of Search Patterns for Distributed Information Gathering. In: Verbraeck, A. (ed.): Proc. Int. Conf. EuroMedia/WEBTEC, London, U.K. December, 19-21. San Diego: The Society for Computer Simulation, 5-12.

Boughman, A. 2005. Realism and Connotations of Data Quality in Computer Displayed Maps. Master's Thesis. University of California.

Bowman, D. A., Koller, D. and Hodges, L. F. 1997. Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques. IEEE Proceedings of Virtual Reality Annual International Symposium (VRAIS). 7, 45-52. http://citeseer.ist.psu.edu/bowman97travel.html

Box, G. E, Hunter, W. G., and Hunter, J. S., 2005. Statistics for Experimenters: Design, Innovation, and Discovery, 2nd Edition, Wiley, ISBN: 0471718130

Bra Paul de. 2004. Adaptive hypermedia and adaptive web systems: third international conference proceeding/AH. The Netherlands. Aug 23-26. Springer (Berlin).Lecture Notes in Computer Science: 3137. ETH(ZH): BIB: 713677:3137.

Brath, R. 1997. Metrics for effective information visualization, infovis, p. 108, 1997 IEEE Symposium on Information Visualization (InfoVis '97).

Brennan, N. and Llyod, R. 1993. Searching for boundaries on map: The cognitive process. Cartography and Geographic Information Science. 20 (4): 222-236.

Brewer, C. A., 1990. The Effect of Color on the Perception of Map Scale, in Student Honors Competition Winning Papers, edited by E. Wingert, Proceedings of the Cartography Specialty Group, Association of American Geographers, Toronto, 1-12.

Brewer, C. A., 1994. Colour use guidelines for mapping and visualization. In Visualization in Modern Cartography (eds A. M. MacEachren and D.R.F. Taylor), 123-147. Oxford: Pergamon.

Brewer, C. A., MacEachren, A., Pickle, L. and Herrmann, D. 1997. Mapping mortaility: Evaluating colour schemes for choropleth maps. Annals of the Association of American Geographers 87: 103-134.

Brinck, I. and Gardenfors, P. 1999. Representation and Self-Awareness in international agents. Synthese. Springer Netherlands, Vol.118, No.1. Jan '99. 89-104.

Brinkman, Willem-Paul. 2003. Is usability compositional? PhD Thesis. Technische Universiteit Eindhoven.

Broadbent, D. E. 1958. Perception and communication. London: Pergamon Press.

Broers, R. and Sebo, L. 2002. Production of the Gazetteer of Australia. Presented at CGNA Annual Meeting Rydges, Caloundra, 21-22. March'02.

Brundson, C. 1998. Exploratory Spatial Data Analysis and Local Indicators of Spatial Association with XLISP-STAT. The Statistician. Vol.47. No.3. 471-484.

Bucher, B., Balley, S., Levin, S., Syed., A. K., Petrelli, D., Weibel. R., Abdelmoty, A. I., Beaulieu, M., Heinzle, F., 2004. User Requirements Specification Reassesment. Technical Report-D3 7101B. Available online at http://www.geo-spirit.org/publications/SPIRITWP7_D3B_final.pdf

Bucher, B., Clough, P., Finch, D., Joho, H., Purves, R. and Syed, A. K. 2005. Evaluation of SPIRIT prototype following integration and testing. Technical Report-D31 7301. Available online at www.geo-spirit.org/publications/SPIRIT_D31_WP7.pdf

Buckland, M. K. 1991. Information as thing. Journal of the American Society for Information Science. 42(5): 351-360.

Buja, A., Cook. D., Swayne, D. F. 1996. Interactive High-Dimensional Data Visualization. Journal of Computational and Graphical Statistics. 5(1). March. 78-99.

Buja, A., Swayne, D. F., Littman, M., Dean, N., and Hoffmann. 2001. XGvis: Interactive Data Visualization with Multidimensional Scaling. Journal of Computational and Graphical Statistics.

Bunch, R. and Lloyd, R. 2006. The cognitive load of geographic information. The Professional Geographer. 58(2): 209-220.

Buono, P., Aris, A., Plaisant, C., Khella, A., and Shneiderman, B., 2005. Interactive Pattern Search in Time Series. Proceedings of Conference on Visualization and Data Analysis, VDA 2005, SPIE, Washington DC. 175-186.

Burgess, N., Becker, S., John, A., King and John O' Keefe. 2001. Memory for events and their spatial context: models and experiments. Phil Tarns. Soc. London. B 356, 1493-1503.

Burnett, G., Smith, D. and May, A. 2001. Supporting the navigation task: Characteristics of "Good" Landmarks. In M.A. Hanson, ed., Contempory Ergonomic 2001", Taylor and Francis, 441-446.

Burrough, P. A. 1986. Principles of Geographical Information Systems for Land Resource Assessment, Oxford: New York.

Butler, D. 2006. Mashups mix data into global service. Is this the future for scientific analysis? Nature. Published online : 4th January. Doi: 10.1038/439006a. http://www.na-ture.com/news/2006/060102/pf/439006a.html.

Buttenfield B. P., and Mackaness W. A. 1991. Visualization, In: Maguire, D. J. Goodchild., M. F. and Rhind. D. W. edited. Geographical Information Systems: Principle and Application, Longman, London. 427-443.

Buyukkokten, O., Cho, J., Garcia-Molina. H., Gravano, L., Shivakumar, N., 1999. Exploiting geographical location information of web pages. In WebDB'99 (with ACM SIGMOD'99), 91-96.

Byström, K. 1999. Task complexity, information types and information sources: examination of relationships. Tampere, Finland: University of Tampere. Doctoral dissertation, Acta Universitatis Tamperensis; 688.

Byström, K. and Järvelin, K. 1995. Task complexity affects information seeking and use. Information Processing & Management, 31(2), 191-213.

Cai, G. 2002. GeoVSM: An integrated retrieval model for geographic information. In M. J. Egenhofer and D. M. Mark (eds.): GIScience, LNCS 2478, 65-79.

Cai, G. and Xue, Y. 2006. Activity-oriented Context-aware Adaptation Assisting Mobile Geo-spatial

Activities. ACM IUI'2006 – International Conference on Intelligent User Interface, Sydney, Australia, January 29 – February 1, 2006.

Cailleteau, L. 1999. Interfaces for visualizing multi-valued attributes: Design and Implementation using starfield displays. Diplöme d'Etudes Approfondies. HCI-University of Maryland (USA) & Ecole des Mines de Nantes (France).

Callaway, C. B. 2003. Integrating discourse markers into a pipelined natural language generation architecture. Proceedings of the 41st Annual Meeting for the Association for Computational Linguistics, July, 264-271.

Canter, D. 1997. The facets of place. In Moore G. T. and Marans R. W. (eds.), Advances in Environment, Behaviour, and Design, Vol.4: Towards the Integration of Theory, Methods, Research and Utilization. NewYork: Plenum, 109-147.

Car, A. 1997. Hierarchical Spatial Reasoning: Theoretical Consideration and its Application to Modelling Wayfinding. PhD Dissertation, Department of Geoinformation, Technical University, Vienna.

Card, S. K., Robertson, G. G. and York. W., 1996. The webbook and the web forager : An information workspace for the world wide web. In proceedings of CHI'96. ACM conference on Human Factors in Software. 111-117. ACM Press.

Card, S. K. and Mackinlay, J. D., 1997. The structure of the information visualization design space. In Proceedings of the IEEE Symposium on Information Visualization '97, pages 92--99, 125. IEEE.

Card, S. K., Jock, D. M. and Schneiderman, B. 1999. Reading in information visualization- using vision to think. Morgan Kaufmann Publishers, Inc. USA.

Carmo, M. B., Freitas, S., Afonso, A. P. and Claudio, A. P. 2005. Filtering Mechanisms for the visualization of Geo-referenced Informaiton. GIR'05. ACM, Nov'4, Bremen, Germany.

Carr, D. B. and Pierson, S. M. 1996. "Emphasizing Statistical Summaries and Showing Spatial Context with Micromaps," Statistical Computing & Graphics Newsletter, 7(3): 16-23.

Carr, D. B., Olsen, A. R., Courbois, J. P., Pierson, S. M., and Carr, D. A. 1998. "Linked Micromap Plots: Named and Described," Statistical Computing & Graphics Newsletter, 9(1): 24-32.

Carr, D. B., Olsen, A. R. Pierson, S. M., and Courbois, J .P. 2000. Using linked micromap plots to characterize Omernik ecoregions. Data Mining and Knowledge Discovery 4: 43-67.

Carr, D. B, Wallin, J. F, and Carr, D. A. 2000. "Two New Templates for Epidemiology Applications. Linked Micromap Plots and Conditioned Choropleth Maps," Statistics in Medicine . 19: 2521-2538.

Carr, D. B. 2001. Designing Linked Micromap Plots for States with many Counties," Statistics in Medicine . 20: 1331-1339.

Cartwright, W. E. 1999. Extending the map metaphor using web delivered multimedia. In-ternational Journal of Geographical Information Science 13(4): 335-353.

Casakin, H., Barkowsky, T., Klippel, A., and Freksa, C. 2000. Schematic maps as way finding aids. In C. Freksa, W. Brauer, C., Habel, & K.F. Wender (Eds.). Spatial Cognition II- Integrating abstract theories, empirical studies, formal models, and practical applications. 54-71. Berlin: Springer.

Casati, R., Varzi, A. eds. 1996. Events. Dartmouth, Aldershot.

Casner, S. M. 1991. A Task-Analytic Approach to the Automated Design of Graphic Presentations, ACM Transactions on Graphics. April, 10(2), 111-151

Casner, S. 1991. BOZ: Task-analytic design and evaluation of visual displays. In Proceedings of 1991 IEEE International Conference on Systems, Man and Cybernetics, 1195-1200.

Castillo C. 2004. Effective Web Crawling. PhD Thesis, Dept. of Computer Science, University of Chile.

November 2004.

Caudra, C. A. and Karter, R. V. 1976. Experimental studies of relevance judgements. Final report. Vol.I: Project Summary (NSF Report No: TM-3520/001/00). Santa Monica, CA: System Development Corp.

Cave, K., and Wolfe, J. 1990. Modelling the role of parallel processing in visual search. Cognitive Psychology. 22(2): 225-271.

Cecconi, A. and Galanda, M. 2002. Adaptive Zooming in Web Cartography. Computer Graphics Forum, 21(4): 787–799.

Chalmers, M. 2001. Place, media and activity. SIGGROUP Bull., 22(3), 38-42.

Chaves, Marcirio Silveira; Martins, Bruno; Silva, Mário J. 2005. Challenges and Resources for Evaluating Geographical IR. In: Proceedings of the 2nd International Workshop on Geographic Information Retrieval, CKIM 2005. Nov. 2005, Bremen, Germany. 65-69.

Chen, C. and Czerwinski, M. 1997. Spatial ability and visual navigation. New Review of Hypermedia and Multimedia, 3. 67-89.

Chen, H., Smith, T. R., Larsgaard, M., Hill, L. L., and Ramsey, M. 1997. A Geographic Knowledge Representation System for Multimedia Geospatial Retrieval and Analysis, In-ternational Journal on Digital Libraries, Vol.1, No.2, 32-152.

Chen, L. and Sycara, K. 1998. A Personal Agent for Browsing and Searching. Proceedings of the 2nd International Conference on Autonomous Agents (Agents'98), ACM Press, New York, 132-139. http://citeseer.ist.psu.edu/article/chen98webmate.html

Chen, C. 2000. Individual difference in spatial-semantic virtual environment, Journal of the American Society for Information Science, 51(6): 529-542.

Chen, C and Yue, Y. 2000. Empirical studies of information visualization: A meta-analysis. International Journal of Human-Computer Studies. 53(5), 851-866.

Chen, C., Cribbin, T. and Macredie, R. D. 2002. Effects of Spatial-Semantic Interfaces in Visual Information Retrieval: Three Experimental Studies (Library and Information Commission Research Report 137). London: Resource, the Council for Museums, Archiv)

Chen, C., Cribbin, T., Kuljis, J. and Macredie, R. 2002. Footprints of information foragers: behaviour semantics of visual exploration. International Journal of Human Computer Studies. 57, 139-163. Available online at http://www.idealibrary.com.on

Chen, H. 2003. Compound Brushing, infovis, IEEE Symposium on Information Visualization, 23.

Chen, H, M, C. 2003. Comparison of three vertical search spiders. Computer. IEEE Computer Society. Available online at http://ai.bpa.arizona.edu/go/intranet/papers/Compari-son%20of%20Three%20Vertical%20Search%20Spiders.pdf

Chen, Z. and Xu, Y. 2005. User-oriented relevance judgement: A conceptual Model. Proceedings of the 38th Hawaii International Conference on System Sciences. IEEE.

Cheng, T., Moelnaar, M. and Lin, H. 2001. Formalizing fuzzy objects from uncertain classification results. International Journal of Geog. Info. Sci. vol.15 (1), 27-42.

Cheung, C. and Black, J. 2005. Residential location-specific travel preferences in an intervening opportunities model: Transport assessment for urban release area. Journal of the eastern asia society for transportation studies. vol.6. 3773-3788.

Chi, E. H. 2000. A Taxonomy of Visualization Techniques Using the Data State Reference Model. Proceedings of InfoVis 2000 (Salt Lake City UT, October 2000)

Chinchor, N. 1997. MUC-7 Named Entity Task Definition. In Message Understanding Conference Proceedings: MUC 7.

Cho, J. 2001. Crawling the Web: Discovery and Maintenance of a Large-Scale Web Data. Ph.D. Thesis, Stanford University, November 2001.

Choo, C. W., Detlor, B., and Turnbull, D., 2000. Information Seeking on the web: An Integrated model of Browsing and Searching . First Monday. 5(2). Available online at http:/ /firstmonday.org/issues/issue5_2/choo/index.html.

Choo, C. W., Detlor, B., and Turnbull, D., 1998. A behavioural Model of Information seeking on the web - preliminary results of a study of how managers and IT specialists used the web. Available online at http://www.ischool.utexas.edu/~donturn/papers/asis98/ asis98.html

Chrisman, N., 1998. "Rethinking Levels of Measurement in Cartography." Cartography and GIS, 25 (4) 231-242.

Chu, W. W. and Zhang, G. 1997. Associative Query Answering via Query Feature Similarity. Proc. IIS. http://citeseer.ist.psu.edu/chu97associative.html

Chu, W. W and Chen, Q. 1992. Neighbourhood and Associative Query Answering. Intelligent Information Systems. 1(3-4): 355-382.

Ciolfi, L. and Bannon, L. 2005. Space, place and the design of technologically enhanced physical environments. In turner P. and Davenport E. (eds.), Space, Spatiality and Technology. London: Springer, volume 5 of The Computer Supported Cooperative Work Series, 217-233.

Cleary, J. G., Mcwha, D. J. A., Pearson, M., 1996. Timestamp representations for virtual sequences. viewed on 15.03.2006. Available online at http://www.cs.waikato.ac.nz/ timewarp/wengine/papers/wp96_27/ wp96_27.html

Clementini, E., Sharma, J. and Egenhofer, M. J. 1994. Modeling topological spatial relations: strategies for query processing. Computers and Graphics. 18(6): 815-822.

Clementini, E. and DiFelice, P. 1996. A model for representing topological relationships between complex geometric features in spatial databases. Information Sciences, 90, 1-4, 121-136.

Clementini, E. and DiFelice, P. 1996. An algebraic model for spatial objects with intermediate boundaries. Geographic Objects with Inderminate Boundaries (P.A. Burrough and A. U. Frank, editors), Taylor & Francis, London, 155-170.

Clementini, E., DiFelice, P., and Hernandez, D. 1997. Qualitative representation of positional information. Artificial Intelligence, 95, 317-356.

Clough, P. and Sanderson, M. 2004. A proposal for comparative evaluation of automatic annotation for geo-referenced documents. In the proceedings of the workshop on Geographic Information Retrieval (SIGIR'04).

Clough, P., Sanderson, M. and Joho, H. 2004. Extraction of semantic annotations from textual web pages. Spirit Technical Report D15-6201. Available online at http://www.geo-spirit.org/public_deliverables.html. 2004.

Clough, P. 2005. Extracting Metadata for Spatially-Aware Information Retrieval on the Internet. Proceedings of the 2005 Workshop on Geographic Information Retrieval. Bremen, Germany.

Clough, P., Purves, R. S. and Joho. H. 2006. Judging the spatial relevance of documents for GIR. In the proceedings of the 28th European conference on IR research (ECIR'06), London, U.K. April'06, Lecture Notes in Computer Science. 3936, 548-552.

Clough, P. D., Joho. H. and Purves, R. S. 2006. Judging the Spatial Relevance of Documents for GIR, In Proceedings of the 28th European Conference on IR Research (ECIR'06), London, UK, April 2006, Springer-Verlag LNCS Volume 3936, 548-552.

Coleman, A., Smith, T. R., Buchel, O. A. and Mayer, R. E. 2001. Learning spaces in digital libraries. In Research and Advanced Technology for Digital Libraries, 5th European Conference on Digital Libraries (ECDL), Darmstadt, Germany, edited by Panos Constantopoulos and Ingeborg Solvberg, Lecture Notes in Computer Science, Vol.2163. Berlin-Springer, 251-262

Coleman, A. S. and Oxnam, M. 2002. International Digital Libraries: Introduction to a special issue on interactivity in Digital Libraries. Journal of Digital Information, Vol.2. Issue 4. Article No. 132, 2002. 05-31.

Colton, C. E. 1997. The land of lincoln: Genesis of a Vernacular Region. Journal of Cultural Geography.

Colunga, E. and Gasser, M. 1998. Linguistic Relativity and Word Acquisition: A computational Approach. In Proceedings of the Twentieth Annual Conference of the Cognitive Science Society, 244-249.

Conti. G., Grizzard. J., Ahamad. M., Owen. H., 2005. "Visual Exploration of Malicious Network Objects Using Semantic Zoom, Interactive Encoding and Dynamic Queries," vizsec, p. 10, IEEE Workshops on Visualization for Computer Security (VizSec'05).

Corum, J. 2004. Maping votes. http://www.style.org/mappingvotes/

Corum, J. 2006. Parsing of the state of the union. http://www.style.org/stateoftheunion/

Cosijn, E. and Ingwersen, P. 2000. Dimensions of relevance. Information Processing and Management. 36. 533-550.

Couclelis, H., Golledge, R., Gale, N. and Tobler, W. 1987. Exploring the anchor-point hypothesis of spatial cognition. Journal of Environmental Psychology. Vol.7, No.2, 99-122.

Couclelis, H. 1996. Towards an operational typology of geographic entities with ill-defined boundaries. In geographic objects with indeterminate boundaries. eds. Peter A. Burrough and Andrew U. Frank, 45-55. London: Taylor & Francis.

Couclelis, H. 1996. Verbal directions for way-finding: space, cognition and language. In the construction of cognitive maps. Ed.J.Portugali, 133-53. Dordrecht:Kluwer Academic Publishers.

Couclelis, H. 1997. GIS without computers: building geographic information science from the ground up. Innovation in GIS-4: Selected papers from the Fourth National Conference on GIS Research UK (GISRUK). Editor Zarine Kemp, 219-26. London:Taylor & Francis.

Couclelis, H. and Gottsegen, J. 1997. What maps mean to people: denotation, connotation and geographic visualization in land-use debates. Spatial Information Theory: a theoretical basis for GIS International Conference COSIT'97. eds. Steven Hirtle and Andrew U. Frank, 151-62. Lecture Notes in Computer Science, no.1329. Springer-Verlag: Berlin.

Couclelis, H., 1998. World of information: The Geographic Metaphor in the visualization of complex information. Cartography and Geographic Information Systems, 25(4), 209-220.

Cousins, S. B., Paepcke, A., Winograd, T., Bier, E. A., and Pier, K., 1997. The digital library integrated task environment (DLITE). In Proc. of the 2nd ACM International Conference on Digital Libraries. p142-151. Philadelphia. PA, USA. July 1997.

Coutinho, M., Neches, R., Bugacov, A., Yao, K. T., Kumar, V., Ko, I. Y., Eleish, R., and Abhinkar, S. 1999. GeoWorlds: A Geographically Based Situation Understanding and Information Management System for Disaster Relief Operations. In Proceedings of the 5th International Conference on Information Systems, Analysis and Synthesis (ISAS 99), Orlando, Florida, July.

Cox, C. 1976. Anchor effects and the estimation of graduated circles and squares. The American

Cartographer. 3: 65-74.

Crampton, J. W., 2002. Interactivity types in geographic visualization. Cartography and Geographic Information Science. 29(2). 85-98.

Crestani, F., Lalmas, M., Van Rijsbergen, C. J., and Campbell, I. 1998. "Is this document relevant?... probably": a survey of probabilistic models in information retrieval. ACM Comput. Surv. 30, 4(Dec. 1998). 528-552- DOI=http://doi.acm.org/10.1145/ 299917.299920

Cribbin, T. and Chen, C. 2001. A study of navigation strategies in spatial-semantic visualization. Proceedings of the 9th international conference on human-computer interaction (HCI International 2001). August 5-10, New Orleans, Louisiana, USA, LEA, v.1. 948-952.

Crosswell, T. 2004. Place: A short introduction, Oxford: Blackwell.

Crow, V., Pennock, K., Pottier, M., Schur, A., Thomas, J. J., Wise, J., Lantrip, D., Fiegel, T., Struble, C. and York, J. 1994. Multidimensional Visualization and Browsing for Intelligence Analysis, in GVIZ '94 Proceedings of Graphics and Visualization Conference (Richland, WA).

Crowder, M. J. and Hand, D. J. 1990. Analysis of repeated measures. Chapman and Hall.

Croxton, F. E. and Stryker, R. E. 1927. Bar charts vs circle diagrams. Journal of American Statistical Association. 22: 473-482.

Crudge, S. E. and Frances, C. J. 2004. Using the information seeker to elicit construct models for search engine evaluation. Journal of the american society for information science and technology, 55(9): 794-806.

Cummins, R. 1996. Meaning and Mental Representation. MIT Press, Cambridge, Mass.

Cunningham, H., Maynard, D., Bontcheva, K. and Tablan, V. 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. In proceedings of the 40th Anniversary meeting of the association for computational lin-guistics (ACL'02), Philadelphia, July 2002.

Curry, M. R. 2002. Discursive Displacement and the Seminal Ambiguity of Space and Place. In Lievrouw L. and Livingstone S. (eds.). The Handbook of New Media: Social Shaping and Consequences of ICT. London: Sage Publications, 502-517.

Curry, M. R. 2005. Towards a geography of a world without maps: Lessons from Ptolemy and Postal Codes. Annals of the Association of American Geographers, 95(3), 680-691.

Dai, X. and Hardisty, F. 2002. Conditioned and Manipulable Matrix for Visual Exploration. Proceedings of the National Conference for Digital Government Research, Los Angeles, CA, May 20-22, 2002, 489-492.

Dale, M. R. T., Dixon, P., Fortin, M-J., Legendre, P., Myers, D. E. and Rosenberg, M. S. 2002. Conceptual and mathematical relationships among methods for spatial analysis. Ecography 25: 558- 577, 2002.

Dapos, A, A., and Kocka, M. 2005. Information assurance visualization for specific stages of situational awareness and intended uses: lessons learned. IEEE workshop on visualization for computer security (VizSEC'05). 107-112.

Davidsson, P. 1992. Concept Acquisition by Autonomous agents: Cognitive modelling versus the Engineering Approach, LUHFDA/HFKO-5010-SE

Davis, C. S. 2002. Statistical Methods for the analysis of repeated measurements. Springer-Verlag.

Davis, E. 1986. Representing and acquiring geographic knowledge. London-Pitman. Los-Altos-California: Morgan Kaufman.

Davis, F. D. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly, Vol.13, No.3., 319-340.

De Berg, M., van Kreveld, M., Overmars, M. and Schwarzkopf, O. 1997. Computational geometry, Springer-Verlag Berline Heidelberg.

Delboni, T. M., Borges, K. A. V., and Laender A. F., 2005. Geographic web search based on positioning expressions. In proceedings of the 2005 Workshop on Geographic Information Retrieval, Bremen, Germany, 61-64.

Densham, I. and Reid, J. 2003. A geo-coding service encompassing a geo-parsing tool and integrated digital gazetteer service. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 79-80, Alberta, Canada: ACL.

Dent, B. D. 1998. Cartography: Thematic Map Design. William C Brown Pub. 5th edition. July.

Deno, J. A. 1995. The relationship of previous spatial visualization ability. Engineering Design Graphics Journal. 59(3), 5-17.

Dervin, B. and Nilan, M. 1986. Information needs and uses. Annual review of information science and technology, 21, 3-33.

Dhyani, D., Keong, W. N and Bhowmick, S. S. 2002. A survey of web metrics. ACM Computing Survey. Vol: 34, No.4, December, 469-503. Available online at http:// www.dcs.warwick.ac.uk/ñikos/cs406/webmetrics.pdf.

Diaper, D. and Stanton, N. (Eds.) 2004. The Handbook of Task Analysis for Human-Computer Interaction, London: Lawrence Erlbaum Associates.

DiBiase. D. 1990. Visualization in the earth sciences. Earth and Mineral Sciences. Bulletin of College of Earth and Mineral Sciences. The Pennsylvania state university. 59(2). Available online at http://www.geovista.psu.edu/publications/others/dibiase90/swoopy.html

DiBiase. D., MacEachren. A., Krygier. J. B. & Reeves. C. 1992. Animation and the role of map design in scientific visualization. Cartography and Geographic Information Systems, 19. 201-214, 265-266.

DiBiase, D., Reeves, C., MacEachren, A., Krygler, J. B., von Wyss, M., Sloan, J. L. & Detweller, M. C. 1993. A map interface for exploring multivariate paleoclimate data. In Auto-Carto II Proceedings, 63-71, Bethesda, MD: The American Society for Photo-grammetry and Remote Sensing.

Ding, J., Gravano, L., Shivakumar, N. 2000. Computing geographical scopes of web resources. In Proceedings of the 26th VLDB Conference, Cairo, Egypt. 546-556.

Dobson, M. 1985. The future of perceptual cartograph. Cartographical. 22: 27-43.

Dorling, D. 1993. From computer cartography to spatial visualization: a new cartogram algorithm. In Proc. 11th Int. Symp. Computer-Assisted Cartography, Minneapolis, 208-217. Bethesda: American Society for Photogrammetry and Remote Sensing- American Congress on Surveying and Mapping.

Dorling, D. 1995. A new social atlas of Britain. John Wiley and Sons, U.K.

Dorling, D. 1996. Area Cartograms: their use and creation concepts and techniques. In Modern Geography Series, No:59, University of East Anglia: Environmental Publications.

Dowman, M., Tablan, V., Cunningham, H. and Popov. B., 2005. Web-assisted annotation, Semantic Indexing and Search of Television and Radio News. WWW 2005, May 10-14. Chiba, Japan.

Dransch, D. 2005. Activity and Context - A conceptual framwework for mobile geoservices. In Meng L., Zipf, A. and Reichenbacher. T. (eds.), Map-based mobile services- Theories and Methods and Implementations, Springer, Chapter 1, 1-8.

Dudley, G. 1991. Scale, aggregation and the modified area unit problem. The operational geographer, 9(3), 28-33.

Duncan, J. and Humphreys, G. 1989. Visual search and stimulus similarity. Psychological Review. 96: 433-458.

Duncan, J. and Humphreys, G. 1992. Beyond the search surface: Visual search and attentional engagement. Journal of experimental psychology. Human Perception and Performance. 18(2): 578-588.

Dutta. S., 1988. Approximate spatial reasoning. IEA/AIE, 1, 88.

Dykes. J., MacEachren. A., Kraak, M-J. 2005. Exploring Geovisualization. Elsevier Ltd. On behalf of the International Cartographic Association.

Dykes, J. 1998. Cartographic visualization: Exploratory Spatial Data Analysis with local indicators of Spatial Association using Tcl/Tk and cdv. The Statistician. Vol. 47, No: 3. 485-497.

Eades, P. 1984. A Heuristic for Graph Drawing. Congressus Numerantium 42. 149-60. 1984.

Eagleson, S., Escobar, F. and Williamson, I., 2003. Automating the administration boundary design process using hierarchical spatial reasoning theory and geographical information systems. In: Int. J.Geographical Information Science, 17 (2), 99-118.

Eastman, C. M. and Jansen, B. J. 2003. Coverage, Relevance and Ranking: The impact of Query Operators on Web Search Engine Results. ACM Transactions on Information Systems. Vol.21, No.4, October, 383-411.

Ebdon, D. 1977. Statistics in Geography. Basil Blackwell. Oxford.

Edelkamp, S., Jabbar, S. and Willhalm, T. 2003. Geometric travel planning. IEEE.

Egenhofer, M. J. and Frank. A. U. 1988. Designing object-oriented query language for GIS. Human Interface Aspects. In Proceedings: Third International Symposium on Spatial Data Handling, 79-95. Columbus, Ohio: International Geographical Union, Commission on Geographical Data Sensing and Processing, Dept of Geography, Ohio State University.

Egenhofer, M. J. 2002. Towards the semanic geospatial web. In Proceedings of the 10th ACM International Symposium on Advances in Geographic Information Systems. (McLean, Virginia, USA, November, 08-09, 2002) GIS'02. ACM Press, New York, NY 1-4. DOI=http://doi.acm.org/10.1145/585147.585148

Eick, G. S. and Steffen. J. L. 1992. Visualizing code profiling line oriented statistics. In proceedings of visualization '92. p210-217. Los Alamitos, California. Oct 1992. IEEE CS Press.

Eisenberg, T. A., and McGinty, R. L. 1977. On spatial visualization in college students. The Journal of Psychology, 95, 95-104.

Elias, B. 2003. Extracting landmarks with data mining methods. In Kuhn, W., Worbhoy, M.F. Timpf, S. eds: Spatial Information Theory. Vol: 285 Lecture Notes in Computer Science, Springer, Berlin, 398-412.

Elias, B. and Brenner, C. 2004. Automatic Generation and Application of Landmark in Navigation Data Sets. P.Fisher, (Eds.), Development in Spatial Data Handling, 469-480.Springer-Verlag, Berlin.

Ellegard, K. 1999. A time-geographic approach to the study of everyday life of individuals- a challenge of complexity, Geojournal 48: 167-175.

Ellis. D. 1989. A behavioural model for information retrieval system design. Journal of Information Science, 15: 237-247.

Ellis, D., Cox, D. and Hall, K. 1993. A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences. Journal of Documentation 49, no. 4: 356-369.

Ellis, D. 1996b. The dilemma of measurement in information retrieval research. Journal of Documentation, 45(3), 23-36.

Ellis, D. and Haugan, M. 1997. Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. Journal of Documentation 53, no. 4: 384-403.

Erickson, T. 2003. Designing visualizations of social activity: six claims. CHI'03. extended abstracts on human factors in computing systems. ACM Press. Ft.Lauderdale, Florida, USA. (3): 846-847.

Ettema, D. F. and Timmermans, 2006. Activity-based approaches to travel analysis. Elsevier.

Eyton, J. R. 1984. Complementary-colour, two-variable maps. Annals, Association of American Geographers. 71: 259-276.

Fabritius, H. 1998. Information seeking in the newsroom. Application of the cognitive framework for analysis of the work context. Information Research, 4(2). (viewed on 27 September, 2003, from http://informationr.net/ir/4-2/isic/fabritiu.html)

Fagerberg, P., Espinoza, F. and Persson, P. 2003. What is a place?: allowing suers to name and define places. In CHI'93: CHI'03 extended abstract on Human factors in computing systems, New York, NY, USA: ACM Press, 828-829.

Fairborn, D., Adriano, G., Adriano, N., Busier, G and Dykes, J. 2001 Representation and its relationship with cartographic visualization: a research agenda. Cartography and Geographic Information Science. Jan'08. Vol. 28, No. 1.

Falbo. D. L., Queen. L. P. and Blinn. C. R. 2002. Introduction to Data Analysis using Geographic Information Systems. Regents of the University of Minnesota.

Farrand, W. A., 1973. Information display in interactive design. PhD Dissertation. Dept of Engineering, University of California, Los Angeles, California.

Farrar, R. K. and Lerud, J. V. 1982. Online searching using geographic coordinates. In Second International Conference on Geological Information, 109-116.

Feiner, S. K. and Beshers, C. 1990. Worlds within worlds: metaphors for exploring n-dimensional virtual worlds. In Proceedings of the 3rd Annual ACM SIGGRAPH Symposium on User interface Software and Technology (Snowbird, Utah, United States, October 03 - 05, 1990). UIST '90. ACM Press, New York, NY, 76-83. DOI= http://doi.acm.org/ 10.1145/97924.97933

Femke Reitsma, Jochen Albrecht, 2005. Modelling with the Semantic Web in the Geosciences. IEEE Intelligent Systems. Vol. 20. No. 2, 86-88, Mar/Apr, Available online at http://www.sop.inria.fr/acacia/gtpe/epique-2005/Model-ing%20with%20the%20semantic%20web.pdf.

Feng, B., Li, Q., Ding, D., Wenyin, L., 2004. Exploiting Spatiality, Motility and Interactivity: A Relationship-Centric Flash Retrieval Model, icdcsw, 242-247, 24th International Conference on Distributed Computing Systems Workshops - W1: MNSA (ICDCSW'04), 2004.

Ferres, D., Ageno, A. and Rodrıguez, H. 2005. The GeoTALP-IR System at GeoCLEF-2005: Experiments Using a QA-based IR System, Linguistic Analysis, and a Geographical Thesaurus. In C. Peters, F. C. Gey, J. Gonzalo, G. J. F. Jones, M. Kluck, B. Magnini, H. Mller, and M. de Rijke., editors, CLEF, volume 4022 of Lecture Notes in Computer Science. Springer.

Filatova, E. and Hovy, E. 2001. Assigning time-stamps to event clauses. In proceedings of the 2001 ACL workshop on temporal and spatial information processing. Toulouse, France.

Fischer, G. 2000. User Modelling in Human-Computer Interaction. User Modelling and User-Adapted Interaction, 11, 65-86.

Fisher, P. and Dobson, J. 2003. Who knows where you are, and who should, in the era of mobile geography? Geographic: Journal of the Geographical Association, 88(4), 331-337.

Fisher, P. Cheng, T. and Wood, J. 2004. Where is Helvellyn? Multiscale morphometry and the mountains of the English Lake District. Transactions of the Institute of British Geographers, 29, 106-128.

Fisher, P. and Unwin, D. 2005. Re-presenting Geographical Information Systems. In Unwin P.F.D.J. (ed.), Re-presenting GIS, London:John Wiley & Sons, Chapter 1, 1-17.

Fishkin, K. and Stone, M. C., 1995. Enhanced dynamic queries via movable filters. Proceedings Human factors in Computing Systems CHI'95 Conference, Denver, CO, 415-420.

Fitzpatrick, G., Kaplan, S. M., Mansfield, T. 1996. Physical spaces, virtual places and social worlds: A study of work in the virtual.. In Q. Jones, and C. Halverson, (Eds.) Proceedings of CSCW'96: ACM Conference on Computer Supported Cooperative Work (334-343), Boston, MA.

Flannery, J. J. 1971. The graduated circle: a descriptive, analysis, and evaluation of a quantitative map symbols. Unpublished dissertation. University of Wisconsin, Madison.

Fluit, C., Sabou, M., van Harmelen, F. 2002. Ontology-based information visualization. Proceedings of Information Visualization '02.

Foley, J. and Ribarsky, B. 1994. Next-generation Data Visualization Tools, in Scientific Visualization, Advances and Challenges, Ed: L. Rosenblum, R.A. Earnshaw, J. Encarnacao, H. Hagen, A. Kaufman, S. Klimenko, G. Nielson, F. Post, D. Thalmann , Academic Press.

Fonesca, F. and Egenhofer, M. J. 1990. Knowledge sharing in Geographic Information Systems. In: P. Scheuermann, (Ed.). The Third IEEE International Knowledge and Data Engineering Exchange Workshop, Chicago IL, 85-90.

Fonseca, F. T. and Egenhofer, M. J. 1999. Ontology-driven geographic information systems. 7th ACM Symposium on Advances in Geographic Information Systems, Kansas City, MO. C.Bauzer Medeiros (ed.), Nov'99.

Fonseca, F., Egenhofer, M. *et al.* **2002.** Semantic Granularity in Ontology-Driven Geographic Information Systems. (AMAI) Annals of Mathematics and Artificial Intelligence - Special Issue on Spatial and Temporal Granularity : 36.1-2 (2002): 121-51.

Fonseca, F., Egenhofer, M., Agouris, P., and Camara, G. 2002. Using ontologies for integrated geographic information systems. Transactions in Geographic Information Systems, 6.

Ford, N., Wilson, T. D., Foster, A. Ellis, D., Spink, A. 2002. Information-seeking and Me diated Searching. Part-4. Cognitive Styles in Information Seeking. Journal of the American Society for Information Science and Technology, 53(9): 728-735.

Foster, A., 2004. A Nonlinear model of information-seeking behavior. Journal of the American Society for Information Science and Technology. 55 (30: 228-237).

Frank, A. U. 1998. Different types of "Times" in GIS. In Egenhofer, M. and Golledge, R. (eds.), Spatial and temporal reasoning in geographic information systems, Oxford: Oxford University Press, 40-62.

Frank, A. U. 2001. Tiers of ontology and consistency constraints in geographic information science, 15, No.7: 667-78.

Fredrikson A., North C., Plaisant C., Shneiderman B.,1999. Temporal, Geographical and categorical Aggregations viewed through coordinated displays: A case study of highway incident data. Proceedings of the workshop on new paradigms in information visualization and manipulation, Kansas city, Missouri, November 6, 1999.(in conjunction with ACM CIKM'99), ACM New York, 26-34.Human-computer interaction laboratory, University of Maryland. Available online at http://hcil.cs.umd.edu/trs/99-31/99-31.pdf

Freksa, C. 1999. Spatial aspects of task-specific way finding maps. Visual and Spatial Reasoning in

Design. John S. Gero and Barbara Tversky (eds.), 15-32.

Freundschuh, S. M. and Egenhoder, M. J. 1997. Human conceptions of spaces: Implications for geographic information systems. Transactions in GIS, 2(4): 361-375.

Friedman, A., Kerkman, D. and Brown, N. 2002. Spatial location judgements: A cross-national comparison of estimation bias in subjective North American geography. Psychonomic Bulletin and Review, 615-623.

Fruchterman, T., Reingold, E. 1991. Graph drawing by force-directed placement. Software: Practice and Experience, 21(11):1129–1164.

Fu, G., Abdelmoty, A. and Jones, C. B. 2003. Design of a geographical ontology. SPIRIT Technical deliverable -D5 3101. Available online at http://www.geo-spirit.org.

Fu, G., Jones, C. B. and Abdelmoty, A. I. 2005. Ontology based spatial query expansion in information retrieval. submitted for publication (ODBASE'2005).

Fu, G., Jones, C. B., Abdelmoty, A. I. 2005. Building a geographical ontology for intelligent spatial search on the web. In proceedings of IASTED International Conference on Databases and Applications (DBA'05).

Furnas, G. W., 1986. Generalized fisheye views. In Proceedings of CHI '86. ACM, New York, 16-23.

Furnas, G. W. and Buja. A. 1994. Prosection Views: Dimensional Inference through sections and projections (with discussion). Journal of Computational and Graphical Statistics. 3. 323-385.

Furnas, G. W., and Bederson. B. B., 1995. Space-Scale Diagrams: Understanding multiscale interfaces. Human Factors in computing systems. CHI'95. Conference Proceedings. ACM. 234-241.

Gabbard, J. L., Hix, D. and Swan II, E. J. 1999. User-centred design and evaluation of virtual environments. Nov/Dec'99. IEEE.

Gahegan, M. N. 1996. Visualisation strategies for exploratory spatial analysis. Paper presented at Third International Conference on GIS and Environmental Modelling, January 21 to January 25, 1996, Santa Fe, New Mexico. http://www.ncgia.ucsb.edu/conf/SANTA_FE_CDROM/sf_papers/gahegan_mark/gahegan.html

Gahegan, M. N. 1999. Four barriers to the development of effective exploratory visualization tools for the geosciences. International Journal of Geographic Information Science 13:289-310.

Gahegan, M. N. 2000. Tools to support the visual exploration of relationshps in geographic data. Paper presented at Annual Meeting of the Association of American Geographers, March 2000, Pittsburgh, Pennsylvania.

Gahegan, M., Wachowicz, M., Harrower and Rhyne, T-M. 2001. The Integration of Geographic Visualization with Knowledge Discovery in Databases and Geocomputation. Cartography and Geographic Information Science 28(1): 29-44.

Gale, W. A., Church, K. W., and Yarowsky, D. 1992. One sense Per Discourse. In: Proceedings of the 4th DARPA Speech and Natural Language Workshop, 233-237.

Galvao, A. and Sato, K. 2004. Human-Centered System Architecture: A Framework for Interpreting and Applying User Needs, in Proceedings of 16th International Conference on Design Theory and Methodology, ASME, Salt Lake City, UT, 2004.

Galvao, A. and Sato, K. 2005. Affordances in Product Architecture: Linking Technical Functions and Users' Tasks, in Proceedings of 17th International Conference on Design Theory and Methodology, ASME, Long Beach, CA, 2005.

Gardenfors, P. 1994, Three levels of inductive inferences. In Logic, Methodology and Philosophy of

Science, IX, D. Prawitz, B. Skyrms and D.Westerstahl. Eds. Elsevier Science, Amsterdam, 427-449.

Gardenfors, P. 2000. Conceptual Space: The geometry of thought. Bradford Books. MIT Press.

Garland, K. 1994. Mr. Beck's Underground Map. Capital Transport Publishing.

Gershon, N. and Page, W. 2001. What storytelling can do for information visualization. Commun. ACM 44, 8 (Aug. 2001), 31-37. DOI= http://doi.acm.org/10.1145/ 381641.381653

Gersmehl, P. J. 1990. Choosing tools: Nine Metaphors of four-dimensional cartography. Cartographic Perspectives 5: 3-17.

Getis, A. and Ord, K. 1992. The analysis of spatial information by use of a distance statistic. Geographical Analysis, 24, 189-206.

Gey, F., Larson, R., Sanderson, M., Joho, H., and Clough, P., 2005. GeoCLEF: the CLEF 2005. Cross-Language Geographic Information Retrieval Track. In: Working Notes for the Workshop on the Cross Language Evaluation Forum 2005. CD-ROM.

Gey, F. and Petras, V. 2005. Berkeley2 at GeoCLEF: Cross-Language Geographic Information Retrieval of German and English Documents. In: Working Notes of the 6thWorkshop of the Cross-Language Evaluation Forum, CLEF. Sep. 2005, Vienna, Austria. http://www.clef-campaign.org/ working_notes/

Gibson, J. J. 1977. The Theory of Affordances. In R.E.Shaw & J. Bransford (eds.), Perceiving, acting and knowing. Lawrence Erlbaum Associates, Hillsdale, NJ.

Gibson, J. J. 1979. The ecological approach to visual perception. Houghton Mifflin, Boston.

Gieryn, T. 2000. A space for place in Sociology. Annual Review of Sociology, 26, 463-496.

GML. 2004. Geography Markup Language (GML). Technical Report ISO/TC/211/WG4/ PT 19136, Open GIS Consortium, Inc.

Goel. A., Baker. C. A, Shaffer C. A., Grossman. B., Haftka R. T., Mason W. H., Watson L. T., 2001. VizCraft: A Problem Solving Environment for Configuration Design of a High Speed Civil Transport, IEEE Computing in Science & Engineering, Vol. 3(1): pages 56-66.

Golovchinsky, G. 1996. From Information Retrieval to Hypertext, and Back Again: The Role of Interaction in the Information Exploration Interface. Ph.D. Thesis, University of Toronto, November, 1996.

Golledge, R. G. and Stimson, R. J., 1997. Spatial behaviour: A geographic perspective. The Guilford Press.

Golledge, R. G. 1998. The Relationship between Geographic Information Systems and Disaggregate Behavioural Travel Modelling. Geographical Systems. 5: 9-17.

Golledge, R. G. 1999. Wayfinding behaviour. The Johns Hopkins University Press, 1999.

Good, L., Bederson, B. B., Stefik, M., and Baudisch, P., 2002. Automatic text reduction for changing size constraints. CHI'02. April. 20-25. ACM.

Goodchild, M. F. 1979. The aggregation problem in location-allocation. Geographical Analysis 11: 240-255.

Goodchild, M. F., Montello, D. R., Fohl, P. and Gottsegen, J. 1998. Fuzzy spatial queries in digital spatial data libraries. Fuzzy systems Proceedings, IEEE World Congress on Computational Intelligence. Vol. 1. 205-210. May'98.

Goodchild, M. F. and Zhou, J. 2003. Finding Geographic Information: Collection-level Metadata. GeoInformatica, Kluwer Academic Publishers, 7: 2, 95-112.

Goodell, H, Jr., Chiang, C. H., and Grinstein, G. G. 2006. Taxonomized Automated Monitoring of Interactive Visualization. Unpublished Technical Report No. 2006-006, University of Massachusetts Lowell, Dept. of Computer Science, Lowell, MA 01854. Available online at: http://teaching.cs.uml.edu/~heines/techrpts/details.jsp?Year=2006&SeqNo=006

Golbeck, J. and Mutton, P. 2006. Spring-Embedded Graphs for Semantic Visualization. Chapter 10 In Vladimir Geroimenko and Chaomei Chen, Visualizing the Semantic web. Second Edition. Springer London. 172-182. http://www.springerlink.com/content/nlr464326114r071/

Gordan, M. and Pathak, P. 1999. Finding information on the World Wide Web: the retrieval effectiveness of search engines. Information processing and management. 35, 141-180.

Gould, M. D. and McGranaghan, M. 1990. Metaphor in geographic information systems. In Proceedings of the 4th International Symposium on Spatial Data Handling. (K. Brassel & H. Kishimoto, eds., 433-442) Columbus: Ohio State University, 1990.

Graham, M. J. 2001. Visualizing multiple overlapping classification hierarchies. PhD Thesis. Napier University. December'01.

Graupmann, J. and Schenkel, R., 2006. GeoSphereSearch: Context-Aware Geographic Web Search. In Proceedings of the 2006 Workshop on Geographic Information Retrieval, Seattle, USA, 64-67.

Gravano, L. 1997. Querying multiple document collections across the Internet. Ph.D. Dissertation, Stanford University, August 1997.

Gravano, L., Hatzivassiloglou, V and Lichtenstein, R. 2003. Categorizing web queries according to geographical locality. In CIKM'03: Proceedings of the twelfth International Conference on Information and Knowledge management, 325-333, New York, NY, USA, 2003, ACM Press.

Green, D. 1998. First through the portal: the business potential of highly trafficked websites. Business Information Review. Vol.15, No. 3.

Green, D. 2000. The evolution of the web searching. Online information review. MCB University Press. Vol. 24, No. 2, 124-137.

Green. M. 1999. Toward a perceptual scene of multidimensional data visualization: Bertin and beyond. http://www.ergogero.com/dataviz/dviz0.html.

Green, R. T. and Courtis, M.C. 1966. Information theory and figure perception: the metaphor that failed. Acta psychol. 25, 12-36.

Green, S. J. 1997. Automatically generating hypertext by computing semantic similarity. Ph.D. Thesis. University of Toronto, 1997.

Gregory Conti, Julian Grizzard, Mustaque Ahamad, Henry Owen, 2005. Visual Exploration of Malicious Network Objects Using Semantic Zoom, Interactive Encoding and Dynamic Queries, vizsec, p. 10, IEEE Workshops on Visualization for Computer Security (VizSec'05)

Griffin, A. 2003. An empirical study of interactive maps in the context of a geographical simulation model. In: Cartographic renaissance. Proceedings 21st International Carto-graphic Conference, Durban, South Africa. International Cartographic Association, 358-364.

Griffith, D. A. 1996. Spatial Correlation and eigenfunctions of the geographic weights matrix accompanying geo-referenced data. The Canadian Geographer, 40(4), 351-367.

Griffiths, A. 1989. SAGIS: A proposal for a Sardinian geographical information system and an assessment of alternative implementation strategies. Journal of Information Science, 15, 261-267.

Grimm, L. G. and Yarnold, P. R. 1996. Reading and Understanding Multivariate Statistics. American Psychological association, Washington, DC.

Gruber, T. R. 1995. Towards principles for the design of ontologies used for knowledge sharing. International Journal of Human-computer Studies, 43, 907-928.

Guo, D., Peuquet, D. J. and Gahegan, M. 2003. ICEAGE: Interactive Clustering and Exploration of Large and High-Dimensional Geodata. GeoInformatica, 7: 3, 229-253, 2003.

Gustafson, P. 2001. Meanings of place: everyday experience and the theoretical conceptu-alizations. Journal of Environmental Psychology, 21, 5-16.

Guilford, J. B., Merrifield, P. R., Christensen, P. R., and Frick, J. W. 1960. An investigation of symbolic factors of cognition and convergent production. Report from the Psychological Laboratory, No. 23, University of Southern California.

Guilford, J. P., Merrifield, P. R., Christensen, P. R., and Frick, J. W. 1961. Some new symbolic factors of cognition and convergent production. Education and Psychological Measurement. 21, 515-543.

Haber, R. B. and McNabb, D. A. 1990. Visualization Idioms: A Conceptual Model for Scientific Visualization Systems, in Visualization in Scientific Computing, G. M. Nielson, B. Shriver and L. J. Rosenblum (eds), IEEE Computer Society Press.

Hackl René, Mandl Thomas, Womser-Hacker, Christa. 2005. Mono- and Cross-lingualRetrieval Experiments at the University of Hildesheim. In: Peters, Carol; Clough, Paul;Gonzalo, Julio; Kluck, Michael; Jones, Gareth; Magnini, Bernard (eds): Multilingual Information Access for Text, Speech and Images: Results of the Fifth CLEF Evaluation Campaign. Berlin et al.: Springer [LNCS 3491]. 165-169.

Hamblin, C. L. 1970. Fallacies. London. Methuen.

Hampe, M. and Elias, B. 2003. Integrating topographic information and landmarks for mobile applications. In Gartner G. (ed.). Location Based Services & Telecartography, vol. 66, 147-157.

Hand, D. J. and Taylor, C. C. 1987. Multivariate analysis of variance and repeated measures. Chapman and Hall, XIII, 262, S, 24cm I.

Hansen, P. 1997. An Exploratory study of IR interaction for user interface design. SIGIR'97 Conference, 27-31 July, Washington, USA. In: N.Belkin, A.D. Narasimhalu and Peter Willet (eds.). Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in information retrieval. 27-31 July, Philadelphia, Pennsylvania, USA. 135.

Hartley, R. V. L. 1928. Transmission of information. Bell System Technical Journal, 7: 535-563.

Harmon, G. 1984. The measurement of information. Information Processing and Management. 20(1-2): 193-198.

Harping, P. 2000. Getty Thesaurus of Geographic Names : User's Guide to TGN Data Releases V2.0. The J. Paul Getty Trust.

Harrower, M. A. and Brewer, C. A, 2003. ColorBrewer.org: An Online Tool for Selecting Color Schemes for Maps, The Cartographic Journal 40(1): 27-37. http://www.person-al.psu.edu/cab38/ColorBrewer/ColorBrewer.html

Harter, S. P. 1996. Variations in relevance assessments and the measurement of retrieval effectiveness. Journal of the American Society for Information Science, 47 (1), 37-49.

Haynes, J. R. 1970. Hierarchical Analysis of factors in cognition. American Educational Research Journal. 7(1). Jan. 55-68.

Hayward, S. 1984. Computers for animation. London: Focal Press.

Havre, S., Hetzler, B. and Nowell, L. 1999. ThemeRiver: In search of trends, patterns, and relationshops. http://citeseer.ist.psu.edu/havre99themerivertm.html

Havre, S. , Hetzler, E., Whitney, P., Nowell, L. 2002. ThemeRiver: Visualizing thematic changes in large document collections. In IEEE Transaction on Visualization and Computer Graphics, Jan/March'02. Vol.8 (1), 9-20. http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/2945/21152/00981848.pdf?tp=&isnumber=21152&arnumber=981848

Hearst, M. A and Plaunt, C. 1993. Subtopic structuring for full-length document access. In 16th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Pittsburg, June 1993, 59-68. New York, Association for Computing Machinery.

Hearst, M. A. and Chandu. K. 1997. Cat-a-Cone: An interactive interface or specifying searches and viewing retrieval results using a large category hierarchy. In proceedings of the 20th Annual international ACM SIGIR Conference, July 27-31, 1997. Philadelphia. PA. ACM Press.

Hearst, M. A. 1999. User Interfaces and Visualization. In: Baez-Yates, Ricardo A.; Ribeiro-Neto, Berthier (Eds.): Modern Information Retrieval. Reading, MA (Addison Wesley Longman) 1999. 257-324.

Hearst, M. A. 2001. User Interfaces and Visualization. In Modern Information retrieval, Ricardo Baeza- Yates, Bethier Ribeiro-Neto (eds.) Addison Wesley Longman Publications Ch. 10. 257-340.

Heath, D. E. 1993. Highly localized vernacular regionalization in the allien town bethlehem area. PANI. Professional Geographer. August'93. vol.45, no.5, 251-263.

Hedley, N. R. 2003. 3D geographic visualization and spatial mental models. Unpublished Ph.D Thesis. University of Washington, Department of Geography.

Heine, M. H. 1999. Reassessing and Extending the Precision and Recall Concepts. MIRA'99. Glasgow, Scotland 14-16th April'99.

Heinström, J. 2003. "Five personality dimensions and their influence on information behaviour" Information Research, 9(1). Paper 165. [Available at http://InformationR.net/ ir/9-1/paper165.html]

Heinzle, F and Sester, M. 2004. Derivation of implicit information from spatial data sets with data mining. In XXth congress of the international society for photogrammetry and remote sensing (ISPRS), Istanbul, Turkey.

Hendry, D. G and Harper, D. J. 1997. An informal information-seeking environment. Journal of the American Society for Information Science. 48(11): 1036-1048.

Henzinger, M. R., Heydon, A., Mitzenmacher, M., Najork, M., 1999. Measuring index quality using random walks on the web. Proceedings of the 8th International worldwide conference. Toronto Canada. Elsevier Science, May, 213-225.

Henzinger, M. R., Motwani, R., Silverstein, C., 2002. Challenges in web search engines. SIGIR Forum, Fall, Vol.36, No.2.

Herzog, D. 2003. Mapping the News: Case studies in GIS and Journalism. May'03. ESRI Press, Redlands, California.

Hetzler, B., Michelle, H. W., Havre, S. and Whitney, P. 1998. Visualizing the Full Spectrum of Document Relationships. In Proceedings of the Fifth International Society for Knowledge Organization (ISKO) Conference.

Hetzler, B., Whitney, P., Martucci, L., Thomas, J. 1998. Multi-faceted Insight Through Interoperable Visual Information Analysis Paradigms. In Proceedings of IEEE Symposium on Information Visualization, InfoVis '98, October 19-20, 1998, Research Triangle Park, North Carolina. 137-144.

Hetzler, B., Miller, N. 1998. Four Critical Elements for Designing Information Exploration Systems. Presented at Information Exploration workshop for ACM SIGCHI '98. Los Angeles, CA. April'98.

Hauff, C., Trieschnigg, D., and Rode, H. 2006. University of Twente at GeoCLEF 2006: Geofiltered Document Retrieval. Working Notes of the Cross-Lingual Evaluation Forum (CLEF) 2006.Alicante, Spain.

Heymann, S., Tham, K., Kilian, A., Wegner, G., Rieger, P., Merkel, D., Freytag, J. C. 2002. Viator- A tool family for graphical networking and data view creation. In proceedings of the 28th VLDB conference, Hongkong, China.

Heywood, 1998. Introduction to Geographical Information Systems. New York: Addison Wesley Longman.

Hibino, S. L. 1999. A task-oriented view of information visualization. In CHI '99 Extended Abstracts on Human Factors in Computing Systems (Pittsburgh, Pennsylvania, May 15 - 20, 1999). CHI '99. ACM Press, New York, NY, 178-179. DOI= http://doi.acm.org/ 10.1145/632716.632829

Hill, L. L., Frew, J., and Zheng, Q. 1999. Geographic Names. The implementation of a gazetteer in a georeferenced digital library. Digital Library. 5(1).

Hill, L. L. 2000. Core Elements of Digital Gazetteers: Placenames, Categories and Footprints. In proceedings of the 4th European Conference on Research and Advanced Technology for Digital Libraries (ECDL'00), London, UK: Springer-Verlag, Volume 1923 of Lecture Notes in Computer Science, 280-290.

Himmelstein, M. 2005. Local Search: The Internet Is the Yellow Pages, IEEE Computer Society Journal, 0018-9162/05, 26-34.

Hiramatsu, K. and Reitsma. F., 2004. Georeferencing the semantic web: ontology based markup of geographically referenced information. In EuroSDR/EuroGeographics.

Hobona, G., James, P. and Fairbairn, D. 2005. Multidimensional visualization of degree of relevance of geographic data. International journal of geographical information science. Vol.20., No.5. May 2006. 469-490.

Hochheiser, H., Shneiderman, B. 2001. Visual Specification of Queries for Finding Patterns in Time-Series Data Proceedings Discovery Science 2001, University of Maryland, Computer Science Dept. Technical Report #CS-TR-4326. UMIACS-TR-2001-25.

Hochheiser, H., Baehrecke, E. H., Mount, S. M., and Shneiderman, B. 2003. Dynamic Querying for Pattern Identification in Microarray and Genomic Data Proeedings IEEE Multimedia Conference and Expo, July, Baltimore, MD.

Hojholt, P. 1995. Generalization of built-up areas using Kohonen networks. In proceedings of Eurocarto XIII, 2-4 October, ISPRA, Italy.

Hollands, J. G., Carey, T. T., Mathews, M. L., and McCann, C. A. 1989. Presenting a graphical network : A comparison of performance using fisheye and scrolling views. In Designing and Using Human-Computer Interface and Knowledge Based Systems, G. Solvency and M.Smith, Eds. Elsevier, Amsterdam, 313-320.

Holmes, D. O. 1990. Computers and geographic information access. Meridian, 4, 37-49.

Holmquist, L. and Ahlberg, C., 1997. Flip Zooming: A Practical Focus+Context Approach to Visualizing Large Information Sets,In Holmquist, L.E. and Ahlberg, C. (1997) Flip Zooming: A Practical Focus+Context Approach to Visualizing Large Information Sets, in Design of Computing Systems: Social and Ergonomics Considerations (Eds. Smith, M., Salvendy, G.).

Holzmann, C. and Ferscha, A. 2007. Towards Collective Spatial Awareness Using Binary Relations. In Proceedings of the Third international Conference on Autonomic and Autonomous Systems (June 19 - 25, 2007). ICAS. IEEE Computer Society, Washington, DC, 36. DOI= http://dx.doi.org/10.1109/CONIELECOMP.2007.110

Hornbaek, K., Bederson, B. B., and Plaisant, C. 2003. Navigation patterns & usability of zoomable user interfaces: with and without an overview. interactions 10, 1 (Jan. 2003), 11-12. DOI= http://doi.

acm.org/10.1145/604575.604582

Hornbaek, K. and Frokjer, E. 1999. Do Thematic maps improve Information retrieval? Human-Computer Interaction - INTERACT'99. Angela Sasse and Chris Johnson (eds.) Published by IOS Press.

Howard. D. and MacEachren. A. 1996. Interface design for geographic visualization: Tools for representing reliability. Cartography and GIS. 23(2): 59-77.

Huang Xingang and Zhao Feng, 2000. Relation-based aggregation: finding objects in large spatial datasets. Intelligent Data Analysis 4, 129-147. IOS Press.

Huff, D. 1954. How to lie with statistics. New York, W.W. Norton

Hwang, J., Lay, S. and Lippman, A. 1994. Nonparametric multivariate density estimation: a comparative study.

Ingwersen. P., 1992. Information retrieval interaction, Taylor Graham Publishing, London, UK.

Ingwersen, P. 1994. Poly-representation of information needs and semantic entities: elements of a cognitive theory for information retrieval interaction. In Proceedings of the 17th Annual international ACM SIGIR Conference on Research and Development in information Retrieval (Dublin, Ireland, July 03 - 06, 1994). W. B. Croft and C. J. van Rijsbergen, Eds. Annual ACM Conference on Research and Development in Information Retrieval. Springer-Verlag New York, New York, NY, 101-110.

Ingwersen, P., 1996. Cognitive perspectives of information retrieval interaction. Journal of Documentation, 52(1), 3-50.

Isbell, C. L. Jr. and Viola, P. 1998. Restructuring Sparse High Dimensional Data for Effective Retrieval. Advances in Neural Information Processing Systems, May. ftp://publica-tions.ai.mit.edu/ai-publications/

Jacsó, P. 2005, Visualizing overlap and rank differences among web-wide search engines: some free tools and services. Online Information Review. 1468-4527, Vol.29. Issue-5, p.554-560, Emerald Group Publishing Limited. Article URL: http://www.emeraldinsight.com/10.1108/14684520510628927

James, T. Becker and Robin G. Morris, 1999. Working Memory(s), Brain and Cognition, Vol.41(1), Oct'99, 1-8.

Jamieson, S., 2004. Likert Scales: how to (ab)use them. Medical Education, Blackwell publishing ltd. 1212-1218, 38.

Jansen, B. J., Spink, A., Bateman, J. 1998. Real Life Information Retrieval: a Study of User Queries on the Web. In: ACM SIGIR Forum, 32 (1), 5-17.

Jansen, B. J., Spink, A., Bateman, J., *et al.*, **1998.** Searchers, the Subjects They Search, and Sufficiency: A Study of a Large Sample of EXCITE Searches. Aus: Maurer, Hermann; Olson, Richard O. (Eds.): Proceedings of WebNet 1998 - World Conference on the WWW, Internet & Intranet Conference: Orlando, FL, November 7- 12. CD-ROM.

Jansen, J. B. and Spink, A. 2006. How are we searching the World Wide Web?. A Comparison of nine search engine transaction logs. Information processing and management. 42, 248-263.

Järvelin, K. and Wilson, T. D. 2003. On conceptual models for information seeking and retrieval research, Information Research, 9(1) paper 163 [Available at http://InformationR.net/ir/9-1/paper163.html]

Jensen, E. C. 2006. Repeatable Evaluation of Information Retrieval Effectiveness in Dynamic Environments. PhD Thesis, Illinois Institute of Technology. May 2006.

Jiang, B. and Harrie, L. 2004. Selection of streets from a network using self-organizing maps. Transactions in GIS, 8(3): 335-350.

Johnson, F. C., Griffiths, J. R. and Hartley, R. J. 2003. Task dimensions of user evaluations of information retrieval systems. Information Research. 8(4), paper no.157. Available at http://informationr.net/ir/8-4/paper157.html

Johnson. B. and Shneiderman. B., 1991. Tree Maps.Tree-Maps - A Space filling approach to the visualization of hierarchical information structures. In Proc. IEEE visualization. 284-291. Sna Diego. California. IEEE Computer Society. Available Online http:// www.cs.umd.edu/hcil/treemaps

Joho, H. and Sanderson, M. 2003. Terrabyte-Scale Test Collections. In proceedings of the workshop on defining evaluation methodologies for Terabyte-Scale Collections. Sheffield.

Joho, H., and Sanderson, M. 2004. The SPIRIT collection: an overview of large web collection. SIGIR'Vol.38. No.2., Dec'04.

Joho, H., Clough, P., and Sanderson, M. 2004. A working searching system. SPIRIT Technical deliverable- D10 2101. Available online at www.geo-spirit.org. Jones, C. B., Purves, R., Ruas, A., Sanderson, M., Sester, M., vanKreveld, M., & Weibel, R. 2002. "SPIRIT Project Annex 1 - Description of Work", IST-2001-35047.

Jones, C. B., Alani, H. and Tudhope, D. 2002. Geographical terminology servers- closing the semantic divide, In M.F. Goodchild,Duckham, M and Worbhoys M.F (eds). Foundations of Geographic Information Science, Taylor and Francis, 201-218.

Jones, C. B., Purves, R., Ruas, A., Sanderson, M., Sester, M., van Kreveld, M., Weibel, R. 2002. Spatial Information Retrieval and Geographical Ontologies: An overview of the SPIRIT Project.In: Proceedings of SIGIR'02, the 25th Annula International ACM SIGIR Conference on Research and Development in Information Retrieval, ACM Press, 387-388. Aug 11-15, Tampere, Finland.

Jones, C. B., Purves, R., Sanderson, M. 2002. Spatial Information Retrieval and Geographical Ontologies: An overview of the SPIRIT Project. In the proceedings of 25th ACM Conference of the Special Interest Group in Information Retrieval, 389-390.

Jones, C. B., Abdelmoty, A. I., Fu, G. 2003. Maintaining ontologies for geographical information retrieval on the web. In proceedings of OTM confederated international conferences CoopIS, DOA and OOBase, 934-951. 934-951 Available online at www.geo-spirit.org

Jones, C. B., Purves, R. S. 2004. Workshop on geographic information retrieval, SIGIR 2004. SIGIR Forum 38(2): 53-56.

Jones, C. B., Abdelmoty, A. I. Finch, D., Fu, G., Vaid, S. 2004. The SPIRIT spatial search engine: Architecture, Ontologies and Spatial Indexing. In Geographic Information Science, Third International Conference Proceedings, GIScience, 2004. Adelphi, Md, USA, Oct 20-23. 125-139.

Jones, C. B. and Purves, R. S. 2006. Editorial: Geographic Information Retrieval. Computers, Environment and Urban Systems. Elsevier, 30(2006). 375-377.

Journel, A. G. and Huijbregts, C. J. 1978. Mining Geostatistics. (London: Academic Press).

Kadmon, N. and Shlomi, E. 1978. A polyfocal projection for statistical surfaces. Cartography. J. 15,1, 36-41.

Kalbach. J. 2000. Designing for Information Foragers: A Behavioural Model for Information seeking on the World Wide Web. Internet Technical Group. ITG Publication. Dec 2000: 3.3. Available Online at http://www.sandia.gov/itg/newletter/dec00/ article_information_foragers.html.

Kalashnikov, D. V. 2004. Querying Imprecise Data in Moving Object Environments. IEEE Trans. on Knowl. and Data Eng. 16, 9 (Sep. 2004), 1112-1127. DOI= http://dx.doi.org/10.1109/TKDE.2004.46

Kalashnikov, D. V., Ma, Y., Mehrotra, S., Hariharan, R., Venkatasubramanian, N., 2006. SAT: Spatial Awareness from Textual Input. EDBT http://sami.ics.uci.edu/papers/32.pdf

Kalashnikov, D. V., Ma, Y., Mehrotra, S., and Hariharan, R. 2006. Index for fast retrieval of uncertain spatial point data. In Proceedings of the 14th Annual ACM international Symposium on Advances in Geographic information Systems (Arlington, Virginia, USA, November 10 - 11, 2006). GIS '06. ACM, New York, NY, 195-202. DOI= http://doi.acm.org/10.1145/1183471.1183504

Kalashnikov, D. V., Ma, Y., Mehrotra, S., Hariharan, R., and Butts, C. 2006. Modeling and querying uncertain spatial information for situational awareness applications. In Proceedings of the 14th Annual ACM international Symposium on Advances in Geographic information Systems (Arlington, Virginia, USA, November 10 - 11, 2006). GIS '06. ACM, New York, NY, 131-138. DOI= http://doi.acm.org/10.1145/1183471.1183494

Kamada, T. 1988. On visualization of abstract objects and relations. Ph.D. Dissertation. Dept of Information Science, University of Tokyo, Dec.'88.

Kamada, T and Kawai, S. 1989. An algorithm for drawing general indirect graphs. Information Processing Letters. 31(1): 7-15.

Kamada, T. and Kawai, S. 1991. A general framework for visualizing abstract objects and relations. ACM Trans. Graph. 10, 1 (Jan. 1991), 1-39. DOI= http://doi.acm.org/10.1145/99902.99903

Kamada, T. and Kawai, S. 1991. An algorithm for drawing general indirect graphs. Information Processing Letters 31(1): 7-15, 1989.

Karnath, H-O., Ferber, S. and Himmelbach, M. 2001. Spatial Awareness is a function of the temporal not the posterior lobe. Nature. 411, 950-953.

Kawagishi, H., Koiso, K. and Tanaka, K. 2000. Dynamic zone retrieval and landmark computation for spatial data. In digital cities, technologies, experiences, and future perspectives (the book is based on an international symposium held in Kyoto, Japan in Sep 1999 T. Ishida and K. Isbister, Eds. Lecture Notes in Computer Science, Vol.1765. Springer-Verlag, London, 303-313).

Keim, D. A., North, S. C., Panse, C., and Schneidewind, J. 2003. Visualizing geographic information: VisualPoints vs CartoDraw. Information Visualization, 2, 58-67.

Keim. D. A., Panse C., Sips. M., 2005. Information visualization: Scope, Techniques and Oppurtunities for Geovisualization. Chapter 2. 23-52. In Dykes. J., MacEachren. A., Kraak, M-J. 2005. Exploring Geovisualization. Elsevier Ltd. On behalf of the International Cartographic Association.

Kekäläinen, J. and Järvelin, K. 2002. Evaluating information retrieval systems under the challenges of interaction and multidimensional dynamic relevance. In: Bruce, H., Fidel, R., Ingwersen, P. & Vakkari, P., eds. Emerging Frameworks and Methods, Seattle, 2002. Colerado: Libraries Unlimited, 253-270.

Kelle, U., Prein, G. and Bird, K. 1995. Computer-aided qualitative data analysis: Theory, Methods and Practice. SAGE Publications. London.

Keller, P. R. and Keller, M. M. 1992. Visual cues, practical data visualization. Piscataway. New Jersey : IEEE Press.

Kenneth L. Summers, Timothy E. Goldsmith, Steve Kubica, Thomas P. Caudell, 2003. "An Experimental Evaluation of Continuous Semantic Zooming in Program Visualization," infovis, p. 20, 2003 IEEE Symposium on Information Visualization.

Keppel, G. 1982. Design and Analysis: A researcher's handbook. London: Prentice Hall, 669.

Kim, J. 1976. Events as property exemplifications. In Action Theory (Dordrecht: Reidel). 159-177.

Kimler, M. 2004. Geo-coding: Recognition of geographical references in unstructured text, and their visualization. Diploma thesis. University of Applied Sciences- Hof, Dept of Computer Sciences & Technology.

Kirwan, B. and Ainsworth, L. K. (Eds.) 1992. A Guide to Task Analysis, Taylor and Francis, London.

Klippel, A. and Winter, S. 2005. Structural salience of landmarks for route directions. In: Proceedings of Spatial Information Theory: International Conference, COSIT 2005. Elli-cottville, NY, USA, Sep 14-18. Anthony G. Cohn and David M. Mark eds. vol.3693 of Lecture Notes in Computer Science, Springer Berlin/Heidelberg.

Klippel, A., Richter, K- F. 2004. Chrometic focus maps. In george gartner (ed.). Location based services & Telecartography. 39-44., Geowissenschaftliche Mittelungen. Technische Universitat, Wien, Wien.

Klippel, A., Richter, K- F., Barkowsky, T., and Freska, C., 2005. The cognitive reality of schematic maps. In L. Meng., A. Zipf & T. Reichenbacher (Eds.), Map-based mobile services - Theories, Methods and Implementations (57-74). Berlin: Springer.

Knight, B. 1993. Information loss in temporal knowledge representations. The computer Journal. Vol. 36. No. 2.

Koenemann, J. and Belkin, N. J. 1996. A case for interaction: a study of interactive information retrieval behaviour and effectiveness. Proceedings of the Human Factors in Computing Systems Conference, Zürich, 205-215.

Kofka. K. 1935. Principles of Gestalt Psychology. New York.

Kohonen. T., 1990. The Self-Organizing map. Proceedings of the IEEE. 78: 1464-1480.

Kohonen. T. 1995. Self-Organizing Maps.Springer, Berlin.

Koiso, K., Mori, T., Kawagishi, H., Tanaka, K. and Matsumoto, T. 2000. InfoLOD and Landmark: Spatial presentation of attribute information and computing representative objects for spatial data. International Journal of Cooperative Information Systems. Vol. 9, No.1-2, 53-76.

Koroscik, J. S. 1982. The effects of prior knowledge, presentation time and task demands on visual art processing. Studies in Art Education, 23(3), 13-22.

Kosala, R. and Blockeel, H. 2000. Web Mining Research. A Survey SIGKDD Explorations Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining. ACM, 2000.

Kosara, R., Sahling, G and Hauser, H. 2004. Linking scientific and information visualization with interactive 3D scatter plots. In proceedings of the 12th international conference in Central Europe on computer graphics, visualization and computer vision (WSCG), 133-140. 2004.

Koshman, S. 2004. A usability study comparing a prototype visualization-based system with a text-based system for information retrieval. Journal of Documentation, 60(5) 265-280.

Koshman, S. 2004. Web-based visualization interface testing: similarity judgments. Journal of Web Engineering, 3(3/4) 281-296.

Koshman, S. 2005. Testing user interaction with a prototype visualization-based information retrieval system. Journal of the American Society for Information Science and Technology, 56(8) 824-833.

Koshman, S. 2006. Visualization-based information retrieval on the web. Library & Information Science Research. Forthcoming.

Koshman, S., Spink, A., and Jansen, B. J. 2006. Web searching on the Vivisimo search engine. Journal of the American Society for Information Science and Technology. Volume 57, Issue 14, 1875 - 1887 http://www3.interscience.wiley.com/cgi-bin/abstract/113386305/ABSTRACT?CRETRY=1&SRETRY=0

Koua, E. L., MacEachren, A. and Kraak , M-J. 2006. Evaluating the usability of visualization methods in an exploratory geovisualization environment. International Journal of Geographic Information

Science. Vol. 20. No. 4, April 2006, 425-448.

Kraak, M-J. 1994. Interactive modelling environment for three-dimensional maps: Functionality and interface issues. Visualization in modern cartography. MacEachren, A. M. and F. D. R. Taylor. Oxford, Elsevier: 269-287.

Kraak, M-J. and MacEachren, A. 1999. Visualization for exploration of spatial data. International Journal of Geographical Information Science 13(4): 285-287.

Kraak, M-J. 2001. Trends in cartography. In: Kraak, M-J. and Brown. A. eds. Web Cartography. London: Taylor & Francis. 9-19.

Kraak, M-J. and Ormeling, F. J. 2003. Cartography : visualization of geospatial data. Harlow, Addison Wesley, Longman, 2003. 205 p. ISBN: 0-13-088890-7.

Kraak, M-J. 2003. Geovisualization illustrated. ISPRS Journal of Photogrammetry and Remote Sensing. 57. 390-399.

Kraft, R. 2005. Cost-Effective Creation of Specialized Search Engines. PhD Thesis, University of California Santa Cruz. March 2005.

Krueger, R. A. 1998. Developing question for focus groups. Focus group kit 3, Thousands Oaks, California: Sage Publications.

Kuhlthau, C. C. 1991. Inside the Search Process: Information seeking from the User's Perspective. Journal of the American Society of Information Science, 42(5), 361-371.

Kuhn, W. 1995. 7+/-2 Questions and Answers about Metaphors for GIS User Interfaces. In Cognitive aspects of Human-Computer Interaction for Geographic Information Systems (Nyerges, T. et al., eds.), Series D: Behavioural and Social Sciences, V I.83, Dordrecht, The Netherlands, Kluwer Academic Publishers, 113-122.

Kuhn, W. 2001. Ontologies in support of activities in geographical space. Int. J. geographical information science. Vol.15. no.7, 613-631.

Kulhthau, C. C., 2004. Seeking meaning. A process approach to library and information sciences. Second ed. Libraries unlimited. Greenwood Publishing.

Kushmerick, N. 1997. Wrapper Induction for Information Extraction. Ph.D. Dissertation, University of Washington. Technical Report UW-CSE-97-11-04.

Kwan, M. P and Lee, J. 2004. Geovisualization of human activity patterns using 3D GIS: A time geographic Approach. In Michael Goodchild and Donald Janelle, eds. Spatially Integrated Science, 48-66. New York: Oxford Press.

Kwok, K. L. and Deng, Q. 2003. GeoName: a system for back-transliterating pinyin place names. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003. Workshop on Analysis of Geographic References, 26-30, Alberta, Canada: ACL.

LaMacchia, B.A. 1996. Internet Fish. Ph.D. Thesis, MIT, Cambridge, MA. Also available as AI Technical Report 1579, MIT Artificial Intelligence Laboratory.

Lamping, J. and Rao, R. 1994. Laying out and visualizing large trees using a hyperbolic space. In ACM Symposium on User Interface Software and Technology, 13-14.

Lamping, J., Rao, R. and Pirolli, P. 1995. A focus + context technique based on hyperbolic geometry for viewing large hierarchies. In ACM SIGCHI Conference on Human Factors in Computing Systems, 401-408.

Lamping, J. and Rao, R. 1996. The Hyperbolic Browser: A Focus + Context technique for visualizing large hierarchies. Journal of visual languages and computing, 7, 33-55

Lamping, J., and Rao, R. 1997. Displaying Node-link structure with region of greater spacing and peripheral branches. US Patent 5619632, Xerox Corporation. Filed 14th sept. 1994, issued 8th April. 1997.

Landau, B. and Jackendoff, R. 1993. "what" and "where" in spatial language and spatial cognition. The behavioural and brain sciences, 16: 217-265.

Langran, G. E. and Poiker, T. K. 1986. Integration of name selection and name placement. In proc. 2nd International symposium on Spatial Data Handling, 50-64.

Larson, R. R. 1996. "Geographic Information Retrieval and Spatial Browsing". In Smith, L & Gluck, M. (eds.) In GIS and Libraries: Patrons, Maps and Spatial Information, edited by Linda Smith and Myke Gluck, Urbana-Champaign: University of Illinois, 1996. (81-124).

Larson, R. R. 2002. Placing cultural events in space and time. In foundations in geographic information science. London: Taylor and Francis.

Larson, R, R. and Frontiera, P., 2004. Spatial Ranking methods for Geographic Information Retrieval (GIR) in Digital Libraries. Research and Advanced Technology for Digital Libraries: 8th European Conference, ECDL 2004, Bath, UK, September 2004. Proceedings, 45-57. Edited by Rachel Heery and Liz Lyon. Published by Springer-Verlag LNCS 3232: 2004.

Larson, R. R. and Frontiera, P., 2004. Geographic Information Retrieval (GIR) Ranking Methods for Digital Libraries. In JCDL 2004: Proceedings of the Fourth ACM/IEEE Joint Conference on Digital Libraries, Tucson, Arizona, June 7- 11, 2004.

Larson, R. R. 2004. Geographic Information Retrieval (GIR): Searching where and what. In Proceedings of the Twenty-Seventh Annual International ACM SIGIR Conference on Research adn Development in Information Retrieval (SIGIR'04), 600. Edited by K. Jarvelin, J. Allan, P. Burza and M. Sanderson. Sheffield, UK, July 25-30.

Larson, R. R. 2005. Chesire II at GeoCLEF: Fusion and Query Expansion for GIR. In: Working Notes of the 6th Workshop of the Cross-Language Evaluation Forum, CLEF. Sep.2005, Vienna, Austria. http://www.clef-campaign.org/2005/working_notes/

Lau, T. and Horvitz, E. 1998. Patterns of Search: Analyzing and Modeling Web Query Refinement. In Proceedings of the Seventh International Conference on User Modeling. ACM Press, http://citeseer.ist.psu.edu/270469.html

Laurini, R., Milleret- Raffort, F. 1990. Principles of Geomatic Hypermaps. Proceedings of the 4th Conference on Spatial Data Handling, 642-651, Zurich, Switzerland, July.

Lawrence, S., Giles, C. L., 2000. Accessibility of information on the web. Intelligence Vol. 11, No. 1, 32-39.

Lee, C. S., Chen, Y- J., and Jian, Z-W., 2003. Ontology based fuzzy event extract agent for chinese e-News Summarization. (SCI) Expert Systems with Applications, vol. 25, no. 3. 431-447.

Lee, F., Bressan, S. and Ooi, B. 2001. Hybrid transformation for indexing and searching web documents in the cartographic paradigm. Information Systems. Vol. 26, No. 2, 75-92.

Lee, P., Klippel, A., and Tappe, T. 2003. The effect of motion in graphical user interfaces. In Butz, A. Kruger, A., Olivier, P., eds: Smart Graphics. Springer Berlin, 12-21.

Leidner, J. L., *et al.* **2003**. Grounding spatial named entities for information extraction and question answering. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 31-38, Alberta, Canada: ACL.

Leidner, J. L., Sinclair, G. and Webber, B. 2003. Grounding spatial named entities for information extraction and question answering. In A. Kornai and B. Sundheim (Eds.), Workshop on the Analysis of

Geographic References at the NAACL-HLT 2003 conference, Edmonton, Canada, 31–38.

Leidner, J. L. 2004. Toponym Resolution in Text: "which Sheffield is it". SIGIR'04, Sheffield, UK (Doctoral Consortium).

Leidner, J. L. 2006. An evaluation dataset for the toponym resolution task. Computers, Environment and Urban Systems, 30(4), 400-417.

Leitner, M. and Curtis, A., 2004. Cartographic guidelines for geographically masking the locations of confidential point data. Cartographic Perspectives, 49, 22-39.

Leitner, M. and Curtis, A. 2006. A first step towards a framework for presenting the location of confidential point data on maps- results of an empirical perceptual study. International Journal of Geographical Information Science, Vol. 20, No. 7, August 2006, 813-822.

Lempel, R. 1999. Finding Authoritative Sites on the WWW (and other hyperlinked media) by Analyzing the Web's Link-Structure. MSc Thesis, Technion, Israel Institute of Technology, July 1999.

Lempel, R. 2002. Link-Structure Analysis and Query Results Caching. PhD Thesis, Technion, Israel Institute of Technology, November 2002.

Leung, Y. K. 1989. Human-Computer interaction techniques for map-based diagrams In Designing and Using Human-Computer Interfaces and Knowledge Based Systems, G. Salvendy and M.Smith, Eds. Elsevier, Amsterdam, 361-368.

Leung , Y. K. and Apperley, M. D. 1993. Extending the perspective wall. In proceedings of OZCHI'93. 110-120.

Leung, Y. K. and Apperley, M. D. 1993. E3: Towards the metrication of graphical representation techniques. In Lecture Notes in Computer Sciences: Human-Computer Interaction, Bass, L.J., Gornostaev, J.G. and Unger, C., Eds. Springer-Verlag, Heidelberg. 125-140.

Leung. Y. K. and Apperley, M. D. 1994. A review and taxonomy of distortion-oriented presentation techniques. ACM Transaction on computer-human interaction. Available on-line at http://www.sims.berkeley.edu: 8000/courses/is247/f05/readings/ Leung_DistortionTechniques_TOCHI94.pdf

Leuski, A. and Allan, J. 2000. Lighthouse: Showing the way to relevant information, IEEE Symposium on Information visualization.

Leuski, A. 2001. Interactive information organization: Techniques and Evaluation. PhD Thesis. University of Massachusetts, Amherst. Department of Computer Science, May.

Levin. G., Nigam, K. and Feinberg, J. 2006. Dumpster - An Online visualization tool that depicts a slice through the romantic lives of American Teenagers. Source: http://art-port.whitney.org/commissions/thedumpster/

Leveling, J., Hartrumpf, S., and Veiel, D. 2005. Using Semantic Networks for Geographic Information Retrieval. In (Peters et al., 2006), 977–986.

Leveling, J. and Veiel, D. 2006. University of Hagen at GeoCLEF2006: Experiments with metonymy recognition in documents. CLEF. Alicante, Spain.

Levy, S. 2004. Making the Ultimate Map. Newsweek. June 7th Issue'04. http://www.msn-bc.msn.com/id/5092840/site/newsweek/

Lewis, C. and Rieman, J. 1994. Task-centered User Interface Design, A practical introduction. 1993-1994. Available Online at hcibib.org/tcuid/

Li, B. 1998. Exploring spatial and temporal patterns with self-organizing maps. In proceedings of GIS/LIS'98. Bethesda, MD. American Society for Photogrammetry and Remote Sensing, 89-96.

Li, H., Srihari, R. K., Niu, C, Li, W. 2002. Location Normalization for Information Extraction. In: Proceedings of COLING 2002, Taipei, Taiwan.

Li, H., *et al.* **2003**. InfoXtract location normalization: a hybrid approach to geographic references in information extraction. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 39-44, Alberta, Canada: ACL.

Li Q., North. C, 2003. Empirical Comparison of Dynamic Query Sliders and Brushing Histograms, infovis, 19, IEEE Symposium on Information Visualization.

Li, W. and Jonsson, M. 2006. Providing user self-contained location information using mobile phones. Tom Pfiefer et al. (Eds.): Advances in Pervasive Computing. Adjunct Proceedings of Pervasive, Dublin 7-10 May, Vol. 207.

Liben, L. S. 1981. Spatial representations and their behaviour: Multiple perspectives. Spatial Representation and Behaviour Across the Life Span. Academic Press, New York.

Lin. X., Soergel. D., Marchionini. G., 1991. A self-organizing semantic map for information retrieval. In Proc SIGIR'91. published as a special case of SIGIR Forum. ACM Press. 262-269. Oct.

Liu, Y. 2003. Geo-referenced Digital Libraries: Experienced Problems of Purpose and In-frastructure. Library Philosophy and Practice, Vol.6, No.1 (Fall'03).

Livnat, Y., Agutter, J., Moon, S. and Foresti, S. 2005. Visual correlation for situational awareness. In IEEE symposium of Information visualization, Minneapolis, MN, USA, 95-102., IEEE, October.

Llobera, M. 2003. Extending GIS-based visual analysis: the concept of visualscapes. International Journal of Geographical Information Science, 17(1), 25-48.

Llyod, R. 1988. Searching for map symbols: the cognitive process. The American Cartographer. 15(4): 363- 378.

Llyod, R. 1997. Visual search processes used in map reading. Cartographica. 34 (1):11-31.

Losee, R. M. 1998. A discipline independent definition of information. J.of the American Society for Information Science, 48(3), 1997, 254-269.

Lopez. N., Kreuseler. M., and Schumann. H., 2001. A scalable framework for information visualization - Special Issue, Computers & Geosciences.

Lovett, M. C. 2002. Problem solving. In H. Pashler & D. Medin (Eds.) Stevens' Handbook of experimental psychology: Memory and cognitive processes. New York: John Wiley & Sons.

Lynch, P. J. and Horton, S. 1997. Imprudent linking weaves a tangled web. Computer, 30, July, 115-117.

Ma, K-L., 2000. Visualizing Visualizations: User Interfaces for Managing and Exploring Scientific Visualization Data, IEEE Computer Graphics and Applications, vol. 20, no. 5, 16-19, Sept/Oct, 2000.

McCormick, B. H., DeFanti, T. A., Brown M. D. (ed), 1987. Visualization in Scientific Computing, Computer Graphics Vol. 21, No. 6, November.

McGee, M. G. 1979. Human spatial abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences. Psychological Bulletin, 86, 889-918.

MacEachren, A., in collaboration with Barbara P. Buttenfield, James B. Campbell, David DiBiase, and Mark Monmonier. 1992. "Visualization." 1992. In Ronald F. Abler, Melvin G. Marcus, and Judy M. Olson, editors, Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography, Rutgers University Press.

MacEachren, A. 1994. Visualization in modern cartography: Setting the agenda. Visualization in modern cartography. MacEachren, A. M. and F. D. R. Taylor. Oxford, Pergamon: 1-13.

MacEachren, A. 1994. Time as a cartographic variable. In: H. Hearnshaw & D. unwin. Eds., Visualization in GIS. London: Wiley & Sons, 115-130.

MacEachren, A. 1995. How maps work: representation, visualization, and design, New York (NY): Guilford Press.

MacEachren, A. and Kraak, M-J. 1997. Exploratory cartographic visualization: Advancing the agenda. Computers & Geosciences 23(4): 335-343.

MacEachren, A., Boscoe, F. P., Haug, D., Pickle, L. W. 1998. Geographic Visualization: Designing Manipulable Maps for Exploring Temporally Varying Georeferenced Statistics, infovis, p. 87, IEEE Symposium on Information Visualization.

MacEachren, A., Wachowicz, M., Edsall, R., Haug, D. and Masters, R. 1999. Constructing knowledge from multivariate spatiotemporal data: integrating geographical visualization with knowledge discovery in database methods. Int. J. Geographical Information Science 13(4): 311-334.

MacEachren, A. 2001. An evolving cognitive-semiotic approach to geographic visualization and knowledge construction. Information Design Journal, 10(1), 26-36.

MacEachren, A., Dai, X., Hardisty, F., Guo, D., and Lengerich, G. 2003. Exploring High-D Spaces with Multiform Matricies and Small Multiples. Proceedings of the International Symposium on Information Visualization, Seattle, WA, Oct. 19-21, 2003, 31-38.

Machlup, F. and Una Mansfield, editors. 1983. The study of information: Interdisciplinary messages. Wiley Interscience, New York.

MacKenzie, S. 1992. Fitts' Law as a Research and Design Tool in Human-Computer Interaction I. Human-Computer Interaction, Vol. 7, No. 1, 91-139 (doi: 10.1207/ s15327051hci0701_3)

Mackinlay, J. D., Robertson, G. G. and Card, S. K. 1991. The Perspective Wall: Detail and Context Smoothly Integrated. In Proceedings of the Conference on Human Factors in Computing Systems (CHI'95), Information Visualization, 173-179. ACM Press.

Maedche, A., Staab, S., Stojanovic, N., Studer, R. and Sure, Y. 2001. Semantic Portal - The SEAL Approach, In creating the semantic web. Fensel,D., Hendler, J., Lieberman, H., and Wahlster, W. (eds.) MIT Press, MA, Cambridge.

Magnetat-Thalmann, N & Thalmann.D. 1990. Computer animation: theory and practice. Berlin : Springer-Verlag.

Malczewski, J. 2006. GIS-based multicriteria decision analysis: a survey of literature. International Journal of Geographical Information Science. Vol.20, No.7, Aug'06, 703-726.

Mandl, T., Schneider, R., Strötgen, R. 2006. A Fast Forward Approach to Cross-lingual Question Answering for English and German. In: Gey, Fredric C.; Gonzalo,Julio; Jones, Gareth J.F; Kluck, Michael; Magnini, Bernardo; Müller, Henning; Peters, Carol; de Rijke, Maarten (eds.). Accessing Multilingual Information Repositories: 6th Workshop of the Cross-Language Evaluation Forum, CLEF. Sep. 2005, Vienna, Austria. Revised Selected Papers. LNCS 4022; Springer, 332-336.

Mann. T. M., 2002. Visualization of www- search results. PhD Thesis. University of Konstanz, Germany.

Manov, D., et al. 2003. Experiments with geographic knowledge for information extraction. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 200. Workshop on Analysis of Geographic References, 1-9, Alberta, Canada: ACL.

Marchionini, G. 1992. Interfaces for end-user information seeking. In: Journal of the American Society for Information Science (JASIS), 43 (1992) 2, p. 156-163.

Marchionini, G. 1995. Information Seeking in Electronic Environments. Cambridge University Press.

Cambridge, England.

Mardis, S. and Burger, J. 2005. Design for an Integrated Gazetteer Database. Technical Description and user guide for a gazetteer to support natural language processing applications. MITRE Technical Report.

Mark, D. M., and Frank, A. U., 1996. Experiential and Formal Models of Geographic Space. Environment and Planning, B, v. 23, 3-24.

Mark, D. M., Smith, B. and Tversky, B. 1999. Ontology and Geographic Objects: An empirical study of cognitive categorization. In COSIT'99: Proceedings of the International Conference on Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science, London, UK: Springer-Verlag, 283-298.

Markowetz, A., Brinkhoff, T. and Seeger, B. 2004. Exploiting the Internet as a Geospatial Database. In Post-Workshop Book of International Workshop on Next Generation Geospatial Information, Cambridge, MA 2003.

Martin, D. W. 1977. Doing Psychology experiments. 4th Edition. Brooks/Cole Publishing company.

Martin. A. and Ward. M. O., 1995. High dimensional brushing for interactive exploration of multivariate data. In Proceedings of the Conference on Visualization '1995, pages 271-278. IEEE Computer Society Press.

Martins, B., Silva, M. J. and Chaves, M.S. 2005. Challenges and resources for evaluating geographic IR. In Proceedings of the 2005 Workshop on Geographic Information Retrieval, Bremen, Germany, 65-69.

Martins, B., Silva, M. J., Frietas, S. and Afonso, A. P. 2006. Handling locations in search engine queries. GIR'06. Held in conjunction with ACM SIGIR, Seattle, Aug 10th, 2006.

McCormick, B. H., Defanti.T. A.and Brown. M. D. 1987. Visualization in scientific computing, ACM Computer Graphics (Special Issue), 21 (6).

McCurley, S. K. 2001. Geospatial Mapping and Navigation of the web. In Proceedings of the 10th International WWW Conference, Hongkong, 1-5 May, ACM Press, 221-229.

McEathron, S. R., McGlamery, P., Shin, D, G., Smith, B., Su, Y. 2002. Naming the landscape: Building the connecticut digital gazetteer. 67th IFLA General Conference, Boston USA, August 16-25. Section on Geography and Map Libraries.

McIntosh, J and Yuan, M. 2005. Assessing similarity of geographic processes and events. Transaction in GIS. 9(2): 223-245.

McKnight, C. 2000. The personal construction of information space. Journal of the American Society for Information Science, 51(8), 730-733.

Meng, L. 2005. Egocentric design of map-based mobile services. The cartographic journal. 42, 1: 5-13, June.

Mersey, J. 1990. Colour and thematic map design: the role of colour scheme and map complexity in choropleth map communication. Monograph. 41. Cartographica. 27:1-157.

Michel de Certeau, 1984. "Space is a practiced place" -Michel de Certeau, The Practice of Everyday Life (Berkeley: University of California Press, 1984). 117.

Michon, P. E. and Denis, M. 2001. When and why are visual landmarks used in giving direction?. D.R. Montello (Ed.), Spatial Information Theory: Foundations of Geographic Information Science, Lecture notes in Computer Science, 2205, Springer-verlag, 292-305.

Mikheev, A., Moens, M. and Grover, C. 1999. Named Entity Recognition Without Gazetteers. In Proceedings of the Annual Meeting of the European Association for Computational Linguistics EACL'99, Bergen, Norway, 1-8.

Miller, A. G. 1956. The Magical Number Seven, plus or Minus Two: Some limits on our capacity for processing information. The Psychological review. 63, 81-97.

Miller, A. G. 1967. The magical number seven, plus or minus two: some limits on our capacity for processing information. In The Psychology of Communication, 14-44, reprinted Psychol. Rev. 63:81-97.

Miller, C. L. 1996. A historical review of applied and theoretical spatial visualization publications in engineering graphics. The Engineering Design Graphics Journal, 60(3), 12-33.

Miller, N., Hetzler, B., Nakamura, G and Whitney, P., 1997. The Need For Metrics In Visual Information Analysis. Workshop on New Paradigms in Information Visualization and Manipulation in conjunction with the Sixth ACM International Conference on Information and Knowledge Management (CIKM '97), November 13-14, Las Vegas Nevada, ACM Press.

Miller E. N., Wong, P. K., Brewster, M. and Foote, H. 1998. Topic Islands- A wavelet- Based Text Visualization system. IEEE Visualization '98.

Miller, H.J. and Han, J. 2001. Geographic Data Mining and Knowledge Discovery. Taylor and Francis, Oct 11.

Miller, H. J. and Wentz, E. A. 2003. Representation and spatial analysis in geographic information systems. Annals of the Association of American Geographers, 93(3), 574-594.

Mitra, G.A., Buckley, C., Singhal, A. and Cardie, C. 1997. An analysis of statistical and syntactic phrases. Proceedings of the 5th International Conference on Computer-Assisted Information Searching on Internet (RIAO'97), 200-214, Montreal, Canada.

Mitta, D. A. 1990. A fisheye presentation strategy: Aircraft maintenance data. In proceedings of interact '90. 875-885.

Mizzaro, S. 1997. Relevance: The whole history. Journal of the American Society for Information Science. 48(9), 810-832.

Mizzaro, S. 1998. How many relevance's in information retrieval. Interacting with computers. 10(3), 303-320.

Mladenic, D. 1998. Machine learning on non-homogeneous, distributed text data. PhD Dissertation. University of Ljubljana, October 1998.

Monmonier, M. 1988. Geographical Representations in Statistical Graphics: A Conceptual Framework. In 1988 Proceedings of the Section on Statistical Graphics, 1-10, Alexandria, VA. American Statistical Association.

Monmonier, M. 1989. Geographic Brushing: Enhancing Exploratory Analysis of the Scatter plot Matrix. Geographical Analysis, 21(1): 81-84.

Monmonier, M. 1996. How to lie with Maps. With a foreword by H. J. de Blijj. 2nd Edition, 222.

Montello, D. R., Lovelace, K. L., Golledge, R. G., and Self, C. M. 1999. Sex-related differences and similarities in geographic and environmental spatial abilities. Annals of the Association of American Geographers, 89, 515-534.

Montello, D. R.,2002. Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches, Cartography and Geographic Information Science, Vol.29, No.3, 2002, 283-304.

Montello, D. R. 2003. Regions in geography: Process and content. In M. Duckham, M. F. Goodchild, & M. F. Worbhoys (Eds.), Foundations of Geographic Information Science (173-189). London: Taylor &

Francis.

Moon, I., Miura, J. and Shirai, Y. 2001. Automatic extraction of visual landmarks for a mobile robot under uncertainty of vision and motion. IEEE International Conference on Robotics and Automation, 1188-1193.

Mordkoff, J. T., Yantis, S. and Egeth, H. E. 1990. Detecting conjunctions of colour and form in parallel. Perception and Psychophysics. 48(2): 157-168.

Morse, E. and Lewis, M. 2000. Evaluating visualizations: using a taxonomic guide. International Journal of Human-Computer Studies. 53. 637-662. Available online at http:// doi:10.1006/ijhc.200.0412.

Morville, P. 2002. In defence of search. Digital web magazine. July 16. Available online at http://www.digital-web.com/articles/in_defense_of_search/

Morville, P. 2005. Ambient Findability: What we find changes who we become. O'Reilly Media, Inc. 1st Edition (September 26, 2005) ISBN:0596007655.

Mountain, D. and Raper, J. 2001. Modelling human spatio-temporal behaviour: A challenge for location-based services. Geocomputation.

Murray, B. H. 2000. Sizing the internet. A Cyveillance White Paper. Cyveillance Inc. July. Source : http://www.cyveillance.com/web/downloads/sizing_the_internet.pdf

Munzner, T. and Burchard, P. 1995. Visualizing the structure of the world wide web in 3D hyperbolic space. In Proceedings of the VRML'95 Symposium (San Diego, CA, December 13-16, 1995), p.33-38. ACM SIGGRAPH, 1995.

Munzner, T. 1997. H3: laying out large directed graphs in 3D hyperbolic space. Proceedings of the 1997 IEEE Symposium on Information Visualization, October 20-21, Phoenix, AZ.

Munzner, T. 2000. Interactive Visualization of Large Graphs and Networks. PhD thesis, Stanford University.

Mutton, P. and Rodgers. P. 1998. Spring embedder preprocessing for www visualization. In Proc. Int'l Symposium on Web Graphics and Visualization IV02-WGV, pages 744-749. http://citeseer.ist.psu.edu/mutton02spring.html

Naaman, M., Song, Y. J., Paepcke, A., Gracia-Molina, H. 2004. Assigning textual names to sets of geographic coordinates. GIR'04. SIGIR'04.

Naaman, M., Song, Y. J., Paepcke, A & Garcia-Molina, H., 2006. Assigning textual names to sets of geographic coordinates. Computers, Environment, and Urban Systems. 418-435.

Nakaya, T. 2000. An information statistical approach to the modifiable areal unit problem in incidence rate maps. Environment & Planning A, Vol. 32 Issue 1.

Neilsen, J. 1993. Usability Engineering. Academic Press, Boston.

Neilsen, J. 1995. Multimedia and Hypertext: The internet and beyond. Morgan Kaufmann Publications. San Francisco. Boston AP Professional. ETH (ZH), BIB-764899/ 001479508.

Nesbitt, K. V. and Friedrich, C. 2002. Applying gestalt principles to animated visualization of network data. Proceedings of the sixth international conference on information visualization. (IV'02). IEEE.

Neumann, A. 2003. Delivering Interactive Topographic Web-maps using Open-Source Database Technology, Proceedings of the second SVG.Open/Carto.net Conference, Vancouver, Canada.

Neumann, A. 2005. Thematic Navigation in Space and Time Interdependencies of Spatial, Temporal and Thematic Navigation for Cartographic Visualization - By Example of a European Map of Cultural Heritage; For the SVG Open 2005 conference Available online at http://www.svgopen.org/2005/papers/abstract_neumann_thematic_navigation_in_space_and_time/

Nivedita, R., Kadaba, P., Irani, P., Toulouse, P. 2005. Visualizing time dependent semantics. An application to quantum algorithms. Ninth International Conference on Information visualization [IV'05], iv. 182-187.

Nogueras-Iso; Javier; Zarazaga-Soria, F; Muro-Medrano; Pedro, R. 2005. Geographic Information Metadata for Spatial Data Infrastructures: Resources, Interoperability and In-formation Retrieval. XXII, Springer-Verlag.

Norman. D. A. 1999. Affordance, Conventions and Design, Interactions, 28-43. http:// www.jnd.org/dn.pubs.html

Norman, D. A. 2006. People: the way I see it: Interaction design is still an art form. ergonomics is real engineering, interactions Volume 13, Number 1 (2006), 45-60.

North, C. and Shneiderman, B. 2000. Snap-together visualization: can users construct and operate coordinated visualizations? International Journal of Human-Computer Studies, 53(5), 715-739.

North, C. and Schneiderman, B. 2000. Snap-Together Visualization: A user interface for coordinating visualizations via Relational Schemata. Proc. AVI '00. Palermo. Italy.

North, C., Conklin, N. and Saini, V. 2002. Visualization schemas for flexible information visualization. Proceedings IEEE Symposium on Information Visualization, Boston, 28-29 Oct., 15- 22.

Notess, G. R. 2005. Search Engine Features Chart. Search Engine Showdown. Available online at http://www.searchengineshowdown.com/features/

Nothegger, C. 2003. Automatic selection of landmarks. Master's thesis. Technical University of Vienna.

Nothegger, C., Winter, S. and Raubal, S. 2004. Selection of Salient Features for Route Directions, Spatial Cognition and Computation, Vol.4, No.2, 113-136.

Nova, N., Girardin, F. and Dillenbourg, P. 2005. 'location is not enough! ': an empirical study of location-awareness in mobile collaboration. In Proceeding of the third IEEE International Workshop on Wireless and Mobile Technologies in Educations (Tokushima, Japan, November 28-30 2005), IEEE Press, 21–28.http://craftsrv1.epfl.ch/publications/2005/paper_wmte2005.pdf

Nova, N. 2007. The influence of location awareness on computer-supported collaboration. n. 3769, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland, March '07. http://biblion.epfl.ch/EPFL/theses/2007/3769/EPFL_TH3769.pdf

Nowell, L. T. 1996. Visualizing Search Results: Some Alternatives to Query-Document Similarity. In Frei, H.P. et al., (Eds.) Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '96) (65-75). New York: ACM.

O' Sullivan, D. and Unwin, D. J. 2003. Geographic Information Analysis, Wiley.

O'Dea, D., Geoghegan, S. and Ekins, C. 2005. Dealing with Geospatial Information in the Semantic Web. Australasian Ontology Workshop (AOW'05): Sydney, Australia, In: Conferences in Research and Practice in Information Technology (CRPIT), Vol. 58. T. Meyer, M. Orgun, eds.

O' Day, V. L. and Jeffries. R. 1993. Orienteering in an information landscape: how information seekers get from here to there. In Proc. Of the INTERCHI '93. Amsterdam. Netherlands. April'93. IOS Press.

Okabe, A., Boots, B. and Sugihara, K. 1992. Spatial Tessellations: Concepts and Applications of Voronoi Diagrams (London: Wiley)

Old. L. J., 2002. Information Cartography: Using GIS for Visualizing Non-spatial Data. Proceeding, ESRI International Users' Conference. San Diego CA, July 2002.

Olson, D. R. and Bialystok, E. 1983. Spatial Cognition: The structure and development of mental representations of spatial relations. London: Lawrence Erlbaum Associated.

Onsrud, H. J., 1999. Liability in the use of geographic information systems and geographic datasets. In D. Maquire, M. Goodchild, D. Rhind and P. Longley (eds.), Geographic Information Systems: Principles, Techniques, Management and Applications (Wiley, New York, 1999), 643-652.

OntoWeb Consortium, 2002. OntoWeb: Ontology-based information exchange for knowledge management and electronic commerce. Technical Report. IST-2000-29243.

Oppenheim, C., Morris, A., McKnight, C. and Lowley, S. 2000. The evaluation of www search engines. Journal of documentation. Vol.56. no.2. March, 190-211.

Openshaw, S. and Taylor, P. 1979. A million or so correlation coefficients: Three experiments on the modifiable area unit problem. In Statistical Applications in the Spatial Sciences, ed. N. Wrigley, (London: Pion), 127-144.

Openshaw, S. and Taylor, P. 1981. The modifiable area unit problem. In Quantitative Geography: A British View, N. Wrigley, ed. (London: Routledge and Regan Paul).

Openshaw, S. 1997. Artificial Intelligence in geography. Chichester, Wiley.

Paasi, A. 2002. Place and region: regional worlds and words. Progress in Human Geography, 26(6), 802-811.

Panse, C. 2004. Visualizing geo-related data using cartograms. Ph.D Thesis. Universität Konstanz, Germany.

Papadias, D. 1994. Relation-based representations for spatial knowledge, PhD Thesis. Knowledge and Database Systems Laboratory, Dept of Electrical and Computer Engineering, National Technical University of Athens. KDBSLAB-TR-94-05.

Papadopoulos, A. N. and Manolopoulos, Y. 2005. Nearest Neighbor Search. Springer, Series in computer Science.

Pask, G., & Scott, B. C. E. 1972. Learning strategies and individual competence. International Journal of Man Machine Studies, 4, 242-253.

Pask, G. and Scott, B. C. E. 1973. CASTE: A system for exhibiting learning strategies and regulating uncertainty. International Journal of Man Machine Studies, 5, 17-52.

Pask, G. 1976a. Conversational techniques in the study and practice of education. British Journal of Educational Psychology, 46, 12-25.

Pask, G. 1976b. Styles and strategies of learning. British Journal of Educational Psychology, 46, 128-148.

Pask, G. 1976c. Conversation theory: Applications in education and epistemology. Amsterdam: Elsevier.

Pask, G. l979. Final report of S.S.R.C. Research programme HR 2708. Richmond (Surrey): System Research Ltd.

Pask, G. 1988. Learning strategies, teaching strategies, and conceptual or learning style. In R.R. Schmeck (Ed.), Learning strategies and learning styles (83-99). New York: Plenum Press.

Perlich, C. and Provost, F. 2003. Aggregation-based feature invention and relational concept classes. In proceedings of the ninth international conference on knowledge discovery and data mining (KDD-2003).

Pearl, J. 1991. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Francisco: Morgan Kaufmann Publishers.

Perlin, K. and Fox, D. 1993. Pad: An Alternative Approach to the Computer Interface. Proc. SIGGRAPH '93, 57-64.

Persson, P. and Fagerberg, P. 2002. GeoNotes: a real-use study of a public location aware community system. Technical report, (T2002:27), Stockholm, SICS.

Peters, C. and Braschler, M., 2001. Cross-language system evaluation: The CLEF campaigns. Journal of the American Society for Information Science and Technology. 52(12), 1067-1072.

Peuquet, D. J. 1994. It's about time: A conceptual framework for the representation of temporal dynamics in geographic information systems. Annals of the association of American geographers, 84(3). 441-461.

Pfitzner, D., Hobbs, V., and Powers, D. 2003. A unified taxonomic framework for information visualization. In Proceedings of the Australian Symposium on information Visualisation - Volume 24 (Adelaide, Australia). T. Pattison and B. Thomas, Eds. Conferences in Research and Practice in Information Technology Series. Australian Computer Society, Darlinghurst, Australia, 57-66.

Pietriga, E. 2005. A Toolkit for addressing HCI issues in Visual Language Environments. IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCI), pages 145-152, Sep'05, Dallas, Tx, USA.

Pillat, R. M., Valiati, E. R. and Freitas, C. M. 2005. Experimental study on evaluation of multidimensional information visualization techniques. In Proceedings of the 2005 Latin American Conference on Human-Computer interaction (Cuernavaca, Mexico, October 23 - 26, 2005). CLIHC '05, vol. 124. ACM Press, New York, NY, 20-30. DOI= http:// doi.acm.org/10.1145/1111360.1111363.

Pinkerton, B. 2000. WebCrawler: Finding what people want. PhD thesis, University of Washington, November 2000.

Pirolli, P. and Card, S. 1995. Information foraging in information access environments human factors in computing systems: Proceedings of CHI '95. Available online at http:// www.acm.org/turing/sigs/sigchi/chi95/Electronic/documnts/papers/ppp_bdy.htm

Pirolli, P., Card, S. K. and van der Wege, M. M. 2000. The effect of information scent on searching information: Visualization of large tree structures, Advanced Visual Interface. May.

Pitkow, J. 1997. Characterizing World Wide Web Ecologies. Ph.D. Dissertation, Georgia Institute of Technology, June 1997.

Poole, C., Miller, S. A., Church, E. B. 2006. Development: Ages & Stages - Spatial Awareness. Early Childhood Today (3), 20(6), p.25-30. April'06. http://www.eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=EJ738512&ERICExtSearch_SearchType_0=no&accno=EJ738512

Porter, M. F. 1980. An algorithm for suffix stripping. Program, 14(3), 130-137.

Pouliquen, B., Steinberger, C. Ignat, and De Groeve. 2004. Geographical Information Recognition and Visualization in Texts Written in Various Languages. In Proceedings of the 19th Annual ACM Symposium on Applied Computing. Nicosia, Cyprus, 1051-1058

Pouliquen, B., Steinberger, R. and Ignat, C., 2004. Automatic Linking of Similar Texts Across Languages. In N. Nicolov, K. Bontcheva, G. Angelova, and R. Mitkov (Eds.), Current Issues in Linguistic Theory - Recent Advances in Natural Language Processing III, Amsterdam. John Benjamins Publishers.

Preece, J; Rogers, Y; Sharp, H. 2002. Interaction Design. Wiley, 1st Edition, Jan.

Presson, C. C., Montello, D. R. 1998. Points of reference in spatial cognition: Stalking the elusive landmark, British Journal of Development Psychology, 6, 378-381.

Purves, R. and Yang, B. 2005. Graphical query and presentation interface. SPIRIT Technical Deliverable -D26 4301.

Purves, R., Syed, A. K., Yang, B. and Weibel, R. 2005. A cartographic visualization interface for spatial information retrieval. International Cartographic Conference. ICC'05, July, La Coruna, Spain.

Purves, R. S., Clough, P. and Joho, H., 2005. Identifying imprecise regions for geographic information retrieval using the web. In: Billen, R., Drummond, J., Forrest, D., and Joao, E (eds.), Proceedings of the GIS Research UK 13th Annual Conference, 313-318, Glasgow, UK.

Purves, R. S., Clough, P., Jones, C.B., Arampatzis, A., Bucher, B., Finch, D., Fu, G., Joho, H., Syed, A. K., Vaid, S. and Yang, B. 2006. The design and implementation of SPIRIT: a Spatially Aware Search Engine for Information Retrieval on the Internet. International Journal of Geographic Information Science. International Journal of Geographical Information Science, Volume 21, Number 7, January 2007 , 717-745 (29).

Purves, R. S. and Clough, P. D. 2006. Judging spatial relevance and document location for Geographic Information Retrieval, extended abstract. In proceedings of the 4th Interna-tional conference on Geographic Information Science (GIScience'06), Münster, Germany, September 2006, 159-164.

Quigley, Eades, P. 2001. FADE: Graph drawing, clustering, and visual abstraction. Proceedings of the 8th International Symposium on Graph Drawing (GD 2000), 197–210.

Raisz, E. 1934. The rectangular statistical Cartogram. Geographical Review. Vol.24(2). April 1934. 292-296. Available online at http://links.jstor.org/sici?sici=0016-7428%28193404%2924%3A2%3C292%3ATRSC%3E2.0.CO%3B2-W

Rajabifard, A., Williamson, I. P., Holland, P. and Johnstone, G. 2000. From Local to Global SDI Initiatives: a pyramid to building blocks. 4th Global Spatial Data Infrastructure Conference, Cape Town, South Africa, 13-15 March 2000.

Ramroop, S. and Pascoe, R. 1999. Processing of spatial metadata queries within federated GISs. SIRC'99- The 11th Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand, December 13-15th.

Randell, D. A., Cui, Z. and Cohn, A. G. 1992. A Spatial Logic based on Regions and Connections. Proceedings of the 3rd International Conference on Principles of Knowledge Representation and Reasoning (KR'92), Cambridge, MA, 165-176, Morgan Kaufmann.

Raper, J. 1999. Spatial representation: the scientists perspective. In Longley, P., Goodchild, M., Maguire, D. and Rhind, D. (eds.)., Geographical Information Systems: principles, techniques, management, and applications, NewYork: Wiley, volume1, 2 edition, 29-38.

Raper, J., Dykes, J., Wood, J., Mountain, D., Krause, A., Rhind, D. 2002. A framework for evaluating geographical information. Journal of Information Science, 28(1). 39-50.

Rashid, A., Shariff, B. M., Egenhofer, M. J. and Mark, D. M. 1998. Natural-language spatial relations between linear and areal objects: The Topology and Metric of English-Language terms. International Journal of Geographical Information Science, 12(3): 215-246.

Ratcliffe, J. H. 2001. On the accuracy of TIGER type geocoded address data in relation to cadastral and census areal units. International Journal of Geographical Information Science. Vol.15. Issue 5: 473-485.

Ratcliffe, J. H. 2002. Aoristic signatures and the spatio-temporal analysis of high volume crime patterns. Journal of Quantitative Criminology, 2002. Vol.18, Issue 1: 23-43.

Ratcliffe, J. H. 2004. Geocoding crime and a first estimate of a minimum acceptable hit rate. International journal of geographical information science, vol.18. issue 1, 61:72.

Raubal, M. and Winter, S. 2003. Enriching way finding instruction with local landmarks. M. Egenhofer and D. Mark (Eds.). Geographic Information Science, Lecture Notes in Computer Science 2478, Springer-Verlag, 243-259.

Rauch, E., Bukatin, M. and Baker, K., 2003. A confidence-based framework for dis-ambiguating geographic terms. In A. Kornai and B. Sundheim [6], pages 50--54. HLT-NAACL 2003 Workshop on Analysis of Geographic References, Alberta, Canada, 50-54. ACL.

Rautberg, M. and Hof, M. 1995. Metaphor Engineering: a participatory approach. In W.Schuler, J. Hannemann and N. Streitz (Eds.), Designing User Interfaces for Hypermedia, Springer, 1995, 58-67.

Ray, R. L. 1995. Geographic information retrieval and spatial browsing. In Linda C. Smith and Myke Gluck. Editors. Geographic Information Systems and Libraries: Patrons, Maps and Spatial Information. 81-123. April 1995. See http://sherlock.berkeley.edu/geo_ir/ PART1.html

Reape, M. and Mellish, C. 1999. Just what is aggregation anyway? In proceedings of the 7th European workshop on Natural Language Generation, 20-29. Toulouse, France.

Rees, W. E. 2001. The concept of ecological footprint. Encyclopaedia of Biodiversity, Vol. 2.

Reichenbacer, T. 2004. Mobile Cartography- Adaptive visualization of geographic information on mobile devices. PhD Dissertation. Dissertation eingereicht bei: Technische Universität München , Fakultät für Bauingenieurwesen und Vermessungswesen, 2003-11-20

Reid, J., Lalmas, M., Finesilver, K., Hertzum, M. 2006. Best entry points for structured document retrieval- Part-II: Types, usages and effectiveness, Information Processing & Management. 42, 89-105.

Reinbacher, I., Benkert, M., van Kreveld, M. J., Mitchell, J. S. B and Wolff, A. 2005. Delineating boundaries for imprecise regions. In proceedings of 13th European symposium on algorithms (143-154), Springer, Berlin.

Relph, E. 1976. Place and Placelessness. London: Pion.

Rennison, E. and Strausfeld, L. 1995. The Millenium Project: Constructing a Dynamic 3+D Virtual Environment for Exploring Geographically, Temporally and Categorically organized Historical Information. In: Spatial Information Theory: A Theoretical Basis for GIS, International Conference COSIT 95, (Semmering, Austria, September 21-23, 1995), Lecture Notes in Computer Science, Vol. 988, Springer, 69- 91.

Retz-Schmidt, G. 1988. Various views on spatial prepositions. AI Magazine. Summer, 95-105.

Richardson, B. C. 1992. Places and regions. In Abler, R. F., Marcus, M.G. and Olson, J. M. (eds.), Geography's Inner Worlds: Pervasive Themes in Contemporary American Geography, 27-49. New Brunswick, NJ: Rutgers University Press.

Ridley, R. and Gross, J. H. 2004. Geographic text search improves homeland security. UC Proceedings abstract.

Rigaux, P., Scholl, M and Voisard, A. 2002. Spatial databases with application to GIS. Morgan Kaufmann.

Robertson, B. 2002. A developer's guide to the resources of the historical event mark-up and linking project. http://www.heml.org

Robertson, G., Mackinlay, J. D. and Card, S. 1991. Cone trees: Animated 3D visualizations of hierarchical information. In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, pages 189-194. ACM.

Robertson, G., and Mackinlay, J. D. 1993. The document lens. In Proceedings of the 6th Annual ACM

Symposium on User interface Software and Technology (Atlanta, Georgia, United States). UIST ‘93. ACM, New York, NY, 101-108. DOI= http://doi.acm.org/10.1145/168642.168652

Robertson. G, Mackinlay, J, and Card. S. K. 1994. Display of Hierarchical Three-Dimensional Structures with Rotating substructures. US Patent 5295243, Xerox Corporation. Filed 29th Dec. 1989, issued 15th March 1994.

Robertson, S. E. and Hancock-Beaulieu, M. M. 1992. On the evaluation of IR systems. Information Processing & Management, 28(4), 457-466.

Robertson, S. E., Walker, S., Jones, S., Hancock-Beaulieu, M. M. and Gatford, M. 1995. Okapi at Trec-3. In proceedings of third text retrieval conference (TREC-3), 109-126.

Robertson, S. E. and Sparck, J. K. 1997. “Simple, proven approaches to text retrieval”. Technical Report. London, Department of Information Science, City University. Available from http://www.ftp.cl.cam.ac.uk/ftp/papers/reports/. [Accessed 25/06/2003].

Robertson, S. E., Walker, S. and Hancock-Beaulieu, M. M. 1998. Okapi at TREC-7: Automatic adhoc filtering, VLC and Interactive Track. In Voorheer, E.M. and Harman, D.K (eds.), NIST Special publication 500-242. The 7th text retrieval conference (TREC-7), Gaithersburg, MD, 253-264. NIST, Gaithersburg, MD.

Robertson, S. E., Walker, S. and Hancock-Beaulieu, M. M. 2000. Experimentation as a way of life: Okapi at TREC. Information Processing and Management, 36, 95-108.

Robinson. A. H. and Bartz Petchenik. B. 1975. The map as a communication system. Cartography Journal. 12, 7-15.

Rodriguez dos Santos, S. 2004. A framework for the visualization of multidimensional and multivariate data. University of Leeds- School of Computing. September.

Rodriguez, M. A. and Egenhofer, M. J. 2000. A comparison of inferences about containers and surfaces in small-scale and large-scale spaces. Journal of visual languages and Computing. 11(6): 639-662.

Rogerson, P. A. 2001. Statistical methods for geography. SAGE Publications.

Rossi, A. M., Varga, M. 1999. Visualisation of massive retrieved newsfeeds in interactive 3D. In E. Banissi, *et al.* (Eds.), Proceedings of the International Conference on Information Visualization, 1999. (12-17)**.**

Rosson, M. B. and Carroll, J. M. 2001. Usability engineering: scenario-based de-velopment of human computer interaction. Redwood City, CA: Morgan Kaufmann, 2001. 575. ISBN 1558607129.

Ruggles, A. J. and Armstrong, M. P. 1997. Towards a conceptual framework for the cartographic visualization of network information. Cartographica. Vol.34, #1, Spring 1997.

Ruthven, I. 2003. Re-examining the potential effectiveness of interactive query expansion. Proceedings of the 26th ACM-SIGIR Conference, Toronto, Canada, 213-220.

SABE, 2004. At http://www.eurogeographics.org/eng/04_sabe.asp.

Sagara, T., Arikawa, M and Sakauchi, M., 2000. Spatial Information Extraction System using Geo-Reference Information. Information processing society of Japan Journal: Database, vol.41. No. SIG 6 (TOD7), 69-80.

Sahling, G. N. 2002. Interactive 3D scatterplots-from high dimensional data to insights. Masters thesis. Forschungszentrum-VRvis in Wein, Available online @ http://www.vrvis.at/via/resources/DA-NSahling

Salton, G. and McGill, M. 1983. Introduction to Modern Information Retrieval. McGrawHill.

Samet, H. 1990. Applications of Spatial Data Structures, Addison-Wesley.

Sanderson, M., van Rijsbergen, C. J. 1999. The impact on retrieval effectiveness of skewed frequency distributions. ACM Transactions on Information Systems 17(4): 440-465.

Sanderson, M. and Joho, H. 2004. Forming test collection with no system pooling. In proceedings of the 27th annual international ACM SIGIR conference on research and development in information retrieval, ACM, Sheffield, UK.

Sanderson, M. and Kohler, J. 2004. Analyzing geographic queries. In the Proceedings of the workshop on Geographic Information retrieval SIGIR'04.

Saracevic, T. 1984. Measuring the degree of agreement between searchers. In: Flood, B, Witiak, J. & Hogan, T.H., eds. Proceedings of the 47th ASIS annual meeting. White Plains, NY, 227-230.

Saracevic, T. 1995. Evaluation of evaluation in information retrieval. In: Fox, E.A., Ingwersen, P., & Fidel, R., eds. Proceedings of the 18th ACM SIGIR Conference on Research and Development of Information Retrieval. Seattle, 1995. N.Y.: ACM Press, 138-146.

Saracevic, T. 1996. Relevance reconsidered '96. In: Ingwersen, P. & Pors, N.O., eds. Proceedings of CoLIS 2, Second International Conference on Conceptions of Library and Information Science: Integration in Perspective. Copenhagen 1996. Copenhagen: Royal School of Librarianship, 201-218

Sarkar, M. and Brown, M. H. 1992. Graphical fisheye views for graphs. In proceedings of CHI'92. ACM. New York, 83-91.

Sarkar, M. Snibbe, S. and Reiss, S. 1993. Stretching rubber sheet: A metaphor for visualizing large structure on small screen. In Proceedings of the ACM symposium on user interface software and technology, ACM, New York.

Schaffer, D. Zuo, Z. Bartram, L, Dill, J. Dubs, S., Greenberg, S., and Roseman, M. 1993. Comparing fish eye and full zoom techniques for navigation of hierarchically clustered networks. In Proceedings of Graphics Interface '93. Morgan-Kaufman, San Mateo, California.

Schaffer, D., Zuo, Z., Greenberg, S., Bartram, L., Dill, J., Dubs, S., and Roseman, M. 1996. Navigating hierarchically clustered networks through fisheye and full-zoom methods. ACM Trans. Comput.-Hum. Interact. 3, 2 (Jun. 1996), 162-188. DOI= http:// doi.acm.org/10.1145/230562.230577

Schlichtmann, H. ed. 1999. Map semiotics around the world. Working Group on Map Semiotics in the Commission on Theoretical Fields and Definitions in Cartography, International Cartographic Association.

Schober, M. F. 1993. Spatial perspective-taking in conversation. Cognition, 47, 1-24.

Schockaert, S., De Cock, M. and Kere, E. E. 2005. Automatic acquisition of fuzzy footprints.

Schonfelder, S. and Axhausen, K. W. 2003. On the variability of human activity space, Arbeitsbericht Verkehrs und Raumplanung, 149. Institut fur Verkehrsplanung und Trans-portsysteme (IVT), ETH Zurich, Zurich.

Schröder, M., Walessa, M., Rehrauer, H., Seidel, K. and Datcu, M. 1999. Gibbs Random Field Models: A Toolbox for Spatial Information Extraction. Preprint submitted to Elsevier Preprint, 8th March 1999.

Schutz, A. and Luckmann, T. 1973. The structures of the life world. Ecanston, IL, North-western University Press.

Schwartz, C. 1998. Web Search Engines. Journal of the American Society for Information Science, 49(11), 973-982.

Sebrechts, M. M., Cugini, J. V. Laskowski, J. S., Vasilakis, J. and Miller, M. S. 1999. Visualization of search results: a comparative evaluation of text, 2d and 3D interface. Annual ACM Conference on research and development in information retrieval. Proceedings of the 22nd annual international ACM SIGIR conference on Research and Development in Information Retrieval. 3-10.

Selberg, E. W. 1999. Towards Comprehensive Web Search. PhD Thesis, University of Washington, June 1999.

Serrano, S. L., Goni-Menoyo, J. M. and Gonzaley-Cristobal, J. C. 2005. MIRACLE's 2005 Approach to Geographic Information Retrieval.

Shaw, J. and McKeown, K. 1997. An architecture for aggregation in text generation. In JIJCAI, poster Session.

Shepherd, A. 1998. HTA as a framework for task analysis, Ergonomics, 41(11), 1537-1552.

Shepherd, A. 2001. Hierarchical Task Analysis, Taylor and Frances, London.

Shepard, I. F. 1994. Symbols with attitude: time-varying symbolism and data visualization. Paper presented at the British Cartographic Society Annual Technical Symposium, University of Manchester, September.

Sherman, C., 2004. Exploring Search Engine Overlap. http://searchenginewatch.com/ searchday/ article.php/3344081, April 28.

Shipman, F. M., Marshall, C. C., and LeMere, M. 1999. Beyond Location: Hypertext Workspaces and Non-Linear Views. In Proceedings of ACM Hypertext '99, Darmstadt, Germany, 121-130, February.

Shipman, F. M. and Marshall, C. C., 1999. Spatial Hypertext: An Alternative to Navigational and Semantic Links. ACM Computing Surveys, 31(4), December 1999. http://www.cs.brown.edu/memex/ ACM_HypertextTestbed/papers/37.html

Shneidermann, B. 1987. Designing the user interface: Strategies for effective human-computer interaction. Addison-Wesley. Reading. MA.

Shneiderman, B. 1994. Dynamic Querying for visual information seeking. IEEE Software, 11, 70-77.

Shneiderman, B. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the IEEE Symposium on Visual Languages, 336-343, Washington. IEEE Computer Society Press, 1996. http://citeseer.ist.psu.edu/409647.html

Shneiderman, B. 1998. Designing the User Interface. Strategies for Effective Human- Computer Interaction. 3rd edition Reading, MA (Addison-Wesley)

Silva, M., Martins, B., Chaves, M., Cardoso, N. and Afonso, A. P. 2006. Adding Geographic Scopes to Web Resources. Computers, Environment and Urban Systems. Elsevier Science, Vol. 30. 378-399.

Silverstein, C., Marais, H., Henzinger, M. and Morciz, M. 1999. Analysis of a very large web search engine query log. SIGIR Forum, 33, 1, Sep'99. 6-12. DOI=http:// doi.acm.org/10.1145/361604.361633.

Silvestri, F. 2004. High Performance Issues in Web Search Engines: Algorithms and Techniques. PhD Thesis, Dipartimento di Informatica, Universit di Pisa. May 2004

Sinton, D. 1979. The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. in Dutton G. (eds.)., Proceedings of the 1st International Advanced Study Symposium on Topological Data Structures for Geographic Information Systems, Volume 7 (Spatial Semantics: Understanding and Interacting with Map Data) of Harvard Papers on Geographic Information Systems, SINTON/1-SINTON/17.

Skinner, C. W. 1974. A heuristic approach to inductive inference in fact retrieval systems. Commun.

ACM 17, 12 (Dec 1974). 707-712. DOI=http://doi.acm.org/10.1145/ 361604.361633.

Skopik, A. and Gutwin, C. 2004. Finding things in fisheyes: Memorability in distorted spaces.

Skopik, A. 2004. Improving memorability in fisheye views. M.Sc Thesis, Department of Computer Sciences, University of Saskatchewan, Saskatoon, Canada. Available online at http://hci.usask.ca/publications/2004/skopik_msc_thesis/Thesis_final.pdf

Skupin, A. 2002. On geometry and transformation in map-like information visualization. In: Börner, K., and Chen,C. (Eds.) Visual Interfaces to Digital Libraries (Lecture Notes in Computer Science 2359). Berlin: Springer-Verlag. 161-170.

Skupin, A. 2004. The world of geography: visualizing a knowledge domain with cartographic means. PNAS, April 6, Vol.101. PNAS. Suppl.1, 5274-5278

Slocum, T, A. 1983. Predicting visual clusters on graduated circle maps. The American Cartographer, 10(1): 59-72.

Slocum, T. A. 1999. Thematic cartography and visualization. Upper Saddle River, New Jersey: Printice Hall.

Slocum, T. A. and Egbert, S. 1993. Knowledge acquisition from chloropleth maps. Cartography and Geographic Information Systems. 20: 83-95.

Smaalen, J. W. N. V. 2003. Automated aggregation of geographic objects. A new approach to the conceptual generalization of geographic databases. NCG Delft, Optima Grafische Communicatie, November 2003.

Smith, B. 1995. On drawing lines on a map. In Frank, A.U and Kuhn, W and Mark, D.M., Eds. Proceedings Spatial Information Theory, 475-484.

Smith, B. and Mark, D. M. 2001. Geographical categories: an ontological investigation. International Journal of Geographical Information Science, 15(7), 591-612.

Smith, B. and Mark, D. M. 2003. Do Mountains Exist?: Towards an ontology of Landforms. Environment and Planning B: Planning and Design, 20(3), 411-427.

Smith, D. A. and Mann, G. S. 2003. Bootstrapping toponym classifiers. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 45-49, Alberta, Canada: ACL.

Smith, T. R. 1996. A Digital Library for Geographically Referenced Materials. Computer, 29, 5 (May, 1996), 54-60, DOI=http://dx.doi.org/10.1109/2.493457

Smith, T. R., Janee, G., Frew, G., Coleman, A., 2001. The Alexandria Digital Earth Prototype System. ACM, JCDL'01. June 24-28, 118-119. Raonake, Virginia, USA.

Solte, C., Tang. D. and Hanrahan. P. 2002. Polaris: A System for Query, Analysis and Visualization of Multi-dimensional Relational Databases. IEEE Transactions on Visualiza-tion and Computer Graphics, Vol. 8, No. 1, January.

Sommervell, J., McCrickard, D. S., North, C. and Shukla, M. 2002. An evaluation of information visualization in attention-limited environments. In Proceedings of the Symposium on Data Visualization. 2002 (Barcelona Spain, May 27-29 2002). ACM International Conference Proceedings Series, vol. 22. Eurographics Association. Aire-la-ville, Switzerland, 211-216..

Song, M. 2000. Visualization in information retrieval: a three-level analysis. Journal of Information Science, 26(1), 3-19.

Soon Tee Teoh, Kwan-Liu Ma, 2002. RINGS: A Technique for Visualizing Large Hierarchies. Graph Drawing 2002: 268-275.

Sorrows, M. and Hirtle, S. 1999. The nature of landmarks for real and electronic spaces. C. Freska and D. Mark (Eds.). Spatial Information Theory: Cognitive and Computational Foundations of Geographic Information Science, Lecture Notes in Computer Science 1661. Springer-Verlag, 37-55.

Soudack, A. 2003. Don't Test users, Test Hypotheses. Boxes and Arrows. October 27.

Southall, H. 2003. Defining and identifying the roles of geographic references within text: Examples from the Great Britain Historical GIS project. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic Ref-erences, 69-78, Alberta, Canada: ACL.

Sparck, J. K., 1981. The cranfield tests. In: information retrieval experiment, K. Sparck Jones (ed.) London: Butterworths. 256-284.

Sparck, J. K and Willet, P. (eds.) 1997. Readings in information retrieval. Morgan Kaufmann, San Francisco, USA.

Spence, R. and Apperley, M. D. 1982. Database Navigation: An office environment for professional. Behav. Inf. Tech. 1,1, 43-54.

Spertus, E. 1998. ParaSite: Mining the Structural Information on the World-Wide Web. PhD Thesis, Department of EECS, MIT, February 1998.

Sphorer, J. C., 1999. Information in places. IBM systems Journal, vol.38. No.4.

Spink, A. 1997. Study of interactive feedback during mediated information retrieval. Journal of the American Society for Information Science, 48(5), 382-394

Spink, A. 2002. A user-centered approach to evaluating human interaction with web search engines: an exploratory study, Information Processing and Management, 38 (2002), 401-426.

SPSS Inc. 2005. SPSS Base 14.0 for Windows User's Guide. SPSS Inc., Chicago IL.

Stafford, T., and Webb, M. 2004. Mind Hacks, O'Reilly, Nov, ISBN-0-596-00779-5.

Stammers, R. B., and Astley, J. A. 1987. Hierarchical Task Analysis: Twenty Years On. In E, D. Magaw (ed.) Contemporary Ergonomics, Taylor and Francis, London, 135-139.

Steel, D. G. and Holt, D. 1996. Rules for random aggregation. Env. and Planning A. 28. 957-978.

Stone M. C., Fishkin K., Bier E A., 1994. The movable filter as a user interface tool. Proceedings of CHI'94. (Boston, MA, April, 24-28) ACM, New York, 306-312.

Su, T. L. 1992. Evaluation measures for interactive information retrieval. Information Processing and Management, 28:4, 503-516.

Sukhov, V. 1970. Application of Information theory in generalization of map contents. International year book of cartography, X: 41-47.

Sullivan, D., Sherman, C., Price, G. 2004. Search Engine Size Wars V Erupts. Available online at http://blog.searchenginewatch.com/blog/041111-084221.

Summers. K. L., Goldsmith. T. E., Kubica. S., Caudell. T. P., 2003. "An Experimental Evaluation of Continuous Semantic Zooming in Program Visualization," infovis, p. 20, 2003 IEEE Symposium on Information Visualization, 2003.

Sung-Soo Kim, Jong-Hyun Park. 2004. Geographic Hypermedia using Search Space Transformation, International Conference on Pattern Recognition (ICPR'04), Vol.4, 368-371.

Sung-Soo Kim, Jong-Hyun Park. 2004. Linking geographic hypermedia using the remotely sensed data. IEEE International Conference on Multimedia and Expo (ICME), 811-814

Sutcliffe, J.T. 1992. The pragmatics of information retrieval experimentation revisited. Information processing and management. 28:4, 467-490.

Swayne. D. F., Cook. D, Buja.A., 1998. XGobi: Interactive Dynamic Data visualization in the X Window System. Journal of Computational and Graphical Statistics. American Statistical Association. Institute of Mathematical Statistics and Interface Foundation of America., 7(1), 113-130

Swanson, E., Galvao, A. and Sato, K. 2003. A Framework for Understanding Contexts in Interactive Systems Development. Proceedings of 7th World Conference on Systemics, Cybernetics and Informatics, FL, USA, Vol. 4, 405-408.

Syed, A. K., Yang, B., Purves, R., Kopczynski, M. 2003. Query interface design. SPIRIT Technical deliverable-D7 4101. Available online at http://www.geo-spirit.org.

Takahashi, K., Yokoji, S. and Miura, N. 2000. Location oriented integration of internet information. Mobile Info search. In digital cities, technologies, experiences and future perspectives. [the book is based on an international symposium held in Kyoto, Japan in September 1999. T. Ishida and K. Isbister, Eds. Lecture Notes in Computer Science. Vol.1765. Springer Verlag, London, 364-377)

Talmy, L. 1983. How language structures space. In pick, H.L. and Acredolo, L.P. (Eds). Spatial Orientation: Theory, Research and Application, 225-282. New York: Plenum press.

Tang, L. and Crovella, M. 2003. Virtual landmarks for the internet. In proceedings of the 3rd ACM SIGCOMM conference on internet measurement (Miami Beach, FL, USA, October 27-29, 2003). IMC'03. ACM Press, New York, NY, 143-152. DOI=http:// doi.acm.org/10.1145/948205.948223.

Taylor, D. R. F. 1991. Geographic information systems: the microcomputer and modern cartography. In Taylor, D.R.F. (ed.), Geographic Information Systems: the microcomputer and modern cartography. Pergamon Press. Oxford. 1-20.

Taylor, M. M., McCann, C. A., and Tuori, M. I. 1984. The interactive spatial information system (DCIEM No.84-R-22). Downsview, Ontario, Canada: Defence and civil institute of environmental medicine.

Taylor, R. S. 1968. Question-negotiation and information seeking in libraries, College and research libraries. 29, 178-194.

Tezuka, T., Takakura, H., Lee, R., and Kambayashi, Y. 2001. Models of conceptual geographical prepositions based on web resources. Journal of geographic information and decision analysis. Vol.5. No.2., 83-84.

Tezuka, T., Lee, R., Takakura, H. and Kambayashi, Y. 2003. Integrated model for a region-specific search system and its implementation, Proceedings of the International Conference on Internet Information Retrieval (ICIIR 2003), 243-248. Koyang, Korea.

Tezuka, T. Yokota, Y., Iwaihara, M. and Tanaka, K. 2004. Extraction of cognitively-significant place names and regions from web-based physical proximity co-occurrences. X.Zhou, S.Su, M.P.Papazoglou, M.E. Orlowska and K.G. Jeffery (Eds.), Web Information Systems- WISE 2004. Lecture Notes in Computer Science 3306, 113-124, Springer-Verlag, 2004.

Tezuka, T. and Tanaka, K. 2005. Temporal and Spatial Attribute Extraction from Web Documents and Time-Specific Regional Web Search System. In Y-J. Kwon, A. Bouju and C. Claramunt. (Eds.): W2GIS 2004. LNCS 3428, 14-25, 2005. Springer-Verlag Berlin Heidelberg.

Tezuka, T. and Tanaka, K. 2006. Visual description conversion for enhancing search engines and navigational systems. In X. Zhou et al. (eds.): APweb 2006, Lecture Notes in Computer Sciences, 3841, 955-960. Springer-Verlag, Berlin Heidelberg.

Thimbleby, H., 1990. User Interface Design. ACM Press. New York.

Timpf, S. 2003. Geographic activity models. In M. Duckham, M. F. Goodchild & M. F. Worboys (eds.), Foundations of geographic information science (241-254). London: Taylor & Francis.

Tjong Kim Sang, E. F. and De Meulder. F., 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. In W. Daelemans and M. Osborne (eds.), Proceedings of CoNLL-2003, 142–147. Edmonton, Canada.

Tobler, W. 1986. Pseudo-Cartograms, The American Cartographer, 13,1, 43-50.

Tobler, W. 2004, Thirty-five Years of Computer Cartograms, Annals, Assoc. Am. Geographers, March.

Tobon C., 2005. Evaluating Geographic Visualization Tools and Methods: An approach and Experiment based upon user tasks. Chapter 34. In Exploring Geovisualization. J. Dykes, A. MacEachren and M-J. Kraak. Elsevier Ltd. Tag der mündlichen Prüfung: 2004-02-06. Available online at http://tumb1.biblio.tu-muenchen.de/publ/diss/bv/2004/ reichenbacher.pdf

Tom, A. and Denis, M. 2003. Referring to landmark or street information in route directions: What difference does it make? In Kuhn, W., Worbhoys, M., Timpf, S. eds: Spatial Information Theory. Vol: 2825 of Lecture Notes in Computer Science, Springer, Berlin, 384-397.

Tombros, A., Sanderson, M. 1998. Advantages of Query biased summaries in information retrieval. Proceedings of the 21st ACM SIGIR conference. Melbourne, Australia, 2-10.

Tomko, M. and Winter, S. 2005. Reconstruction of Scenes from Georeferenced Web Resources. Proceedings of SSC 2005 Spatial Intelligence. Innovation and Praxis: The national biennial conference of the Spatial Science Institute. Sept'05. Melbourne: Spatial Sciences Institute. ISBN-09581366-2-9.

Treinish, L. A., 1999, Task-Specific Visualization Design, IEEE Computer Graphics and Applications, vol. 19, no. 5, 72-77, Sept/Oct, 1999.

Treisman, A., and Sato, S. 1990. Conjunction search revisited. Experimental Psychology: Human Perception and Performance. 16: 459-478.

Tsaparas, P. 2004. Link Analysis Ranking. PhD Thesis, University of Toronto, March 2004.

Tsotras, V., Jensen, C. and Snodgrass, R. 1998. An extensible notation for spatiotemporal index queries. ACM SIGMOD Record 27, 47-53.

Tsukamoto, M. 2000. Image based Pseudo-3D Visualization of Real Space on WWW: In Digital Cities, Technologies, Experiences, and Future Perspectives [the book is based on an international symposium held in kyoto, japan in September 1999. T. Ishida and K. Isbister. Eds. Lecture Notes in Computer Science, Vol. 1765. Springer-Verlag, London, 288-302.

Tufte. E. 1990. Envisioning Information. Graphics Press.

Turnbull D. 2003. New Approaches for Studying and Building Information Seeking Models: A Possible Hybrid Approach. SIG USE Research Symposium at ASIST '03 in Long Beach, CA. Oct 18, 2003. Available online at http://www.ischool.utexas.edu/~donturn/ research/SIGUSE2003Worksheet.html.

Tweedie. L, 1997. Characterizing interactive externalizations, Proceedings of the SIGCHI conference on Human factors in computing systems, 375-382, March 22-27, 1997, Atlanta, Georgia, United States.

Uejima, H., Miura, T., and Shioya, I. 2004. Estimating timestamp from incomplete news corpus. Communications in information and systems. International Press.Vol.4, No.4, 273-288.

Unwin, A. and Unwin, D. 1998. Exploratory spatial data analysis with local statistics. The statistician. Vol.47. No.3. 415-421.

Upstill T. G. 2005. Document ranking using web evidence. PhD Thesis, The Australian National University. August 2005.

Uryupina, O. 2003. Semi-supervised learning of geographical gazetteers from the internet. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 18-25, Alberta, Canada: ACL.

Vaid, S., Jones, C. B., Joho, H. and Sanderson, M. 2005. Spatio-textual indexing for geographic search on the web. In Proceedings of the 9th international symposium on spatial and temporal databases.

van Kreveld, M., van Oostrum, R and Snoeyink, J. 1997. Efficient settlement selection for interactive display. In proceedings of Auto-carto 13: ACSM/ASPRS Annual Convention Technical papers, 287-296.

van Kreveld, M. and Speckmann. B. 2004. On Rectangular Cartograms. S. Albers and T. Radzik (eds.): ESA 2004, LNCS, 3221, 724-735, Springer- Verlag, Berlin, Heidelberg, 2004.

van Kreveld, M., Schramm, E., Wolff, A. 2004. Algorithms for the placement of diagrams on maps. In proceedings of the 12th annual ACM International workshop on Geographic Information Systems. (Washington DC,USA, November 12-23. 2004) , GIS'04. ACM Press, New York, NY. 222-231. DOI=http://doi.acm.org/10.1145/ 1032222.1032255

van Kreveld, M. J., Reinbacher, I., Arampatzis, A., and van Zwol, R. 2004. Distributed ranking methods for geographic information retrieval. In P.F. Fisher (ed.), Developments in spatial data handling: 11 th international symposium on spatial data handling, 231-243.

van Kreveld, M. J. & Reinbacher, I. 2004. GoodNEWS: Partitioning a simple polygon by compass directions. International Journal of Computational Geometry & Applications, 14, 233-259.

van Kreveld, M., Reinbacher, I., Arampatzis, A. and Van Zwol. R. 2005. Multi-Dimensional Scattered Ranking Methods for Geographic Information Retrieval, GeoInformatica 9: 1, 61-84, 2005.

van Rijsbergen, C. J. and Lalmas, M. 1998. Information Calculus for Information Retrieval. Journal of the American Society for Information Science. Vol.47. No.5. 385-398.

van Rijsbergen, K. 2004. Geometry of Information Retrieval. Cambridge University Press, Sept'04.

van Zwol, R., van Kreveld, M. J., Reinbacher, I. and Veltkamp, R. C. 2003. Aspects of spatial similarity measures. Utrecht, The Netherlands: Institute of Information and Computing Sciences, Utrecht University.

Vechtomova, O., Robertson, S. E., Jones, S. 2003. Query expansion with long-span collocates. Information Retrieval, 6(2), 251-273.

Vechtomova, O., Karamuftuoglu, M., Lam, E. 2004. Interactive Search Refinement Techniques for HARD Tasks. Proceedings of the Twelfth Text Retrieval Conference, November 18-21, 2003. Gaithersburg, MD.

Vechtomova, O., and Karamuftuoglu, M. 2004. Comparison of two interactive search refinement techniques. Proceedings of the human language technology of the North American Chapter of the Association for Computational Linguistics: HLT-NAACL.

Veerasamy, A., Heikes, R. 1997. Effectiveness of a graphical display of retrieval results. In: Belkin, Nicholas J.; Narasimhalu, A. Desai; Willett, Peter (eds.): SIGIR 1997: Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Conference: Philadelphia, PA, July 27-31 1997. New York (ACM Press) 1997. 236-245.

Vinson, N. G. 1999. Design Guidelines for landmarks to support Navigation in Virtual En-vironments. In proceedings of CHI'99. ACM Press, 278-285.

Von Wyss, M. 1996. The production of smooth scale changes in an animated map project. Cartographic Perspectives. 23: 12-20.

Voorhees, E. M. and Harman, D. 2000. Overview of the sixth Text RETrieval Conference (TREC-6).

Information Processing and Management, 36(1), 3-35.

Wainer, H., and C. Francolini. 1980. An empirical inquiry concerning human understanding of two-variable colour maps. The American Statistician. 34: 81-93.

Waldinger, R., *et al.* **2003**. Pointing to places in a deductive geospatial theory. In: Kornai, A. and Sundheim, B. (eds.) Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References, 10-17, Alberta, Canada: ACL.

Walton, D. N. 1995. A pragmatic theory of fallacy. Tuscaloosa. University of Alabama Press.

Wang, C., Xie, X. et al. 2005. Detecting geographic locations from web resources. Proceedings of the 2005 Workshop on Geographic Information Retrieval. Bremen, Germany.

Wang, C., Xie, X., Wang, L., Lu, Y and Ma Wei-Ying. 2005. Web Resource Geographic location classification and detection. WWW'05. May 10-14, Chiba, Japan, ACM 1-59593-051-5/05/2005.

Wang, F. 2000. A fuzzy grammar and possibility theory-based natural language user interface for spatial queries. Fuzzy sets and systems, Vol.113, 147-159.

Ward, M. O., 1994. XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data. In R. Daniel Bergeron and Arie E. Kaufman, editors, Proceedings of the Conference on Visualization, pages 326-336, Los Alamitos, CA, USA. IEEE Computer Society Press.

Ware, J. M., Anand, S., Taylore, G. E. and Thomas, N. 2006. Automated productions of schematic maps for mobile applications. Transactions in GIS, 10(1), 25-42.

Waterworth, J. A., and Chignell, M. H. 1991. A model for information exploration. Hypermedia. 3(1). 35-58.

Wattenberg, M. 2005. Baby Names, Visualization and Social Data Analysis. Infovis. Proceedings of the 2005 IEEE Symposium on Information visualization (INFOVIS '05). p.1. http://www.babynamewizard.com/namevoyager/lnv0105.html.

Watters, C. and Amoudi, G. 2002. GeoSearcher: Location-Based Ranking of Search Engine Results. Journal of the American Society for Information Science and Technology, 54(2), 140-151.

Weerahandi, S. 2004. Generalized Inferences in Repeated Measures. John Wiley & Sons, Inc.

Wehrend, S. 1990. A categorization of scientific visualization techniques. MS Thesis, Department of Computer Science, University of Colorado, Boulder, CO.

Wehrend, S. and Lewis, C., 2000. A problem-oriented classification of visualization techniques. Proceedings of the 1St IEEE conference on Visualization, 1990, 139-143, San Francisco, California.

Westphal, C., and Blaxton. T., 1998. Data Mining solutions. New York, New York: Wiley.

Wiegand, N. Zhou, N., Ventura, S. and Cruz, I. F. 2003. Extending XML web querying to Heterogeneous Geospatial Information. Proceedings Annual International Conference on Digital Government Research, Boston, MA, USA, 279-282.

Wildbur, P. and Burke, M. 2000. Information Graphics: Innovative Solutions in Contemporary Design.

Wilhelm, A. and Steck, R. 1998. Exploring spatial data by using interactive graphics and local statistics. The statistician. Vol.47. No.3, 423-430.

Wilkinson. L. 1999. The grammar of graphics. Newyork: Springer-Verlag.

Williamson, C. and Shneiderman, B. 1992. The Dynamic Home Finder: Evaluating Dynamic Queries in a Real-Estate Information Exploration System. 15th Annual International SIGIR'92, Denmark, ACM.

Wills. G. J. 1996. Ways to Say ``This is Interesting''. In Proceedings of the IEEE Symposium on Information Visualization 1996 (InfoVis 1996), 54-61. IEEE Computer Society Press.

Wilson, T. D. 1981. On user studies and information needs. Journal of Documentation, 37(1), 3-15.

Wilson, T. D. 1999. Models in information behaviour research. Journal of Documentation. 55(3), 249-270.

Winckler, M. A., Palanque, P., and Freitas, C. M. 2004. Tasks and scenario-based evaluation of information visualization techniques. In Proceedings of the 3rd Annual Conference on Task Models and Diagrams (Prague, Czech Republic, November 15 - 16, 2004). TAMODIA '04, vol. 86. ACM Press, New York, NY, 165-172. DOI= http://doi.acm.org/ 10.1145/1045446.1045475.

Wise, J. A., Thomas, J. J., Pen-nock, Kelly, Lantrip, David, Pottier, Marc, and Schur, Anne. 1995. Visualizing the non-visual: Spatial analysis and interaction with in-formation from text documents. In Proc. of the Information Visualization Symposium, 51-58. IEEE Computer Society Press.

Wilkie R. W. 1976. Maps and cartograms, Statistical abstracts of Latin America, Vol.17, 1-23.

Wilkie R. W. 1977. Population cartograms and political subdivisions: Latin American populations in the 1970s, Statistical abstracts of Latin America, Vol.18, 1-26.

Wilkie R. W. 1980. Latin American cartogram, Statistical abstracts of Latin America, Vol.21, p.xxvii.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. 1977. Field dependent and field independent cognitive styles and their educational implications. Review of Educational Research, 47, 1-64.

Wiza, W., Walczak, K., and Cellary, W. 2004. Periscope: a system for adaptive 3D visualization of search results. In Proceedings of the Ninth international Conference on 3D Web Technology (Monterey, California, April 05 - 08, 2004). Web3D '04. ACM Press, New York, NY, 29-40. DOI= http://doi.acm.org/10.1145/985040.985045.

Wood, C. H. 1993. Visual search centrality and minimum map size. Cartographica. 30(4): 32-44.

Woods, J. and Douglas, W. 1989. Fallacies: Selected Papers 1972-1982. Dordrecht, Holland & Providence, RI: Foris.

Woodruff, A. G. and Plaunt, C. 1994. GIPSY: automated geographic indexing of text documents. J.American Society Information Science. 45, 9 (Oct, 1994), 645-655. DOI=http://dx.doi.org/10.1002/(SICI)1097-4571(199410) 45:9 <645:AID-ASI2>3.0, CO, 2-8.

Woodruff, A. G., and Olston, C., 1998. Iconification and Omission in Information Exploration. CHI'98. Workshop on Innovation and Evaluation in information exploration interfaces. Los Angeles. April.

Worbhoys, M. F. 1994. Object oriented approaches to geo-reference information. International Journal of Geographical Information Systems. 8(40), 385-399.

Worbhoys, M. F. 2005. Event-oriented approaches to geographic information. International Journal of Geographic Information Science, 19,1-28.

Worboys, M. F. 2001. Nearness relations in environmental space. International Journal of Geographic Information Science, 15, 7, 633-651.

Wurman, R. S. 1989. Information Anxiety. DoubleDay: Newyork.

Xie, H. 2000. Shifts of Interactive Intentions and Information-seeking strategies in interactive information retrieval, Journal of the American society of information science. 51(9): 841-847.

Xie, N., Cao, C., Guo, H. Y. 2005. A knowledge fusion model for web information. WI. The 2005 IEEE/WIC/ACM International Conference on web intelligence (WI'05). http://doi.ieecomputersociety.

org/10.1109/WI.2005.4

Xu, W., Li, W., Wu, M., Li, W and Yuan, C. Deriving event relevance from the ontology constructed with formal concept analysis. Lecture Notes in Computer Sciences-3878.

Yamada, N., Lee. R., Kambayashi, Y., 2002. Classification of web pages with geographic scope and level of details for mobile cache management. Proceeding of the Third International Conference on Web Information System Engineering (Workshops). IEEE.

Yang K. 2002. Combining Text-, Link-, and Classification-based Retrieval Methods to Enhance Information Discovery on the Web. PhD thesis, University of North Carolina, May 2002.

Yang, B., Purves, R. S., Syed, A. K. and Weibel, R. 2006. Web-based visualization tools for spatial information retrieval.GISRUK'06. April 5-7, 2006.

Yang, Y., Carbonell, J. G., Brown, R. D., Pierce, T., Archibald, B. T., Liu, X. 1999. Learning approaches for detecting and tracking news events. Intelligent Systems and Their Applications. IEEE [See also IEEE Intelligent Systems] Volume 14, Issue 4, July-Aug. 1999. 32-43.

Yattaw, N. J. 1999. Conceptualizing space and time: a classification of geographic movement. Cartography and Geographic Information Science, 26, 85-98.

Yip, K. and Zhao, F. 1996. Spatial Aggregation: Theory and Applications. Journal of Artificial Intelligence Research, 5, 1-26. AI Access foundation and Morgan Kaufmann Publishers.

Zacks, J., Tversky, B. 2001. Event structure in perception and conception. Psychological Bulletin. 127, 3-21.

Zamir, O. E. 1999. Clustering Web Documents: A Phrase-Based Method for Grouping Search Engine Results. PhD Thesis, University of Washington, 1999.

Zhai, Cheng, Xiang and Lafferty, J. 2006. A risk minimizing framework for information retrieval. Information processing and management. 42, 31-55.

Zhang, JingXiang. 2002. Uncertainity in geographical information. London: Taylor & Franacis, Research Monographs in GIS.

Zhang, P. 2000. The effects of animation on information seeking performance on the World Wide Web: securing attention or interfering with primary tasks? Journal of the association for Information Systems. 1:1, March.

Zhang, V. W., Rey, B., Stipp, E and Jone, R. 2006. Geomodification in query rewriting. GIR'06, Held in conjunction with ACM SIGIR, Seattle, Aug 10th. 23-27.

Zhao, H. and Shneiderman, B. 2002. Image-based highly interactive web mapping for georeferenced data publishing. ISR Technical Reports-TR 2005-35.

Zhou, Y., Xie, X., Wang, C., Gong, Y., and Ma, W. 2005. Hybrid index structures for location-based web search. In proceedings of the 14th ACM international Conference on information and Knowledge Management (Bremen, Germany, October 31 - November 05, 2005). CIKM '05. ACM Press, New York, NY, 155-162. DOI= http:// doi.acm.org/10.1145/1099554.1099584

Zien, J., Meyer, J., Tomlin, J. and Liu, J. 2000. Web query characteristics and their implication on search engines. Almaden Research Center, Research Report, RJ 10199 (95073).

Zubin, D. 1989. Natural language understanding and reference frames. In: Mark, D., Frank, A., Egenhofer, M., Freundschuh, S., Mcgranaghan, M. & White, R (eds.). Languages of spatial relations: Initiative 2 specialist meeting, Santa Barbara, CA: National Center of Geographic Information and Analysis. Technical Report, 2A, 13-16.

Appendix A

Publications

- Ross S. Purves, Paul Clough, Christopher, B. Jones, Avi Arampatzis, Benedicte Bucher, David Finch, Gaihua Fu, Hideo Joho, Awase Khirni Syed, Subodh Vaid and Bisheng Yang. The design and implementation of SPIRIT: A Spatially Aware Search Engine for Information Retrieval on the Internet. International Journal of Geographic Information Science, Volume 21, Issue 7 January 2007, pages 717-745.
- Bisheng Yang, Ross S. Purves, Awase Khirni Syed and Robert Weibel. Web-based visualization tools for Spatial Information Retrieval, GISRUK'06, April 5-7, 2006.
- Ross S. Purves, Awase Khirni Syed, Bisheng Yang and Robert Weibel. A Cartographic Visual Interface for Spatial Information Retrieval, XII International Cartographic Conference (ICC 2005) Coruna, Spain, 11-16, July, 2005.
- Benedicte Bucher, Paul Clough, Hideo Joho, Ross S. Purves and Awase Khirni Syed, Geographic IR Systems: Requirements and Evaluation. XII International Cartographic Conference (ICC 2005), Coruna, Spain, 11-16, July, 2005.
- Geovisual representations for Spatially-Aware Information Retrieved from the Internet. EuroConference on Methods to Support Interaction in Geovisualization Environments. Kolymbari, Crete, Grece, 13-18, March, Short Paper Presentation.

Appendix B

Curriculum-vitae

Syed Awase Khirni

born November 3rd 1975 in Vijayawada, Andhra Pradesh, India.

Citizen of Republic of India.

Education	
Sep 21 2006 - till date	Faculty member in Geographic Information Science at International Institute of Information Technology, Bangalore, India.
Sep'02-Aug'06	Research Assistant and PhD Candidate on SPIRIT (sponsored by EU/IST 5th Framework Programme), GIS Division, Dept of Geography, University of Zurich, Switzerland.
Aug'99-Dec'00	M.E. (Civil Eng), Birla Institute of Technology & Sciences, Pilani, Rajasthan, India-333031.
July '93-July '97	B.E. (Civil Eng), Karnatak University, Dharwad, Karnataka, India
Aug '91-May '93	Board of Intermediate Education, Andhra Pradesh, India
1991	Central Board of Secondary Education Certificate, New Delhi- N.St. Mathews Public School, Vijayawada, Andhra Pradesh, India.

Appendix C

Internet Statistics

Table C.1 Top 20 Countries with highest number of internet users

Country or Region	Internet users, Latest Data	Population (2006 Est.)	Internet Penetration	Source and Date of Latest Data	% Users of World
1.USA	205,326,680	299,093,237		Nielsen/NR Jan'06	
2.China	111,000,000	1,306,724,067		CNNIC Dec/05	
3.Japan	86,300,000	128,389,000		eTForecasts Dec/05	
4.India	50,600,000	1,112,225,812		C.I.Almanac Mar/05	
5.Germany	48,721,997	82,515,988	59.0%	Neilsen/NR Jan'06	
6.UK	37,800,000	60,139,274	62.9%	ITU/Oct05	3.7%
7.Korea (South)	33,900,000	50,633,265	67.0%	eTForecast Dec/05	3.3%
8.Italy	28,870,000	59,115,261	48.8%	48.8% ITU Sept/05 2.8%	2.8%
9.France	26, 214,173	61,004,840	43.0%	Neilsen/NR Jan'06	2.6%
10.Brazil	25,900,000	184,284,898	14.1%	eTForcasts Dec/05	2.5%
11.Russia	23,700,000	143,682,757	16.5%	eTForcasts Dec/05	2.3%
12.Canada	21,900,000	32,251,238	67.9%	eTForcasts Dec/05	2.2%
13.Indonesia	18,000,000	221,900,701	8.1%	eTForcasts Dec/05	1.8%
14.Spain	17,142,198	44,351,186	38.7%	Nielsen/NR Jan'06	1.7%

Notes: (1). World Internet User Statistics were updated as of March 31, 2006. (2) Data for users in individual countries and regions may be found by clicking each country name. (3) Population numbers are based on data contained in the world-gazetteer page. (4) The most recent user information comes from data published by Neilsen/NetRatings, ITU and other trustworthy research sources. (Source : InternetWorldStats.com)

Table C.1 Top 20 Countries with highest number of internet users

Country or Region	Internet users, Latest Data	Population (2006 Est.)	Internet Penetration	Source and Date of Latest Data	% Users of World
15.Mexico	16,995,400	105,149,952	16.2%	AMIPCI Nov/05	1.7%
16.Australia	14,189,557	20,750,052	68.4%	Nielsen/NR Jan'06	1.4%
17.Taiwan	13,800,000	22,896,488	60.3%	C.I.Almanac Mar/05	1.4%
18.Netherlands	10,806,328	16,386,216	65.9%	Nielsen/NR Jan'06	1.1%
19.Poland	10,600,000	38,115,814	27,.8%	ITU Sept 05	1.0%
20.Turkey	10,220,000	74,709,412	13.7%	IWS-Mar.31/06	1.0%
Top 20 Countries	811,986,333	4,064,319,458	20.0%	IWS-Mar.31/06	79.45
Rest of the world	210,876,974	2,435,377,602	8.7%	IWS-Mar.31/06	20.6%
Total world users	1,022,863,307	6,499,697,060	15.7%	IWS-Mar.31/06	100.0%

Notes: (1). World Internet User Statistics were updated as of March 31, 2006. (2) Data for users in individual countries and regions may be found by clicking each country name. (3) Population numbers are based on data contained in the world-gazetteer page. (4) The most recent user information comes from data published by Neilsen/NetRatings, ITU and other trustworthy research sources. (Source : InternetWorldStats.com)

Table C.2 World Internet Usage and Population Statistics (Source : InternetWorldStats.com)

World Regions	Population (2006 Est.)	Population % of World	Internet Usage Latest data	% Population (Penetration)	Usage % of the world	Usage Growth 200-2005
Africa	915,210,928	14.1%	23,649,000	2.6%	2.3%	423.9%
Asia	3,667,774,066	56.4%	380,400,713	10.4%	36.5%	232.8%
Europe	807,289,020	12.4%	294,101,844	36.4%	28.2%	179.8%
Middleast	190,084,161	2.9%	18,203,500	9.6%	1.7%	454.2%
North America	331,473,276	5.1%	227,470,713	68.6%	21.8%	110.4%
Latin America	553,908,632	8.5%	79,962,809	14.7%	7.8%	350.5%
Oceanic/ Australia	33,956,977	0.5%	17,872,707	52.6%	1.7%	134.6%
World total	6,499,697,060	100%	1,043,104,886	16.0%	100%	189%

Notes: (1). World Internet User Statistics were updated as of March 31, 2006. (2) Data for users in individual countries and regions may be found by clicking each country name. (3) Population numbers are based on data contained in the world-gazetteer page. (4) The most recent user information comes from data published by Neilsen/NetRatings, ITU and other trustworthy research sources. (Source : InternetWorldStats.com)

Table C.3 Top ten languages used in the web (Source : InternetWorldStats.com)

Top ten languages in the internet	% of all internet users	Internet users by language	Internet Penetration by Language	Internet Growth for Language (2000-2005)	World Population 2006 Estimate for the language
English	30.0%	312,924,679	27.8%	128.1%	1,125,664,397
Chineese	13.8%	144,301,513	10.8%	346.7%	1,340,767,863
Japaneese	8.3%	86,300,000	67.2%	83.3%	128,389,000
Spanish	7.5%	78,166,075	17.9%	216.7%	429,293,281
German	5.6%	58,214,778	60.7%	110.9%	95,982,043
French	4.4%	45,807,499	12.0%	275.5%	381,193,149
Korean	3.2%	33,900,000	45.8%	78.0%	73,945,860
Portugese	3.1%	32,372,000	14.0%	327.3%	230,846,275
Italian	2.8%	28,87,000	48.8%	118.7%	59,115,261
Russian	2.3%	23,700,000	16.5%	664.5%	143,682,757
Top ten languages	81.0%	844,556,544	21.0%	160.7%	4,008,879,867
Rest of the world langauges	19.0%	198,548,342	8.0%	436%	2,490,817,193
World total	100%	1,043,104,886	16.0%	189.0%	6,499,697,060

Notes: (1). World Internet User Statistics were updated as of March 31, 2006. (2) Data for users in individual countries and regions may be found by clicking each country name. (3) Population numbers are based on data contained in the world-gazetteer page. (4) The most recent user information comes from data published by Neilsen/NetRatings, ITU and other trustworthy research sources. (Source : InternetWorldStats.com)

Appendix D

Early stage requirements gathering exercises

Testing of Geo-search Interfaces
SPIRIT Consortium

Dear Sir/Ma'm,

Many thanks for volunteering; your participation will helps us in designing a web based geographical information retrieval system. We are a group of European researchers working together on a common project, SPIRIT. Our final goal is to design it to meet user's needs.

The set of activities you do, will allow us to understand better, how current geo-search engines and text search engines are used and perceived. You will receive some scenarios describing a situation where a person needs to search for geographical information and you could use a set of tools to solve the problem.

Before searching, you will be requested to fill in questionnaire and after the search to answer few questions on your experience with different tools.

This is roughly the schema of your participation lasting for 1 hour:

- Questionnaire (15 minutes)
- Self-learning process of some selected geo-search engines on the web (15 minutes)
- 3 searching (45 minutes, 15 minutes each).

Please feel free to ask any information or clarification at any time of the evaluation.

We thank you very much for your co-operation.

The SPIRIT Consortium.

You can contact us:

Ross Purves

GIS Division, Dept of Geography,

University of Zürich-Irchel,

rsp@geo.unizh.ch

Syed Awase Khirni

GIS Division, Dept of Geography,

University of Zürich-Irchel

sak@geo.unizh.ch

Job

Describe your job:

Company / Institution:

Years of Experience:

Education and Qualification:

Age

Less than 25	25-34	35-44	45-54	over 54

Sex	Male	Female	

Which languages(s) do you speak, read or write?

(List the language(s) and tick the box that apply).

	Native	Fluent	Fair	Basic

How often do you use another language than your native one?

(tick all that apply and write the language(s) in the space provided).

Daily	Seveal times a week	Once a week	Occasionally	Never

Do you search in foreign languages?

Yes	No

If “YES”, please specify

As a computer user, do you consider yourself: (tick one)

1 Novice	2	3	4	5 Expert

Do you use technological equipment other than a desktop PC to access the internet?				
Yes		No		
If "YES", please What do you use?				
Laptop	Personal Organizer (as Palm Pilot)	PDA (such as Nokia Communicator)	WAP	Others, Please specify

As a searcher do you consider yourself:				
1 Novice	2	3	4	5 Expert

How often do you search for information for work purposes?				
Daily	Seveal times a week	Once a week	Occasionally	Never

If you do not get the result you expect, you normally:
(tick all that apply)
Slightly change the query and search again
Completely change the query and search again
Switch to an "advanced" search mode
Change the search engine
Ask somebody else for help
GIve up
Other, please specify

Do you use directories as support for searching (e.g. categories of objects)?	
Yes	No

Which web search engines do you use?
(tick all that apply and give a percentage of use)
Altavista
Yahoo
NothernLight
Excite
About
Ask Jeeves
Go
Kartoo.net
Others, please specify

Have you ever used geographical information services on the web?

Never	
Multimap	How often?
Map24.com	How often?
Somewherenear.com	How often?
Upmystreet.com	How often?
Mapquest.com	How often?
Mapblast.com	How often?

If you had used one of the services above, in which context?
Planning a business travel
Planning a leisure travel
Looking for accomodation
To find a map of a place
To find a service on the map
Others, please specify

Essential information on geographical services on the web

Multimap.com provides maps and location-based services, mainly in Europe. Multimap offers street-level maps; travel directions; aerial photographs; and other local information (e.g.hotels, restaurants).

Map24.com offers an interactive map navigation of Europe. It allows to browse the map without asking for precise address. In addition to maps a graphical routing service is available.

Mapquest.co.uk is another map provider. It offers driving directions too.

Upmystreet.com finds services (e.g. shops, hotels, etc.) ont he bases of a specific addtress.

Essential information on text based search on the web

Kartoo.com a text search engine with a special visualization interface of the result. We would like you to test for us.

Google.com a text search engine which is widely used.

You should now help us understand how a geographical search service has to be designed on the basis of your judgement on currently available map services and search engines. Listed below are few scenarios that give a description of a situation and a goal to achieve. We request you to document your thinking process in each step, by writing it down, under the space provided. Please take some time to read the scenarios and answer a few questions.

Scenario-I

You are planning your two week vacation in the Black Forest in Germany. You would like to go cycling thus have to identify the best route cycling around. You would also include visits to small cities as well as big city in your travel plan. Possibly include in your tou a visit to home town of Jan Ulrich, the winner of Tour de France, 1997.

1.Based on the scenarios given above. How do you wish to proceed? Which are the tools you think are useful?

2.Which keywords would you use for searching ?

3.How would you define your intermediate goals to reach your final objective? Please list the intermediate goals?

4.Do you prefer to use free text based search interface as a starting point or map based interface as a starting point?

5.Do you think the information provided above was clear enough, for you to identify your intermediate goals?

6.Please specify the queries formulated by you for the scenario and document your interaction during the search sessions? For e.g. Hotels in Zürich (Google)

Scenario-II

You are planning to spend your vacation on the mediterranean coast of France and wish to stay in reasonably large city with good restaurants, commercial shopping centers. You would prefer the accomodation in a city on the coast since you wish to practice windsurfing in a nearby place. You wish to hire the equipment for windsurfing thus you need a fully equipped beach service.

1.Based on the scenarios given above. How do you wish to proceed? Which are the tools you think are useful?

2.Which keywords would you use for searching ?

3.How would you define your intermediate goals to reach your final objective? Please list the intermediate goals?

4.Do you prefer to use free text based search interface as a starting point or map based interface as a starting point?

5.Do you think the information provided above was clear enough, for you to identify your intermediate goals?

6.Please specify the queries formulated by you for the scenario and document your interaction during the search sessions? For e.g. Hotels in Zürich (Google)

Scenario-III

You want to go on a trip to the Swiss Alps with your family. You have heard from a friend that there is some nice mountaineering near a place called Aigle, and also some pleasant walks that you could do with your children (who are aged 7 and 9 respectively).
You want to book a chalet to stay in with your family, and also get information about mountaineering routes, rock climbing and hiking in the area. You have heard of one mountain called Les Diablerets with nice landscapes.

1.Based on the scenarios given above. How do you wish to proceed? Which are the tools you think are useful?

2.Which keywords would you use for searching ?

3.How would you define your intermediate goals to reach your final objective? Please list the intermediate goals?

4.Do you prefer to use free text based search interface as a starting point or map based interface as a starting point?

5.Do you think the information provided above was clear enough, for you to identify your intermediate goals?

6.Please specify the queries formulated by you for the scenario and document your interaction during the search sessions? For e.g. Hotels in Zürich (Google)

Note: Please check if you have answered all the questions. If you have competed all the questionnaires, please proceed to the interview.

Appendix E

Snap shots of various interfaces widely used

hotels in zurich, Switzerland

Search: ● the web ○ pages from India

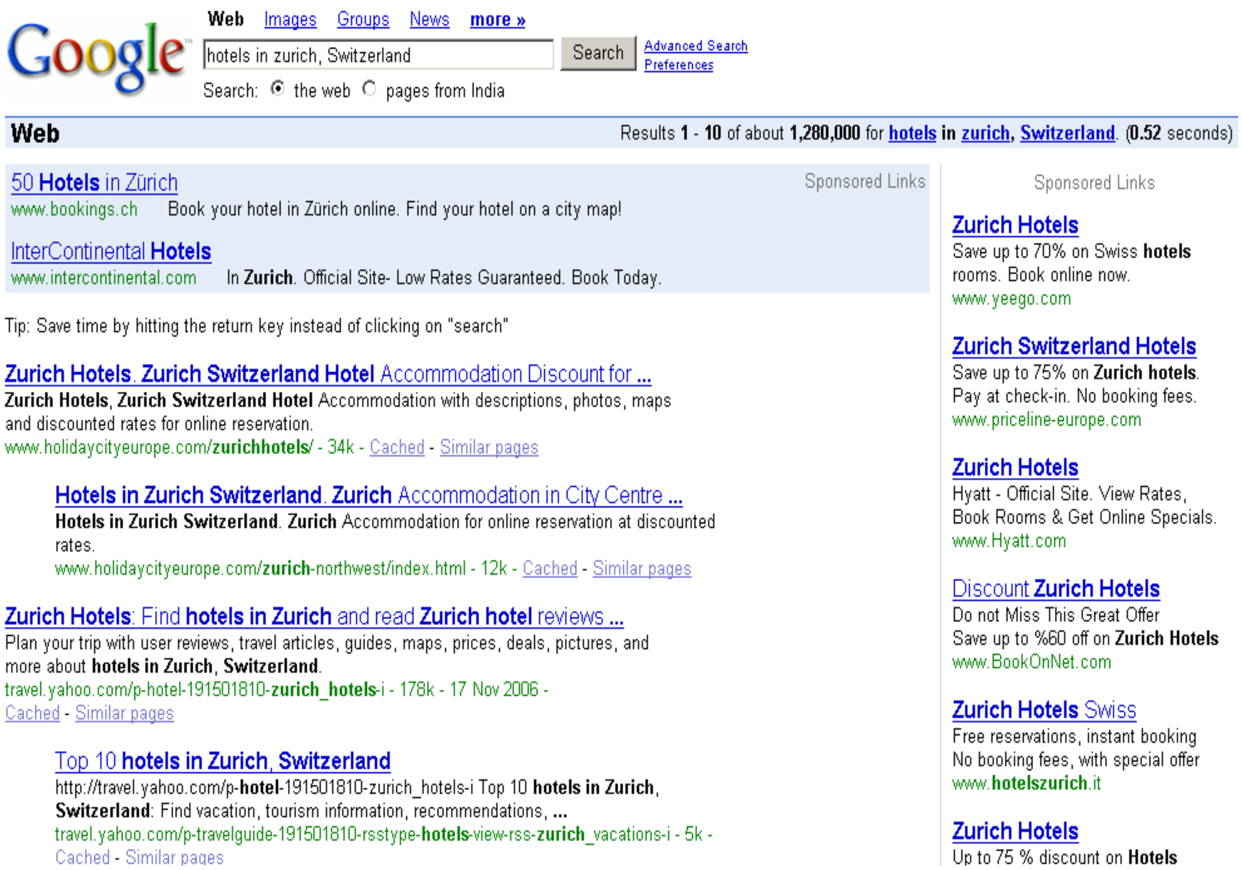
Google.co.in offered in: Hindi Bengali Telugu Marathi Tamil

Advertising Programs - About Google - We're Hiring - Go to Google.com

Make Google Your Homepage!

©2006 Google

Figure E.1 Simple Text based starting point Google (www.google.com)



hotels in zurich, Switzerland

Search: ● the web ○ pages from India

Web Results **1** - **10** of about **1,280,000** for **hotels in zurich, Switzerland**. (**0.52** seconds)

50 **Hotels** in Zürich
www.bookings.ch Book your hotel in Zürich online. Find your hotel on a city map!

InterContinental **Hotels**
www.intercontinental.com In **Zurich**. Official Site- Low Rates Guaranteed. Book Today.

Sponsored Links

Tip: Save time by hitting the return key instead of clicking on "search"

Zurich Hotels, **Zurich Switzerland Hotel** Accommodation Discount for ...
Zurich Hotels, **Zurich Switzerland Hotel** Accommodation with descriptions, photos, maps and discounted rates for online reservation.
www.holidaycityeurope.com/**zurichhotels**/ - 34k - Cached - Similar pages

Hotels in Zurich Switzerland. **Zurich** Accommodation in City Centre ...
Hotels in Zurich Switzerland. **Zurich** Accommodation for online reservation at discounted rates.
www.holidaycityeurope.com/**zurich**-northwest/index.html - 12k - Cached - Similar pages

Zurich Hotels: Find **hotels in Zurich** and read **Zurich hotel** reviews ...
Plan your trip with user reviews, travel articles, guides, maps, prices, deals, pictures, and more about **hotels in Zurich**, **Switzerland**.
travel.yahoo.com/p-hotel-191501810-**zurich_hotels**-i - 178k - 17 Nov 2006 - Cached - Similar pages

Top 10 **hotels in Zurich, Switzerland**
http://travel.yahoo.com/p-**hotel**-191501810-zurich_hotels-i Top 10 **hotels in Zurich**, **Switzerland**: Find vacation, tourism information, recommendations, **...**
travel.yahoo.com/p-travelguide-191501810-rsstype-**hotels**-view-rss-**zurich**_vacations-i - 5k - Cached - Similar pages

Sponsored Links

Zurich Hotels
Save up to 70% on Swiss **hotels** rooms. Book online now.
www.yeego.com

Zurich Switzerland Hotels
Save up to 75% on **Zurich hotels**. Pay at check-in. No booking fees.
www.priceline-europe.com

Zurich Hotels
Hyatt - Official Site. View Rates, Book Rooms & Get Online Specials.
www.Hyatt.com

Discount **Zurich Hotels**
Do not Miss This Great Offer
Save up to %60 off on **Zurich Hotels**
www.BookOnNet.com

Zurich Hotels Swiss
Free reservations, instant booking
No booking fees, with special offer
www.**hotelszurich**.it

Zurich Hotels
Up to 75 % discount on **Hotels**

Figure E.2 Textual listing of results Google (www.google.com)

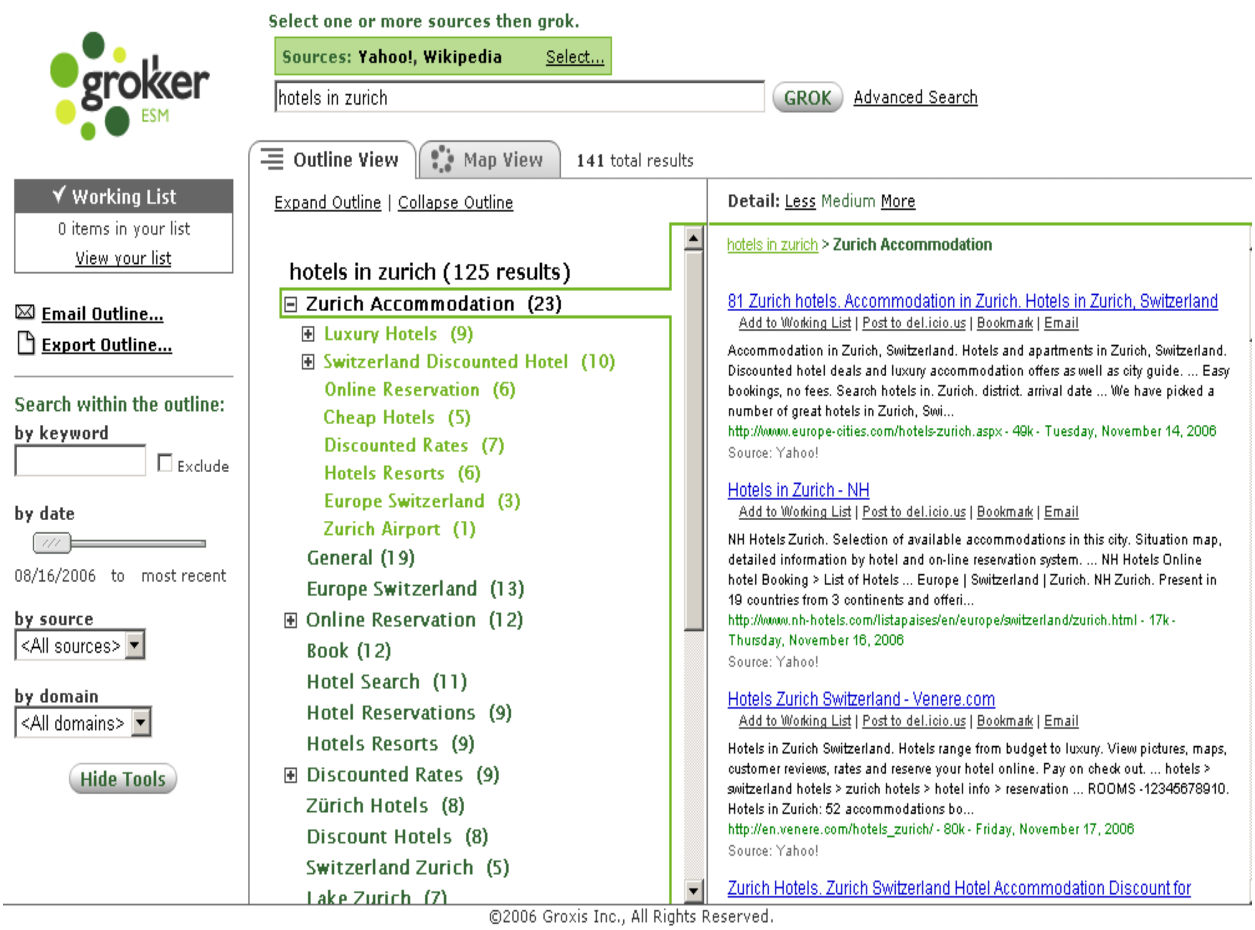
Search the Web

Select one or more sources then grok.

☑ **Yahoo!** ☑ **Wikipedia** ☐ Amazon Books

hotels in zurich

Figure E.3 Simple Text based starting point Grokker(www.grokker.com)



hotels in zurich (125 results)

- Zurich Accommodation (23)
  - Luxury Hotels (9)
  - Switzerland Discounted Hotel (10)
  - Online Reservation (6)
  - Cheap Hotels (5)
  - Discounted Rates (7)
  - Hotels Resorts (6)
  - Europe Switzerland (3)
  - Zurich Airport (1)
- General (19)
- Europe Switzerland (13)
- Online Reservation (12)
- Book (12)
- Hotel Search (11)
- Hotel Reservations (9)
- Hotels Resorts (9)
- Discounted Rates (9)
- Zürich Hotels (8)
- Discount Hotels (8)
- Switzerland Zurich (5)
- Lake Zurich (7)

hotels in zurich > **Zurich Accommodation**

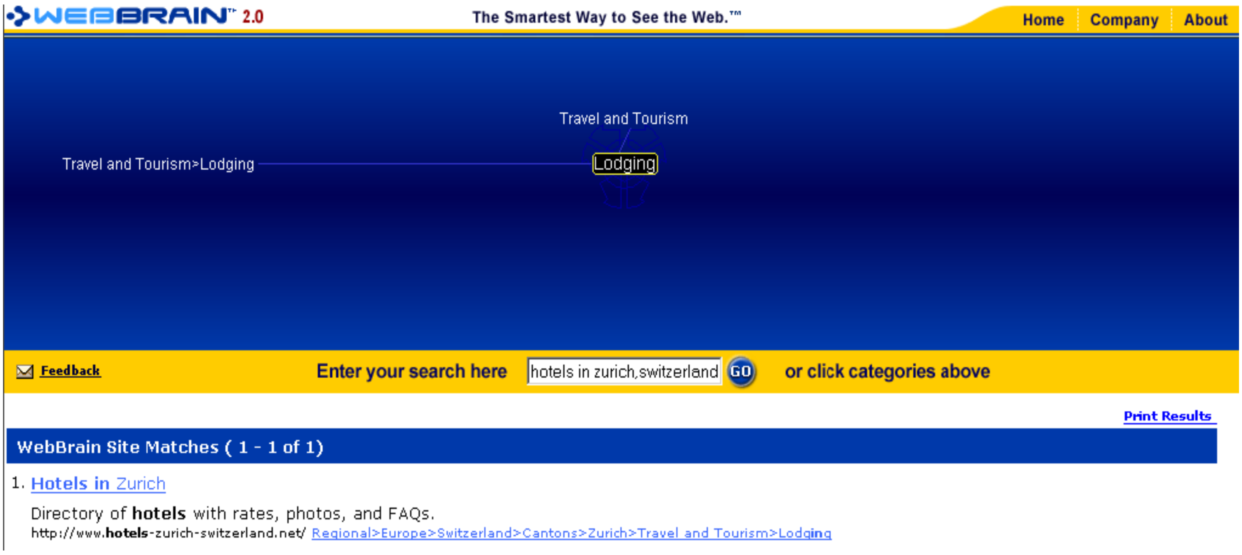
81 Zurich hotels. Accommodation in Zurich. Hotels in Zurich, Switzerland
Add to Working List | Post to del.icio.us | Bookmark | Email
Accommodation in Zurich, Switzerland. Hotels and apartments in Zurich, Switzerland. Discounted hotel deals and luxury accommodation offers as well as city guide. ... Easy bookings, no fees. Search hotels in. Zurich. district. arrival date ... We have picked a number of great hotels in Zurich, Swi...
http://www.europe-cities.com/hotels-zurich.aspx - 49k - Tuesday, November 14, 2006
Source: Yahoo!

Hotels in Zurich - NH
Add to Working List | Post to del.icio.us | Bookmark | Email
NH Hotels Zurich. Selection of available accommodations in this city. Situation map, detailed information by hotel and on-line reservation system. ... NH Hotels Online hotel Booking > List of Hotels ... Europe | Switzerland | Zurich. NH Zurich. Present in 19 countries from 3 continents and offeri...
http://www.nh-hotels.com/listapaises/en/europe/switzerland/zurich.html - 17k - Thursday, November 16, 2006
Source: Yahoo!

Hotels Zurich Switzerland - Venere.com
Add to Working List | Post to del.icio.us | Bookmark | Email
Hotels in Zurich Switzerland. Hotels range from budget to luxury. View pictures, maps, customer reviews, rates and reserve your hotel online. Pay on check out. ... hotels > switzerland hotels > zurich hotels > hotel info > reservation ... ROOMS -12345678910. Hotels in Zurich: 52 accommodations bo...
http://en.venere.com/hotels_zurich/ - 80k - Friday, November 17, 2006
Source: Yahoo!

Zurich Hotels. Zurich Switzerland Hotel Accommodation Discount for

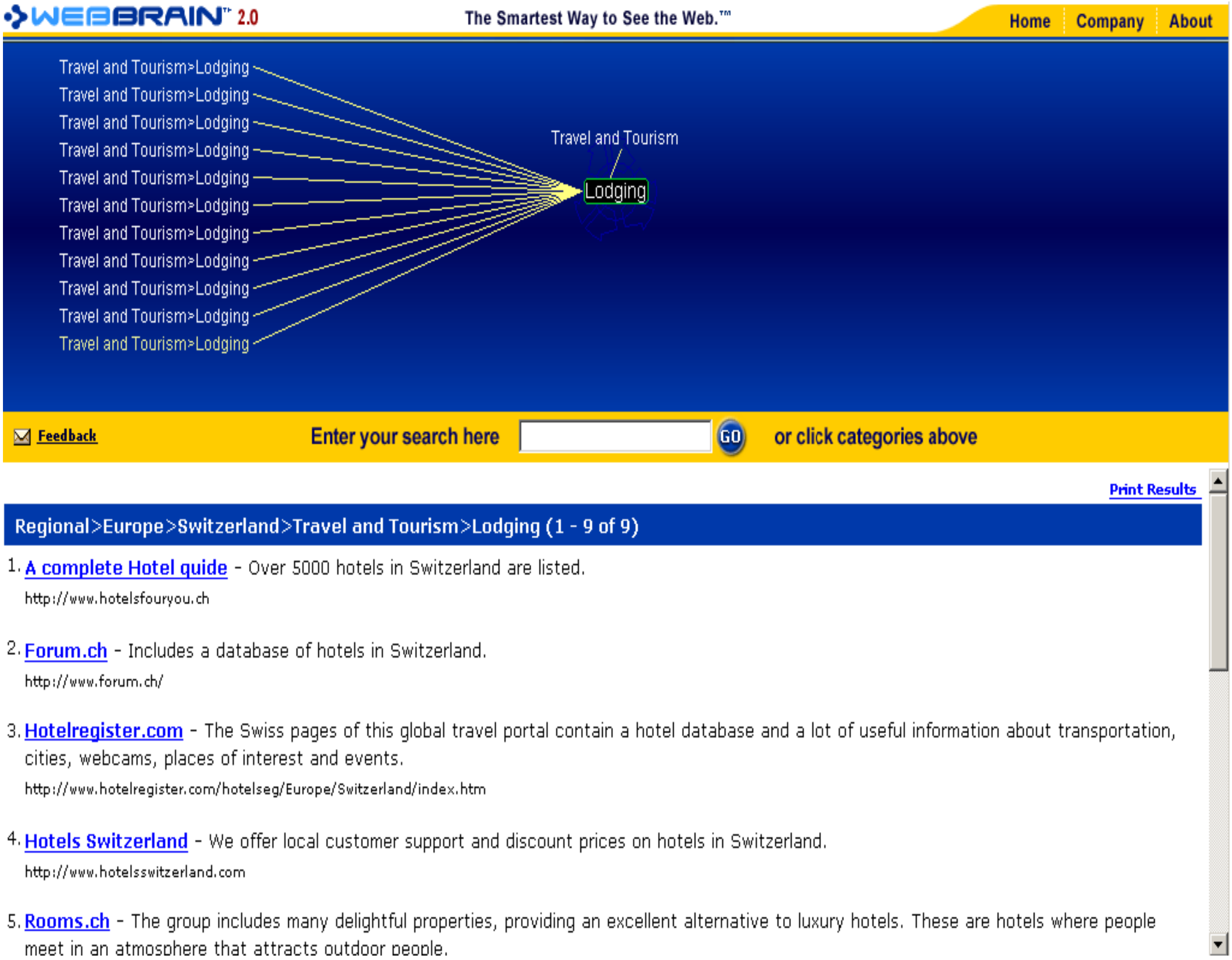
Figure E.4 Result presentation interface Grokker (www.grokker.com)



WebBrain Site Matches (1 - 1 of 1)

1. Hotels in Zurich
Directory of **hotels** with rates, photos, and FAQs.
http://www.**hotels**-zurich-switzerland.net/ Regional>Europe>Switzerland>Cantons>Zurich>Travel and Tourism>Lodging

Figure E.5 Simple text based interface as starting point: Webbrain (www.webbrain.com) Touch-graph visualization



Regional>Europe>Switzerland>Travel and Tourism>Lodging (1 - 9 of 9)

1. A complete Hotel guide - Over 5000 hotels in Switzerland are listed.
http://www.hotelsfouryou.ch

2. Forum.ch - Includes a database of hotels in Switzerland.
http://www.forum.ch/

3. Hotelregister.com - The Swiss pages of this global travel portal contain a hotel database and a lot of useful information about transportation, cities, webcams, places of interest and events.
http://www.hotelregister.com/hotelseg/Europe/Switzerland/index.htm

4. Hotels Switzerland - We offer local customer support and discount prices on hotels in Switzerland.
http://www.hotelsswitzerland.com

5. Rooms.ch - The group includes many delightful properties, providing an excellent alternative to luxury hotels. These are hotels where people meet in an atmosphere that attracts outdoor people.

Figure E.6 Result presentation Webbrain (www.webbrain.com)

KartOO is a metasearch engine with visual display interfaces. When you click on OK, KartOO launches the query to a set of search engines, gathers the results, compiles them and represents them in a series of interactive maps through a proprietary algorithm
>>More information

Search: hotels in Zurich, Switzerland

● english pages ○ world wide web

OK
Flash Player detected

KartOO uses FlashPlayer to draw interactive maps. This extension is already installed on 90% of the browsers. If you cannot see the maps, please click here to install FlashPlayer.
You can also use the HTML version with results in the form of classic lists.

Legal Note - Bookmark - Other products

KartOO.com - KartOO FR - KartOO UK - KartOO ES - KartOO IT - KartOO DE - KartOO PT - KartOO BR

(c) KartOO

Figure E.7 Simple text based interface as starting point: Kartoo (www.kartoo.com)

hotels in Zurich, Switzerland

Geneva Zurich Lucern Hotels, Switzerland Budget Hotel Discount up to ... Switzerland hotels Network - List of Discounts Hotels in Switzerland , with map, hotel information ... Welcome to Switzerland-Hotels.net - "Your Worldwide Travel Partner" , the ...

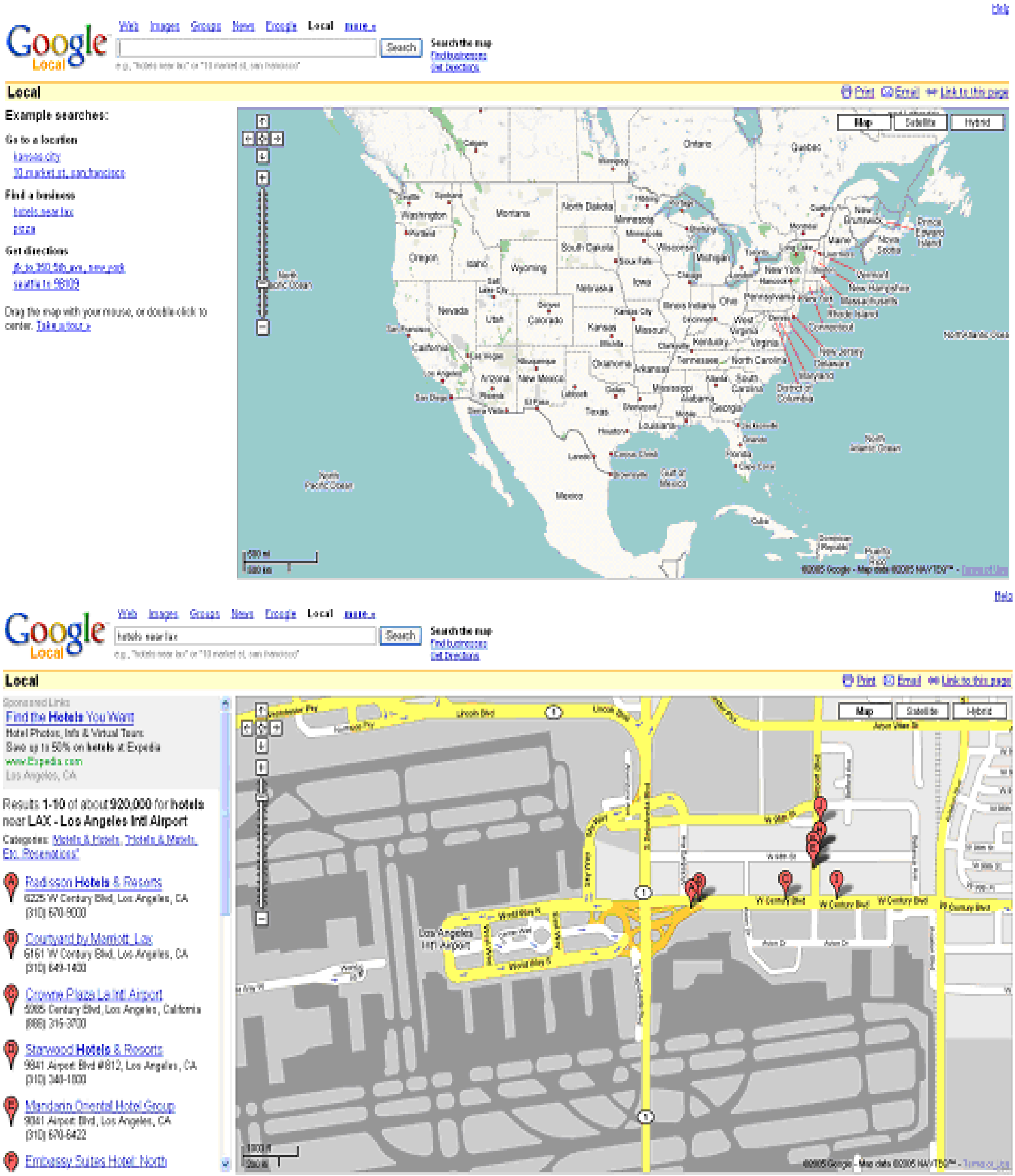
http://www.switzerland-hotels.net/

> MSN

218 000 Found results

1 - 19

Figure E.8 Interactive cartographic result presentation : Kartoo (www.kartoo.com)



e.g., "hotels near lax" or "10 market st, san francisco"

Example searches:

Go to a location
kansas city
10 market st, san francisco

Find a business
hotels near lax
pizza

Get directions
jfk to 350 5th ave, new york
seattle to 98109

Drag the map with your mouse, or double-click to center. Take a tour »

hotels near lax

Sponsored Links
Find the **Hotels** You Want
Hotel Photos, Info & Virtual Tours
Save up to 50% on **hotels** at Expedia
www.Expedia.com
Los Angeles, CA

Results **1-10** of about **920,000** for **hotels** near **LAX - Los Angeles Intl Airport**

Categories: Hotels & Motels, "Hotels & Motels, Etc. Reservations"

A Radisson **Hotels** & Resorts
6225 W Century Blvd, Los Angeles, CA
(310) 670-9000

B Courtyard by Marriott, Lax
6161 W Century Blvd, Los Angeles, CA
(310) 649-1400

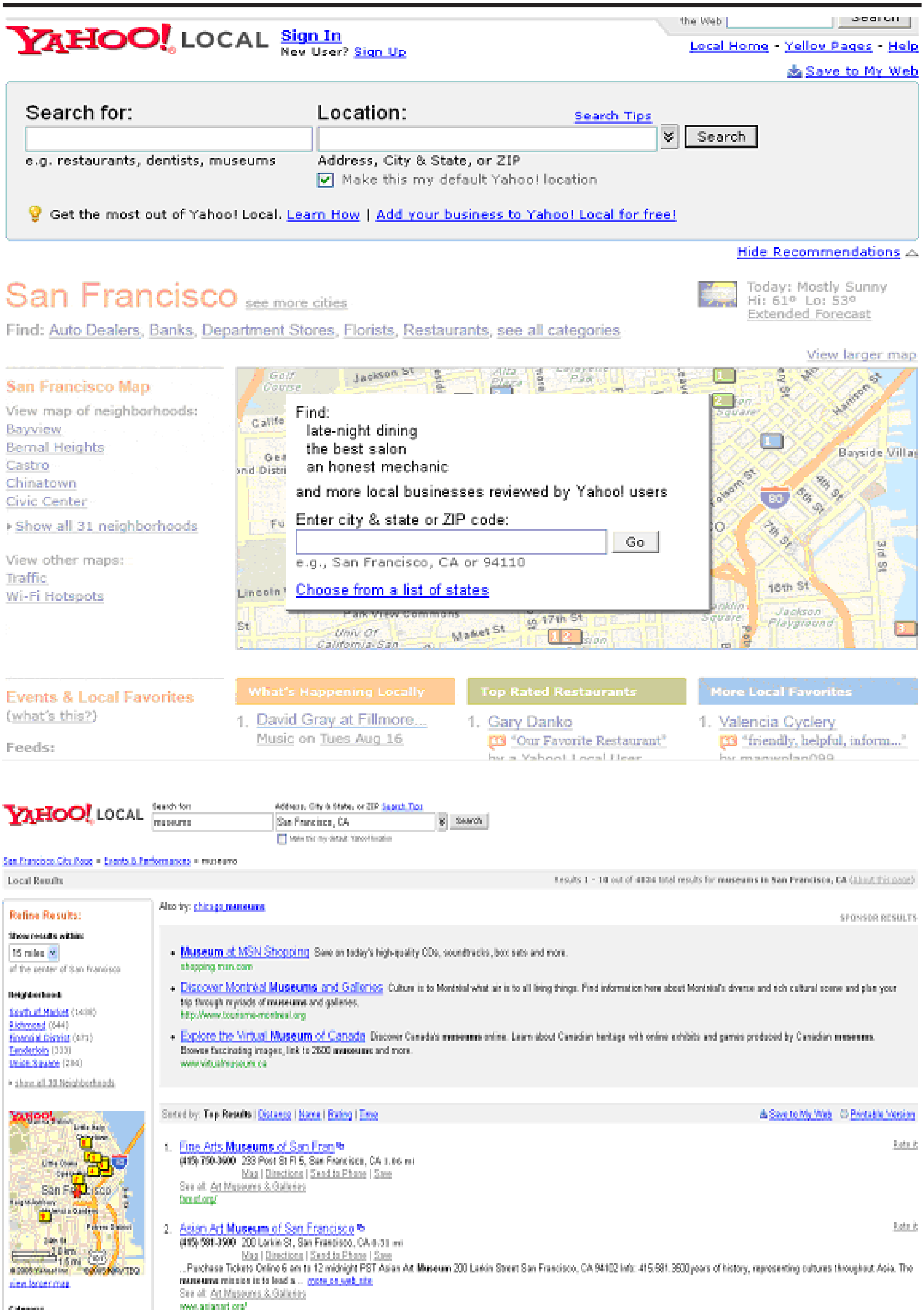
C Crowne Plaza La Intl Airport
5985 Century Blvd, Los Angeles, California
(888) 315-3700

D Starwood **Hotels** & Resorts
9841 Airport Blvd #812, Los Angeles, CA
(310) 348-1000

E Mandarin Oriental Hotel Group
9841 Airport Blvd, Los Angeles, CA
(310) 670-6422

F Embassy Suites Hotel: North

Figure E.9 A snapshot of Starting point and result presentation of spatial search engines launched late 2005: Google Local (http://www.local.google.com)



e.g. restaurants, dentists, museums

Address, City & State, or ZIP

☑ Make this my default Yahoo! location

Get the most out of Yahoo! Local. Learn How | Add your business to Yahoo! Local for free!

Today: Mostly Sunny
Hi: 61° Lo: 53°
Extended Forecast

Find: Auto Dealers, Banks, Department Stores, Florists, Restaurants, see all categories

View map of neighborhoods:
Bayview
Bernal Heights
Castro
Chinatown
Civic Center
Show all 31 neighborhoods

View other maps:
Traffic
Wi-Fi Hotspots

Find:
late-night dining
the best salon
an honest mechanic
and more local businesses reviewed by Yahoo! users

Enter city & state or ZIP code:
e.g., San Francisco, CA or 94110
Choose from a list of states

1. David Gray at Fillmore...
Music on Tues Aug 16

1. Gary Danko
"Our Favorite Restaurant"
by a Yahoo! Local User

1. Valencia Cyclery
"friendly, helpful, inform..."
by magwolan099

museums
Address, City & State, or ZIP Search Tips
San Francisco, CA
☐ Make this my default Yahoo! location

San Francisco City Page » Events & Performances » museums

Results 1 - 10 out of 4836 total results for museums in San Francisco, CA (about this page)

Show results within:
15 miles
of the center of San Francisco

Neighborhood
South of Market (1408)
Richmond (644)
Financial District (471)
Tenderloin (333)
Union Square (284)
show all 33 Neighborhoods

Also try: chicago museums

- Museum at MSN Shopping Save on today's high-quality CDs, soundtracks, box sets and more.
  shopping.msn.com
- Discover Montreal Museums and Galleries Culture is to Montreal what air is to all living things. Find information here about Montreal's diverse and rich cultural scene and plan your trip through myriads of museums and galleries.
  http://www.tourisme-montreal.org
- Explore the Virtual Museum of Canada Discover Canada's museums online. Learn about Canadian heritage with online exhibits and games produced by Canadian museums. Browse fascinating images, link to 2800 museums and more.
  www.virtualmuseum.ca

Sorted by: Top Results | Distance | Name | Rating | Time
Save to My Web   Printable Version

1. Fine Arts Museums of San Fran
(415) 750-3600 233 Post St Fl 5, San Francisco, CA 1.06 mi
Map | Directions | Send to Phone | Save
See all: Art Museums & Galleries
famsf.org/

2. Asian Art Museum of San Francisco
(415) 581-3500 200 Larkin St, San Francisco, CA 0.31 mi
Map | Directions | Send to Phone | Save
...Purchase Tickets Online6 am to 12 midnight PST Asian Art Museum 200 Larkin Street San Francisco, CA 94102 Info: 415.581.3500years of history, representing cultures throughout Asia. The museums mission is to lead a... more on web site
See all: Art Museums & Galleries
www.asianart.org/

Figure E.10 A snapshot of Starting point and result presentation of spatial search engines launched late 2005: YahooLocal

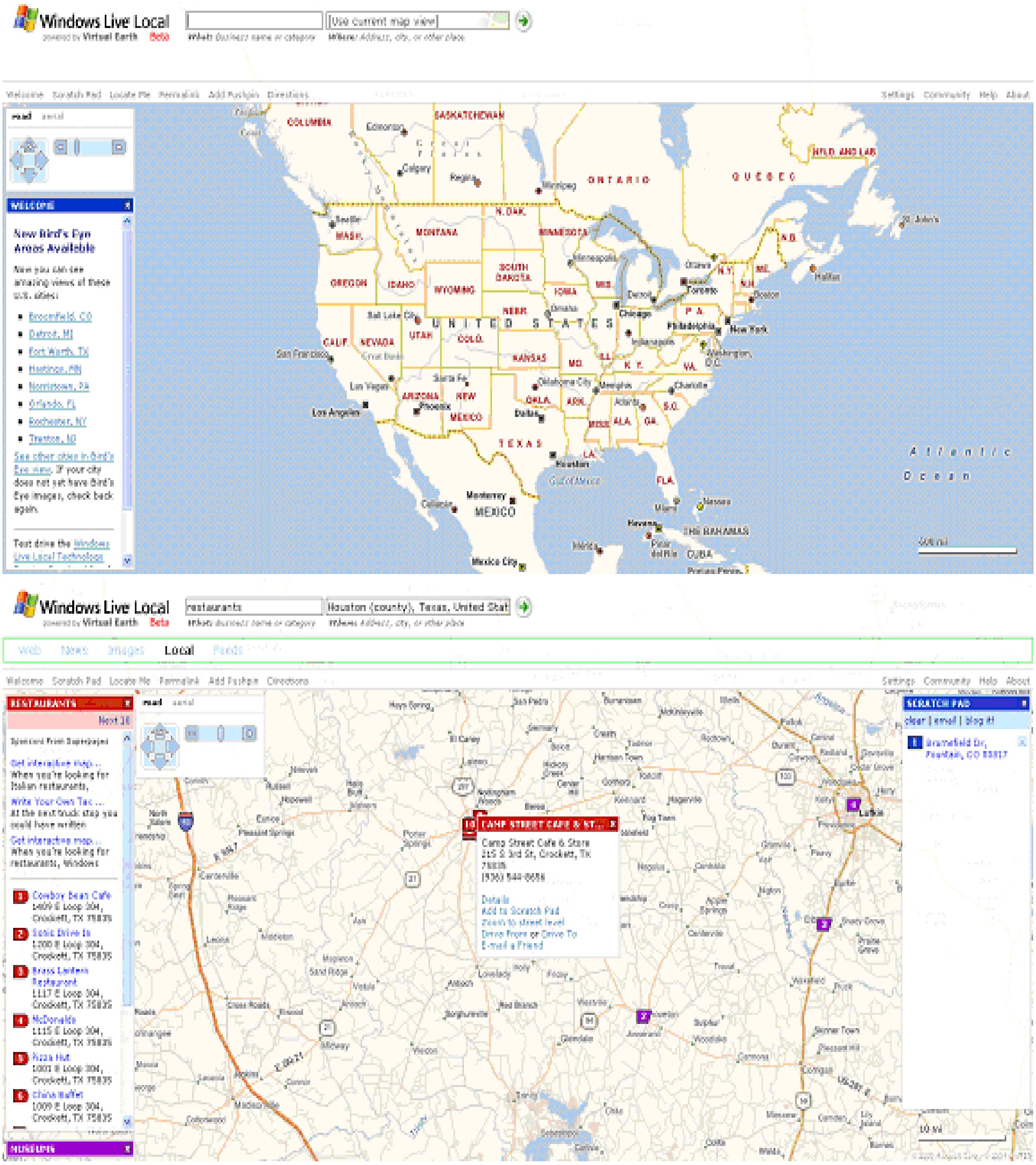


Figure E.11 A snapshot of Starting point and result presentation of spatial search engines launched late 2005: Windows Live local

Appendix F

Geographic data sources (Purves, 2003)

Country	Precision/Scale	Format
France		
BDTopo (feature based vector data set)- 8 principal themes: transport, routes, electricity network, hydrography, buildings and vegetation; description of relief, administrative boundaries, and place names. The data set is divided into two elements BDAgglo (towns) and BDPays (countryside).	~1m	EDiGeo Same as BD Carto
BDCARTO- Cartographic dataset : Themes include road networks, rail networks, place names, hydrology, adminstrative units, diverse elements Derivatives: Route120, GeoFLCP, GEOFla	~15m - source scales are 1:50,000 and 1:100,000	EDIGeo, Shapefile, MIF/MID, ARC/INFO export, Geoconcept export, DXF
GeoFLA/GeoFLCP- Adminstrative boundaries of France down to the level of communes. GeoFLCP also contains polygons describing French postal districts	As BDCarto	EDIGeo, Shapefile, MIF/MID, ARC/INFO export, Geoconcept export,
GEOROUTE- Road dataset for navigation purposes with a precision of 5-10m in urban zones and 10-20m in the rural zones.	5-10m (urban) and 10-20m (rural)	EDIGeo, Shapefile, MIF/MID, Geoconcept export.
ROUTE120-Road network data of France 120,000 km. Features include railways, towns, etc. Similar product Route500 describes 500,000 km of road network.	As BDCarto	EDIGeo, ARC/INFO MIF/MID, Shapefile, Geoconcept export
BDNYME- A total of 90 possible place types are listed in 1.7 million place names which are included in the dataset.	1:25,000	ASCII Text format, Image
Germany		

Country	Precision/Scale	Format
DBLM - Vector data set at a scale of 1:25,000	3-15m for features and the scale ranges from 1:10,000 to 1:30,000	EDBS, ARC/INFO Export, ARC/INFO Coverage, ARC/INFO Shape, AutoCAD DXF.
DLM250,000- Vector data set from 1:250,000 paper maps	1:150,000 and 1:30,000	ARC/INFO coverage, ARC/INFO Export, ARC/INFO Shape
DLM 500,000- Vector data set produced by scanning 1:500,000 rastermaps and digitzing.	1:750,000 to 1:500,000	ARC/INFO Export, ARC/INFO Generate, ARC/INFO Shape
Netherlands		
Top10L, 1:10,000 object based model for the whole of Netherlands delivered as GML.	1:10,000	GML
Top10Vector -cartographic vector dataset available as layers such as main routes, rail and air infrastructure, other routes, vegetation and landuse, hydrology and borders.	1:10,000	NEN1878, ARC/INFO, DXF, DGN
Top50 Vector	1:50,000	NEN1878, ARC/INFO, DXF, DGN
Top250Vector	1:250,000	NEN1878, ARC/INFO, DXF, DGN
Switzerland		
Vector25-9 thematic layers and has a national coverage. The layers include rail, road networks and other traffic networks. Hydrology, primary surfaces (landuse), building, hedges and trees. POI	3-8m, 1:25,000	ARC/INFO e00, ArcviewShapefile, Interlis 1, AutoCAD DXF.
Vector200- cartographic base dataset 1:200,000. It consists of 6 layers, traffic network, hydrographic network, primary surfaces, boundaries, building and single objects (POI).	1:200,000, 20-60 m	ARC/INFO e00, Arcview shapefile, AutoCAD DXF.
GG25- Adminstrative boundaries at a nominal scale of 1:25,000	1:25,000 and 3-8 m	ARC/INFO e00, Arcview Shapefile, Interlis 1, AutoCAD DXF
SwissNames- geographical names database- 190,000 Names	1:25,000 and 3-8 m	Arcview Shape, Text
United Kingdom		
Mastermap- Seamless object database	1:1,250 (urban) to ~1:2,500 (rural)	GML

Country	Precision/Scale	Format
Meridian2-midscale dataset - 2 themes -Communication and topography. Avaiable as national coverage, England, Scotland or Wales. Derived from OSCAR (a road management dataset)	Nominal 1:50,000	NTF, Shape, DXF and MIF
Strategi- small scale dataset. 7 classes of features: Communications, water features, settlements, adminstrative boundaries, woods , landuse and tourist featues.	1:250,000	NTF, Shape, DXF and MIF
Address point - grid reference of 26 million addresses in the U.K. Used in association with the post office address file	Nominal 0.01 m	CSV
CodePoint - 1.6 million U.K postcodes to grid coordinates.	1 m	ARC/INFO E00, MIF
POI-3 million points of interest with over 750 types of POI, uppermost 10 types subdivided into 56 individual categories	~1 m (positional accuracy flag provided in the data)	ASCII
Pan European datasets		
TeleAtlas MultiNet - Plethora of information, including geometry, geocoding, routing, navigation and POI	5-12 m	GDF(ASCII), Shape and Oracle Spatial
Eurostreets- GeoDan- Derived from MultiNet-Layers include Roads & Steets, incl. house number, waterareas and waterways, ferries, built up areas, nature areas, recreation areas, islands, industrial areas, adminstrative areas and POI	as MultiNet	Shape/mapinfo or any common GIS format
Major roads of Europe- extracted from MultiNet	as MultiNet	as MultiNet
SABE DATA	30 or 200m	ArcInfo Export.
Europe 4/5 digit postcode points	1:2500, 000 125 m	Shape/MapInfo
1 Digit postcode- consists of country codes	1:2500,000, 125 m	Shape/MapInfo

Appendix G

Experiment Design

Evaluation of Spatially Aware Information Retrieval System (SPIRIT)

www.geo-spirit.org

1-ABCD

Dear Assessor,

Many thanks for volunteering; your participation will help us in assessing the usability of visualizations developed for SPIRIT project (Spatially Aware Information Retrieval on the Internet). SPIRIT is a state-of-the-art spatially aware information retrieval system designed and implemented by a group of European researchers. SPIRIT's purpose is to allow users to search the internet using both geographic and conceptual terms.

Our objective is to measure the practical usability of the visualizations developed to display the results of searches.

The set of activities you do will allow us to understand better how the SPIRIT search engine is used and perceived by users. We will ask you to work through a set of scenarios describing situations where users need to search for geographical information. Online "help" is provided to assist you in understanding the functionalities of the SPIRIT system.

The SPIRIT system has restrictions to query regions because the geographic data that has been used is more detailed around Cardiff, Edinburgh and Zurich. A brief questionnaire accompanies this email, to allow us to better understand your background, exposure to search engines and information needs.

Subsequent to filling in this questionnaire you will be asked to design a scenario, where you wish to perform a search for geographical located information based on your interests and you think you could use the SPIRIT search engine to solve the problem.

After the search you will once more be asked to answer few questions based on your experience of usage of the SPIRIT system.
This is a rough timetable of your participation lasting for 1 and half hours.

- Fill in initial questionnaire (5 minutes)
- Self-learning process of SPIRIT engine - Tutorial(15 minutes)

- 2-Scenarios presented by us, for search process (30 minutes)
- User scenario formulation – 15 minutes.
- User scenario search process -15 minutes.
- Personal Interview/Feedback -15 minutes

Please feel free to ask for any information or clarification at any time during the evaluation.

We thank you very much for your co-operation. The results of this study will form an important part of my PhD, and it is also intended that they shall appear in refereed journal articles. No information allowing you to be identified will be published.

You can contact us through:
Syed Awase Khirni,25 L 94, GIS Division, Department of Geography
University of Zurich
Tel: +41-44-635 52 56
Email:sak@geo.unizh.ch

Instructions

Before examining aspects of SPIRIT search engine service, you are requested to familiarize yourself with the functionalities at your disposal. For assistance, please refer to the online help.

1. You can find the two interfaces at the following addresses:

Simple Interface
http://cheese.geo.unizh.ch:28080/tmp/visinterface/index.jsp

Account : guest
Password: guest

Advanced Interface
http://cheese.geo.unizh.ch:28080/applet/graph.jsp

2. You shall be provided with a questionnaire to gather the your background information. This task usually takes 5 minutes to complete, helping us understand your experience and background

3. A standard tutorial is provided to help you understand and learn the SPIRIT system functionalities. Please answer the questionnaire attached with the tutorial while carrying out the learning process. This task usually takes 15 minutes.

4. You are requested to carry out the tasks assigned to you, with the specified user interface, for 2 scenarios illustrated. We request that you strictly follow the order in which corresponding visualizations have to be used. A set of questionnaire are provided during the search process, for you to document the process. This task usually takes about 30 minutes.

5. For one of the scenarios ("Hotels near Portree") given to you, snap shots of the visualization are provided, to help you understand how the system works.

6. Imagine a scenario, where in SPIRIT is used to meet your needs. We recommend you to illustrate the scenario in detail, so as to help us to better understand your needs. This task roughly involves 15 minutes.

7. Considering the scenario formulated by you, please carry out the search process using SPIRIT and the interface of your choice. Please specify why a specific user interface has been chosen. This task roughly involves 15 minutes.

8. Please proceed to Personal Interview once the above jobs have been completed. This usually takes 15 minutes.

In the event of any difficulties, in understanding approach your test coordinator for more information or help.

Appendix –A

Background information about you

(5 Minutes)

1.1 Which best fits your current position?
Please tick one box.

Diploma student	
If yes, which semester:	
If yes, what is your "Hauptfach"?	
If yes, what is your "Nebenfach"?	
Academic teaching staff	
Librarian	
PhD Student	
Other (Please Specify)	

1.2 How often do you use search engines such as Google, Yahoo, etc?
Please tick one box.

Daily	
Several times a week	
Once a week	
Once or twice a month	
Never	
Please specify the search engine you normally use:	

1.3 How often do you use internet services that include mapping, e.g. Google Maps, www.map24.com, map.search.ch, etc?

Daily	
Several times a week	
Once a week	
Once or twice a month	
Never	
Please specify the mapping service you normally use:	

1.4 For what purposes do you the mapping services described above:

1.5 How familiar are you with the use of paper maps?
Please tick one box.

Very familiar	
Quite familiar	
Occasional use	
Use rarely	

Never use	
For what do you most commonly use them(e.g. ski touring, hiking, travel maps, research..)	

Simple Interface Questionnaire

(15 Minutes)

Instructions:

- **To help you understand the questionnaire better screenshots of simple interface are provided with question numbers highlighted on each of them.(Refer to the Screenshots Provided)**
- **Incase of any confusion, you are advised to refer back to these screenshots for reference or approach the test organizer for assistance.**
- **Using the "Simple interface" make a query for "hotels in Scotland" and "hotels near Glasgow", then answer the following questions.**

1. It was easy to get started with the system and make my query

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

2. It was obvious what to type in the different boxes of the simple interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If you had problems, please detail these:

____

3. The list of spatial relationships provided was sufficient:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If you would like to have more relationships available, please specify them:

____

4. Did you find the interface interesting?

Yes,very much	A little	No, not at all

5. It was easy to find the locations of documents listed to the right of the map on the map

Yes,very much	A little	No, not at all

If you found this difficult, please explain why:

____

6. Did you have any other problems using or understanding the interface?

7. The background mapping was sufficiently detailed

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. The basic map functionality (pan, Zoom, etc) was easy to use

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. What other basic map functionality might have been useful?

Group I

Scenario-I

Cottages in “Western Scotland”
Advanced Interface

You wish to go for a holiday with your family in the ‘western Scotland’. You wish to book a cottage for a week long stay. You have chosen SPIRIT to achieve your objective of finding “cottages in ’western Scotland’ ”.

Please carry out the tasks described for the illustrated scenario

Scenario-II

Hotels near Portree.
Simple Interface

You wish to go on holiday to Portree and plan to book a hotel for your family. You have chosen SPIRIT to achieve your objective of finding “hotels near Portree”

Please carry out the tasks described for the illustrated scenario

Advanced Interface

-A-

“ “

Post Retrieval Questionnaire

Simple Map based Representations

1. List the top 5 documents you found.

2. List the names and addresses of 3 places of interest found from relevant documents for your task.

3. Identify places where many relevant documents were found with respect to your query?

4. List the 5 documents, from the results list, which are most relevant with respect to your query?

5. This visualization was easy to use

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example:

____

7. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. Zooming was a useful tool in looking at the results:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. It was easy to find documents which shared the same location

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

12. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

13. This visualization made it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

14: How could this interface be improved?

-B-

Density Surface Representation

1. Identify the regions with highest document density?

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If you can identify these regions, please describe them (e.g. north of Scotland, etc.)

____

2. The visualization makes it easy to see how document density varies over space?

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

3. This visualization was easy to use

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

4. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example:

____

5. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

7. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. Zooming was a useful tool in looking at the results:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. It was easy to find documents which shared the same location

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. This visualization made it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

12. How could this interface be improved?

¬¬¬¬¬¬¬¬¬¬¬¬

-C-

Footprint based Filtering

1. It was easy to find the regions of the following documents

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1st ranked document					
2nd ranked document					
3rd ranked document					

2.Identify the top five documents with the smallest footprint (geographical extent)?

3.Identify the top five documents with largest footprint (geographical extent)?

4. List 5 documents which you perceive as most relevant with respect to your query after using footprint based filtering technique?

5. This visualization was easy to use

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example: ____

7. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. Zooming was a useful tool in looking at the results:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. It was easy to find documents which shared the same location

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

12. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

13. This visualization made it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

14. How could this interface be improved?

-D-

Cartogram

1. A single region had a significantly greater document frequency than all the others

2. It was easy to find the regions of the following documents

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1st ranked document					
2nd ranked document					
3rd ranked document					

3. This visualization was easy to use

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

4. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example:

____

5. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

7. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. Zooming was a useful tool in looking at the results:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. It was easy to find documents which shared the same location

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. This visualization made it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

12. How could this interface be improved?

Simple Interface
"Hotels near Portree"

Post Retrieval Questionnaire

Simple Map based Representations

1. List the top 5 documents you found.

2. List the names and addresses of 3 places of interest found from relevant documents for your task.

3. Identify places where many relevant documents were found with respect to your query?

4. List the 5 documents, from the results list, which are most relevant with respect to your query?

5. This visualization was easy to use

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example:

____

7. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. Zooming was a useful tool in looking at the results:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. It was easy to find documents which shared the same location

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

12. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

13. This visualization made it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

14: How could this interface be improved?

-E-

Illustrate a detailed scenario, where SPIRIT can be helpful in obtaining the required geographic information with respect to it.

Enlist the tasks involved in the illustrated scenario and carry out the search process using SPIRIT.

How you been successful in obtaining the required information for the illustrated scenario?

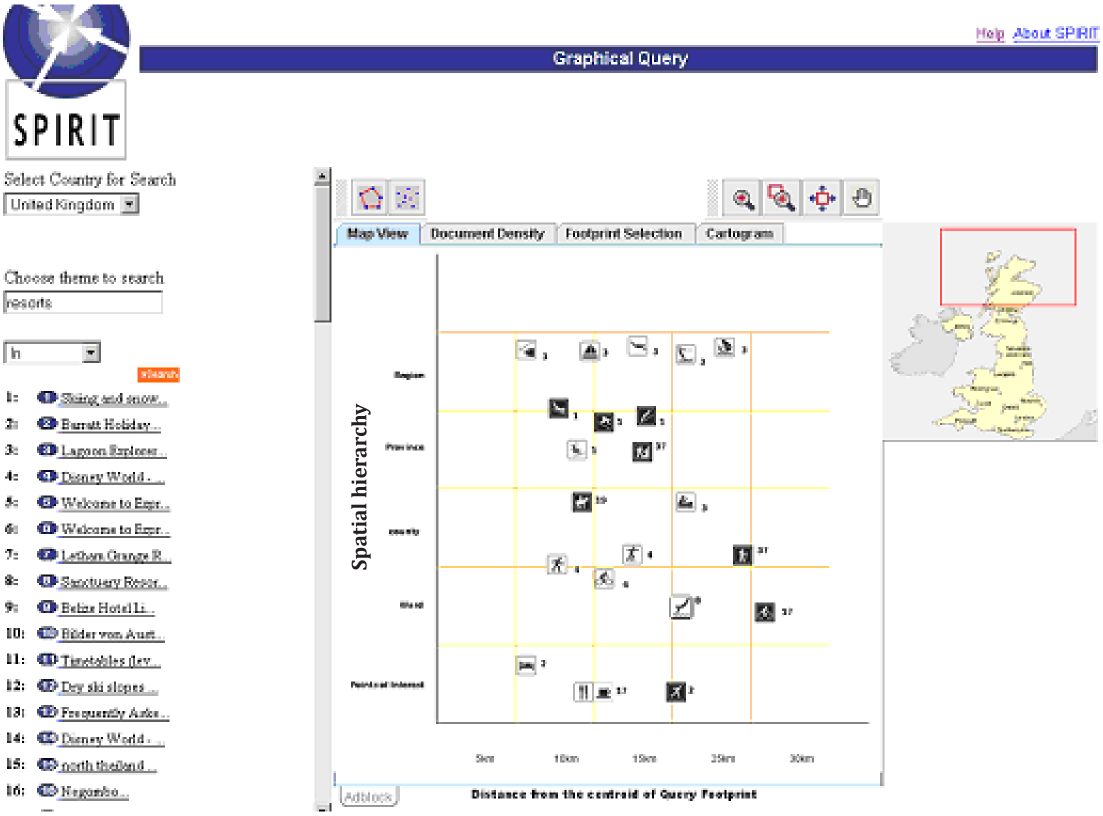
Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

-F-

Spatial Activity Plot

Instructions:

- The following visualization depicts a list of activities that could be performed at a specific region of interest specified in the information request. “Resorts in Scotland”
- A two dimensional graph plot, showing (centroids of the geographical extents) of the footprint with in individual and between documents are represented in the visualization.
- Each of these document footprints are plotted against the distance from the centroid of the query footprint expressed by the user and spatial hierarchy on X and Y-axis.
- A decision grid of 5x5 km is constructed in order to help users visualize the positioning of the document footprints with in the spatial hierarchy.



Select Country for Search
United Kingdom

Choose theme to search
resorts

In

1: Skiing and snow...
2: Barratt Holiday...
3: Lagoon Explorer...
4: Disney World - ...
5: Welcome to Expr...
6: Welcome to Expr...
7: Letham Grange R...
8: Sanctuary Resor...
9: Belize Hotel Li...
10: Bilder von Aust...
11: Timetables (lev...
12: Dry ski slopes ...
13: Frequently Aske...
14: Disney World - ...
15: north thailand ...
16: Negombo...

1. It was easy to understand the information presented in the figures above.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

2. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example:

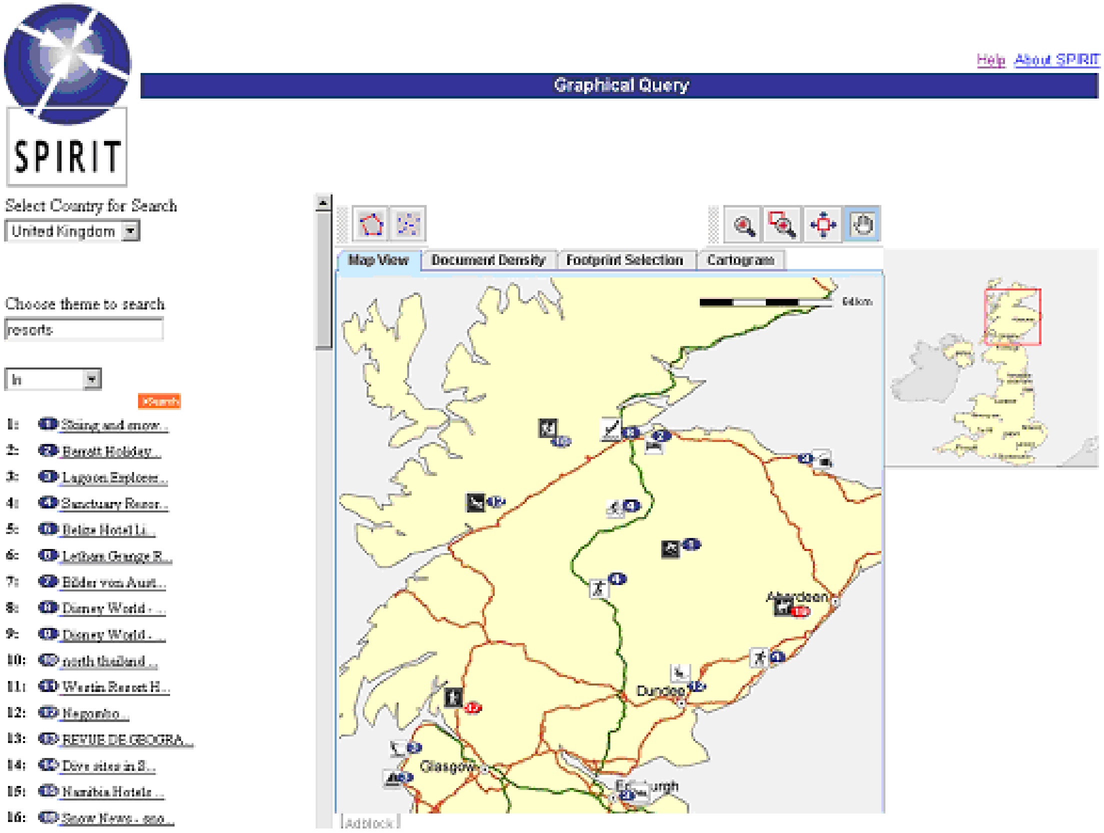
____

3. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

4. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree



Select Country for Search: United Kingdom

Choose theme to search: resorts

In

1: Skiing and snow...
2: Barratt Holiday...
3: Lagoon Explorer...
4: Sanctuary Resor...
5: Belize Hotel Li...
6: Letham Grange R...
7: Bilder von Aust...
8: Disney World - ...
9: Disney World - ...
10: north thailand ...
11: Westin Resort H...
12: Negombo...
13: REVUE DE GEOGRA...
14: Dive sites in S...
15: Namibia Hotels ...
16: Snow News - sno...

Figure Spatial activity map using real datasets with symbols depicting the contextual details of each document for a query “resorts in Scotland, United Kingdom”.

5. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. The use of symbols to illustrate activities was effective and appropriate

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

7. Is contextual classification based on activities an effective way of representing information

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. Are decision grids an effective means of identifying documents based on proximity and direction

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

10. This visualization makes it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. It is an effective way of abstraction of the results.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

11. Identify the difficulties in understanding the visualization?

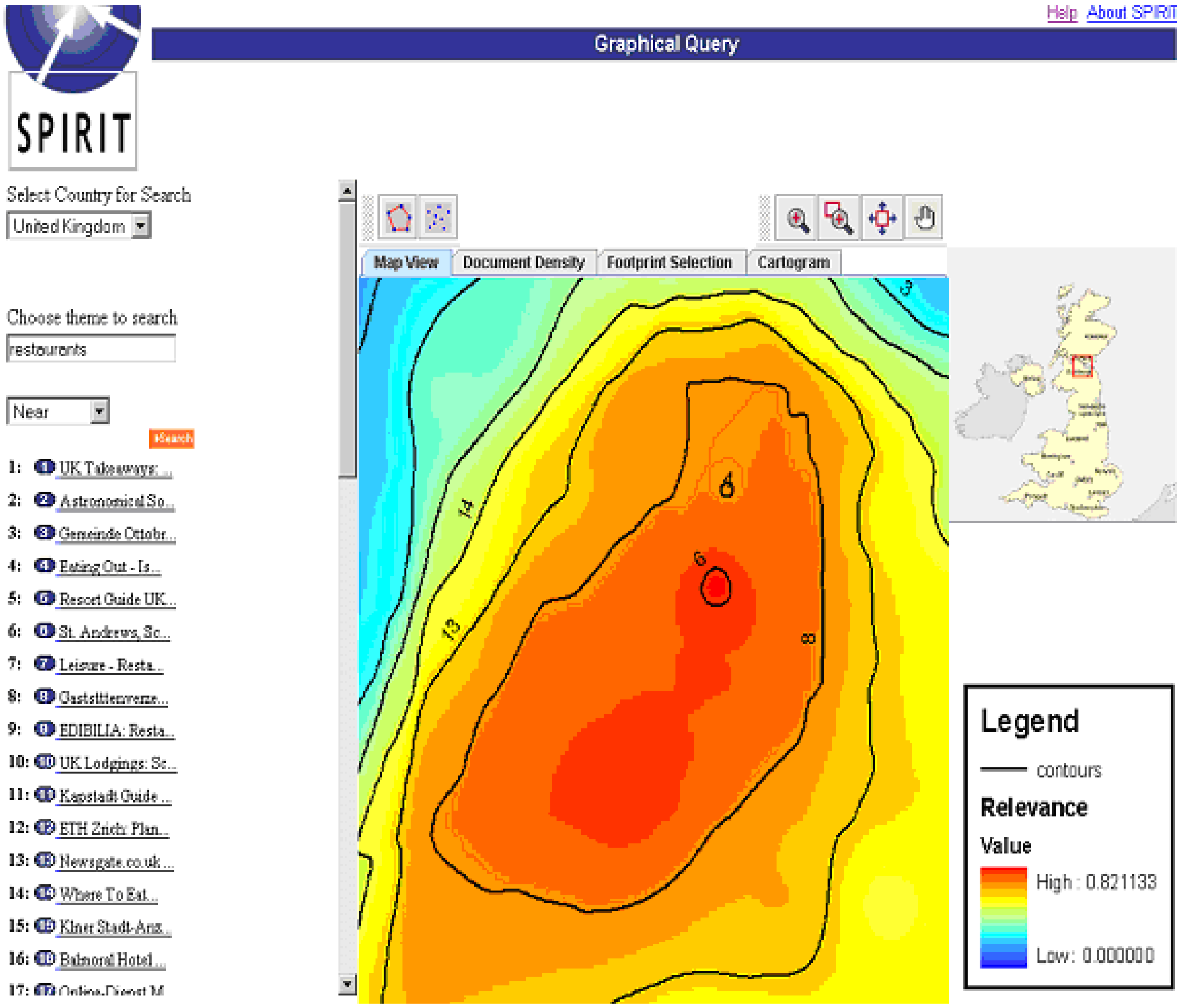
12. How could this interface be improved?

-G-

Isoline Representation

Instructions:

- The following visualization is a design concept, aimed to evaluate the effectiveness of the concept.
- The “mockup visualization” presents the results for the query “Hotels near Edinburgh Castle, Edinburgh”.
- An isoline map is constructed with continuous lines joining points of the same value (combined relevance). As combined relevance varies in space. Isoline mapping is used to interpret the information on some thematic maps.
- The peaks or regions with higher color density are regions of higher combined relevance.



Select Country for Search: United Kingdom

Choose theme to search: restaurants

Near

1: ① UK Takeaways: ...
2: ② Astronomical So...
3: ③ Gemeinde Ottobr...
4: ④ Eating Out - Is...
5: ⑤ Resort Guide UK...
6: ⑥ St. Andrews, Sc...
7: ⑦ Leisure - Resta...
8: ⑧ Gaststättenverze...
9: ⑨ EDIBILIA: Resta...
10: ⑩ UK Lodgings: Sc...
11: ⑪ Kapstadt Guide ...
12: ⑫ ETH Zrich: Plan...
13: ⑬ Newsgate.co.uk ...
14: ⑭ Where To Eat...
15: ⑮ Kiner Stadt-Anz...
16: ⑯ Balmoral Hotel ...
17: ⑰ Online-Dienst M

1. It was easy to understand the information presented in the figure above.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

2. It was easy to understand the information presented by this interface

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

If not, please give an example:

____

3. The use of legends and annotation made it easier to understand the visualization

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

4. The use of colours was useful and appropriate:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

5. This visualization provided a good way of looking at the results

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

6. It was easy to find documents which shared the same location

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

7. It was easy to get confused when using this interface:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

8. This visualization makes it easy to find the most relevant documents for my task:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree

9. Identify the difficulties in understanding the visualization?

10. How could this interface be improved?

Appendix H

Factorial Design to impart Uniform Learning effect

1	A	B	C	D
2	A	C	B	D
3	A	C	D	B
4	A	D	B	C
5	A	B	D	C
6	A	D	C	B

A- Map based interface
B- Density Surface Representation
C- Footprint based Filtering
D- Cartogram Representation

Appendix I

Table I.1 Real needs expressed by participants during evaluation of visual representations developed for SPIRIT.

S.No	User defined scenario	Geographic prepositions	Keywords used
1.	Climbing in Croatia/Scotland near the sea	near	Climbing, Croatia, Scotland
2.	Restaurants in Zürich	In	Restaurants, Zürich
3.	Surfschool in the south of England/Near Cardiff	In, South of, Near	Cardiff, England, Surf School
4.	Climbing area in the south of England (region). Switzerland (Interlaken) and France (St.Malo), South of Scotland.	In, South of	England, Climbing, Switzerland (Interlaken), France (St.Malo)
5.	Parks in Zernez, Switzerland, Swiss National Park in/near by Zernez, Switzerland	In, near by	Zernez, Switzerland, Parks, Swiss National Park
6.	Where to go in Scotland to catch Salmon? Salmon Fishing.	In	Salmon Fishing, Scotland, United Kingdom.
7.	Shopping facilities in Zurich (Clothing Stores)	In	ClothStores, Shopping malls, Zurich.
8.	Trekking in/west Scotland	In,near, west	Hiking, Trekking, Scotland
9.	Finding flats, Koblenz, Germany, Museums in Luzern	In, near	Koblenz, Germany, Flats/Apartments, Museums, Luzern
10.	River rafting in Scotland	In, near	Rafting, Scotland, United Kingdom, Rivers
11.	House boat holidays in Scotland	In	House boat, Scotland, United Kingdom
12.	Skiing near the area of chur, Switzerland, Ski Slopes in and around Zermatt or Davos, Switzerland	In, near, around	Zermatt, Davos, Chur, Switzerland.

S.No	User defined scenario	Geographic prepositions	Keywords used
13.	Holdiays in the south west of Britain, United Kingdom, Near Plymouth	South west, Near	Holidays, Plymouth, Britain, United Kingdom.
14.	Bike Tour beside the river Daneub, Germany, Austria and Romania	Beside, by the	Daneub, German, Austria, Romania, Bike tour
15.	Mountain biking near Zurich, Switzerland, Creel Fishing/ Trout fishing, Stalking in Scotland	Near, In	Mountain biking, Zurich, Switzerland, Creel fishing, Trout Fishing, Stalking, Scotland
16.	Paragliding School around Zurich, Switzerland	Around	Paragliding School, Zurich, Scotland.
17.	Free Climbing near Southampton, Sheep farming in and around Southamptoon, United Kingdom	Near, in and around	Free climbing, Sheep farming, Southampton, United Kingdom.
18.	Hotels, supermarkets, restaurants		
19.	Climbing sites near the city of Sheffied(United Kingdom).	Near	Climbing, Sheffied, United Kingdom
20.	SnowShoeing in Switzerland	In	Snow Shoeing
21.	Photocopy shops in and around Zürich, Switzerland	In and around	Photocopy Shops, Zürich, Switzerland
22.	Tourism information centers in Zürich	In	Tourism information centers
23.	Climbing in the Swiss Central Alps.	In	Climbing, Swiss Alps
24.	Car hire in London	In	Car Hire, London
25.	Wine producers in the north of Zürich	In, north of	Wine producers, Zürich
26.	Hotels campsites for holiday around Zürich.	around	Hotels, campsites for holdiay, Zürich, Switzerland.
27.	Hiking in Switzerland	In	Hiking, Switzerland.
28.	Hiking trips/huts for given destination.		
29.	Mountaineering clubs in Edinburgh, Scotland, United Kingdom.	In	Mountaineering clubs, Edinburgh, Scotland, United Kingdom.
30.	Holdiays in Italy	In	Italy
31.	Skiing places in Switzerland.	In	Skiing places, Switzerland.
32.	Hotels in Zürich, Switzerland	In	Hotels, Zurich, Switzerland.
33.	Geography department in Zürich		

S.No	User defined scenario	Geographic prepositions	Keywords used
34.	Develop an IT development index based on web site density.		
35.	Hotels on islands.		
36.	Field trips/holidays in a specific area.		
37.	Camping along the west coast of Scotland, Fish food along the west coast of Scotland, United Kingdom	Along, west coast of	Camping, Fishfood, Scotland, West coast, United Kingdom
38.	Accomodation in the Swiss Mountains	In	Accomodation, Swiss
39.	Hiking routes, B&B Location		
40.	Opera house in Southern Bavaria where verid's "traviata" will be performed between july 1st and August 1st. Find sport arenas where the "San Francisco Sharks" play in 2006 that are no father away from the Zurch than CHF 1500. Stadiums within 10kms of York, Germany	In, around, near within 10 kms of	Stadiums, Opera Houses, Southerm Bavaria, Zurich, York, Germany.
41.	B&B in a city		
42.	Hiking South of a region	South of	Hiking
43.	No use interface doesn't work.		
44.	Hotels in city or places.	In	Hotels
45.	Travel to bern, Hotels in Bern Switzerland.	In	Travel, hotels, Bern, Switzerland.
46.	Beach Holidays in Dorset	In	Beach Holidays, Dorset

Appendix J

SPIRIT Prototype Evaluation Questionnaire (Benedicte Bucher)

Background Information

1.Name :

e.g. Madonna

2.Job. Background :

e.g. Singing, Acting

3.Age :

e.g. 45

4.Nationality, Gender:

eg American woman

5.How well do you speak English ?

(please choose only one)

Native - Fluent - Basic

6.How often do you use Internet search engines (e.g. Google, AltaVista, Yahoo)?

(please choose only one)

Never - Occasionally - Once a week - Several times a week - Daily

7.How often do you use commercial online search engines (e.g. Dialog, Lexis-Nexis)?

(please choose only one)

Never - Occasionally - Once a week - Several times a week - Daily

8.How often do you perform searches on computerized library catalogues (e.g. your library)?

(please choose only one)

Never - Occasionally - Once a week - Several times a week - Daily

9.How often do you perform searches on mapping websites (e.g. multimap.co.uk, streetmap.co.uk)?

(please choose only one)

Never - Occasionally - Once a week - Several times a week - Daily

* * * * *

Express yourself

1.Please use this section to say what you think about SPIRIT and to tell us what you did in your own words.

2.What queries did you try?

3.Were the first documents retrieved relevant to your query? If not, could you find relevant documents?

4.Did anything about SPIRIT puzzle you, or make it hard to complete your search?

5.What do you like about SPIRIT ?

6. What do you dislike about SPIRIT ?

* * * * *

Give us marks

1.For each of these pairs of properties, where would you locate SPIRIT, on a scale from 1 to 5 :

A: Overall reaction to the system

Terrible Wonderful

1 3 4 5 6

Frustrating Satisfying

1 3 4 5 6

Dull Stimulating

1 3 4 5 6

Difficult Easy

1 3 4 5 6

Rigid Flexible

1 3 4 5 6

B : Specific items

Characters on the computer screen are :

hard to read easy to read

1 3 4 5 6

The organisation of information on screen is :

Confusing very clear

1 3 4 5 6

Sequence of screens is :

Confusing very clear

1 3 4 5 6

Use of terms throughout system is :

Inconsistent consistent

1 3 4 5 6

Useful messages told you what was happening :

Never always

1 3 4 5 6

Error messages were :

Unhelpful helpful

1 3 4 5 6

Help messages were :

Unhelpful helpful

1 3 4 5 6

Learning to operate the system was :

Difficult easy

1 3 4 5 6

Exploring new features by trial and error was

Difficult easy

1 3 4 5 6

Tasks can be performed in a straight-forward manner :

Never always

1 3 4 5 6

Supplemental reference materials is :

Confusing clear

1 3 4 5 6

System speed is :

too slow fast enough

1 3 4 5 6

System is :

Unreliable reliable

1 3 4 5 6

Do you want to comment on why you gave some marks?

* * * * *

Comparisons

Please, if you have some more time, try out one of your queries on some of the following Web sites :

http://local.google.co.uk

As compared to SPIRIT,..

Were the retrieved documents more relevant?

Was writing the query / refining of the query easier ?

Was results presentation better (mapping, ranking)?

http://www.yell.co.uk

As compared to SPIRIT,..

Were the retrieved documents more relevant?

Was writing the query / refining of the query easier ?

Was results presentation better (mapping, ranking)?

http://www.map24.com

Were the retrieved documents more relevant?

Was writing the query / refining of the query easier ?

Was results presentation better (mapping, ranking)?

Please, mail this form to :

benedicte.bucher@ign.fr

or post it to :

Bénédicte Bucher

IGN - COGIT

2 avenue Pasteur

94165 Saint Mandé Cedex

FRANCE

Appendix K

Additional Background Information: Users Preferences

Users background preferences with regarding to preferred search engine, preferred mapping services normally used and the purpose for which these services are normally used

No	Please specify the search engine you normally use	Please specify the mapping service you normally use	For what purpose do you use the mapping services described above	For what do you most commonly use them(e.g. ski touring, hiking,travel, maps, research)
1.	Google	map24.ch	find my way	hiking,climbing
2.	Google		private purposes	hiking & travel
3.	Google	map24.ch	find my way	hiking, climbing
4.	Google	google earth		research
5.	Google		trip planning, finding locations, hiking	research, hiking, skitouring, travel
6.	Google		find locations	hiking, travel maps
7.	Google			Skiing, hiking, sailing
8.	Google	map24.ch, google maps, google earth		general orientation, route planning, holidays, hiking, skiing
9.	Google		Route Planning	Hiking
10.	Google	google maps	to see how people have adopted it	research, travel, hiking.
11.	Google	map24.ch	putting an overview of where something is.	city map to find myself around in town.
12.	Google		wayfinding, address localization	travel, geographic related information.
13.	Google	map24.ch, stadtplandienst.de		travel hiking
14.	Google	mapquest.com, map24.com, map.search.ch, google earth	route planning, street searching, local viewings, orientation.	hiking,ski touring, travelling, study orientation.

No	Please specify the search engine you normally use	Please specify the mapping service you normally use	For what purpose do you use the mapping services described above	For what do you most commonly use them(e.g. ski touring, hiking,travel, maps, research)
15.	Google	map.search.ch, GIS-Browser: Zürich	Orienteirung	Travel maps, research
16.	Google	map24.ch	to find some places	at university, travel, for sport
17.	Google		to find a street, place look for address/ location, mountainous areas	hiking
18.	Google	map.search.ch	planning, knowledge about a geographic region	research, travel, location, maps.
19.	Google	google maps	searching locations	snowboard, touring, climbing.
20.	Google, Vivisimo.com			hiking, skiing, touring
21.	Google	GIS Browser of city of Zurich	identifying locations	skiing
22.	Google			ski-touring, hiking, travel, city maps
23.	Google	go2map.sina.com.cn, www.map24.com	travel	travel and research
24.	Google	map24.com, gis.zh.ch	looking for places (how to get there)	hiking
25.	Google	map.search.ch, swissgeo.ch	finding addresses, locations, how to get there address lcoation.	road travel
26.	Google	search.ch	location, holiday, general spatial info	hiking, military service
27.	Google	map24.ch, mapquest.com, map.search.ch	routing, tour planning, curiosity	travel
28.	Google, wikipedia	map.search.ch/ gis.zh.ch	to identify place where i have never been before	for orientation abroad hiking or visiting a city.
29.	Google, pubmed		navigation and travelling from on plae to other	Wander fahradfahre, travel

No	Please specify the search engine you normally use	Please specify the mapping service you normally use	For what purpose do you use the mapping services described above	For what do you most commonly use them(e.g. ski touring, hiking,travel, maps, research)
30.	Google		finding locations,car driving instructions	travel maps, research, car driving
31.	Google.	map.search.ch	personal information / addresses	skiing, hiking, travel maps.
32.	Google	map.search.ch, visualization of www.sbb.ch, www.zvv.ch		all kinds of outdoor activities, hobbies, hiking, jogging, cycling, rollerblading, travelling, etc.
34				hiking, travel
35.	Google	map.search.ch,www.swissgeo.ch, www.gis.zh.ch	searching addresses, localise, places, regions, get an idea of landscape.	hiking, travel maps.
36.	Google	map.search.ch	searching places, where i have to go, planning routes	hiking+ travel
37.	Google			travel
38.	Google	map.search.ch, google earth		travel, learning activities
39.	Google	gis.zh.ch, gis browser	to find out location of addresses	hiking, biking, work
40.	Google			Walk & recreation travelling and hiking.
41.		map24.ch	searching direction or addresses	travelling and hiking, research in cartography(studying)
42.	mostly google			travel maps, biking and hiking
43.	Google	map24.com	finding location and routes to the locations	travel maps, route maps
44.	Google	map24.com	travelling, find places	travelling, hiking, flying
45.	Google		searchign for address	travel maps, mountain biking.
46.	Google	map24.com, GIS system of canton of Zürich	searching for address	travel maps
47.	Google, Search.ch	map.search.ch, google maps.	negotiations	travel maps, hiking

No	**Please specify the search engine you normally use**	**Please specify the mapping service you normally use**	**For what purpose do you use the mapping services described above**	**For what do you most commonly use them(e.g. ski touring, hiking,travel, maps, research)**
48.	Google	map.search.ch	get some address location	hiking, travelling
49.	Google	Google and map24.com	route search or finding a map and other information about this place	travelling (sport)
50.	Google	http://portale.web.de/auto/routen/planner, googlemaps	routeplanning, search for addresses	biking, travel with car.