Citation: **Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies**
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006. 130 pages.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Introducing Terminology-based Ontologies

Papers and Materials presented by the authors at the
workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel)
at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006

Winfried Schmitz-Esser
University of Applied Sciences, Hamburg, Germany
http://www.schmitz-esser.de

Alexander Sigel
University of Cologne, Dept. of Information Systems & Information Management, Germany
http://www.wim.uni-koeln.de/19.0.html

14th of July, 2006

Abstract:
This work-in-progress communication contains the papers and materials presented by Winfried Schmitz-Esser and Alexander Sigel in the joint workshop (with Roberto Poli) "Introducing Terminology-based Ontologies" at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.

Keywords:
Ontologies, Thesauri, Semantic Relations, KOS, Topic Maps, ICLO

Readers are advised to check for a revised and updated version of this paper and/or its parts, since the authors are working on streamlining and integrating the presentation of ICLO (the Integrative Cross-Language Ontology), on the use of a particular upper ontology within ICLO, and on how to model ICLO with topic maps.

Table of Contents:

The workshop aimed at moving from traditional engineering of Knowledge Organization Systems (KOS) (e.g. classifications and thesauri) towards more recent developments. For this purpose, it drew from a number of fields, bringing them closer together, including:
refined engineering of semantic relations in thesauri, improved provision for natural language constructs, implications of research in categories and upper ontologies for modern KOS, and the application of semantic web technologies like the Topic Maps paradigm.

At the workshop, the following contributions were presented:

1. From traditional Knowledge Organization Systems
   (authority files, classifications, thesauri)
   towards ontologies on the web (Alexander Sigel) [Tutorial]
2. Introduction to Integrative Cross-Language Ontology (ICLO):
   Formalizing and interrelating textual knowledge to enable intelligent action
   and knowledge sharing (Winfried Schmitz-Esser)
3. First Idea Sketch on Modelling ICLO with Topic Maps
   (Alexander Sigel) [Work in progress paper]
4. Upper levels (Upper Ontologies hold it together) (Roberto Poli)

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Alexander Sigel (sigel@wim.uni-koeln.de)
University of Cologne, Department of IS and IM, Cologne, Germany

From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web

Abstract: Within the three-part workshop "Introducing Terminology-based Ontologies" by Schmitz-Esser, Poli and Sigel, this paper accompanies the first part, an overview tutorial with discussion on how to migrate from traditional Knowledge Organization Systems (KOS) towards ontologies using semantic web technologies. The focus lies on Topic Maps. More up-to-date details can be found in the slides. Participants will:

1. understand what (computer science) ontologies are
2. understand how ontologies relate to traditional Knowledge Organization Systems (KOS)
3. gain insights into the potential of ontologies for the purpose of knowledge organization, retrieval and use
4. learn about migration paths from traditional KOS to ontologies expressed with semantic web technologies (RDF, OWL, Topic Maps)
5. see practical examples of how modern KOS with rich semantic relations can be represented and queried with current semantic tools, including conversion from RDF to Topic Maps (e.g. semantic relationships will be modelled in a Topic Map, and a thesaurus will be represented in RDF with SKOS and then up-converted to Topic Maps)
6. gain an overview (not an in-depth knowledge!) about ontological tools and their functions which are useful for modern knowledge organization
7. understand the difference between upper ontologies and domain ontologies, how this relates to concept definition and category theory, and what this means in practice
8. receive an update on the state-of-the-art in knowledge organization with ontologies
9. learn about the 10 best books and articles on this exciting topic
10. be prepared to understand the second and third part of this workshop and how they interrelate.

Outline:

1. Motivation, Rationale and Problem Statement
2. Some Characteristics and Shortcomings of Traditional KOS
3. Ontologies
4. Topic Maps as a Semantic Technology
5. Leveraging Traditional KOS
   Example: SKOS (Simple Knowledge Organization System)

I am very much interested in feedback, so please drop me an email (sigel@asigel.de) if you use this tutorial. The slides are available as PDF from the author, and the author is available for presenting this tutorial elsewhere.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web

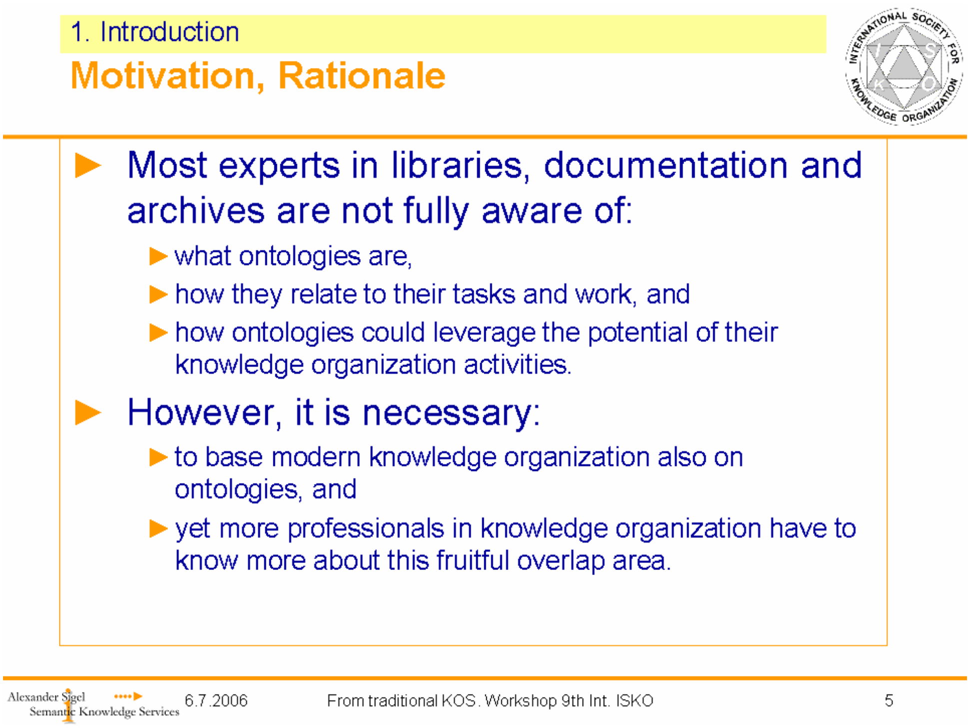
Basic Questions

- How to migrate from traditional Knowledge Organization Systems (KOS) towards ontologies using semantic web technologies?
  - focus: Topic Maps
  - Published Resource Identifiers (PRIs)
  - SKOS (Simple Knowledge Organization Systems)
- Illustration: How to implement key parts of ICLO with the Topic Maps approach as lightweight ontology?

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop “Introducing Terminology-based Ontologies” (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

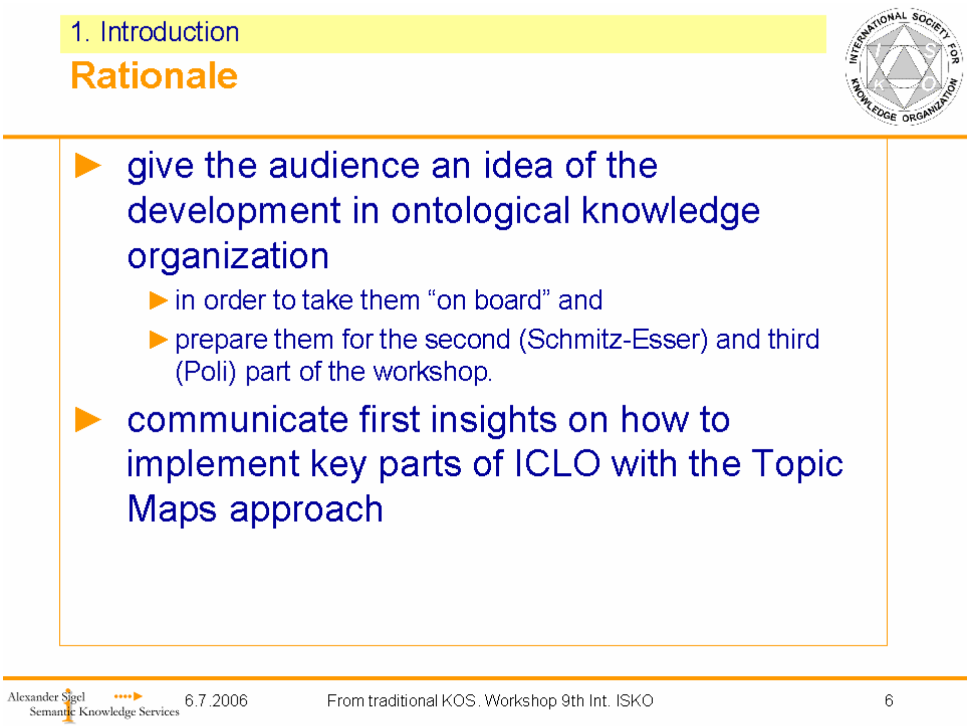
1. Motivation, Rationale, and Problem Statement

In the last years, the field of ontologies (in the sense of computer science and artificial intelligence, cf. (Gruber, 1993)) has mainly developed outside the field of information science. Judging from the contributions to the journal Knowledge Organization, the proceedings of ISKO conferences, and the entries in the bibliography Knowledge Organization Literature, there has not yet been much overlap between the communities of knowledge organization, ontology engineering, and the semantic web. Relevant work is taking place outside our community, but there is not much interaction.
Most experts in libraries, documentation and archives are not fully aware of what ontologies are, how they relate to their tasks and work, and how ontologies could leverage the potential of their knowledge organization activities. However, it is necessary to base modern knowledge organization also on ontologies, and yet more professionals in knowledge organization have to know more about this fruitful overlap area.



Therefore, the **first** rationale for this tutorial is to give the audience an idea of the development in ontological knowledge organization in order to take them “on board” and prepare them for the second (Schmitz-Esser, 2006) and third part (Poli, 2006) of the workshop.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

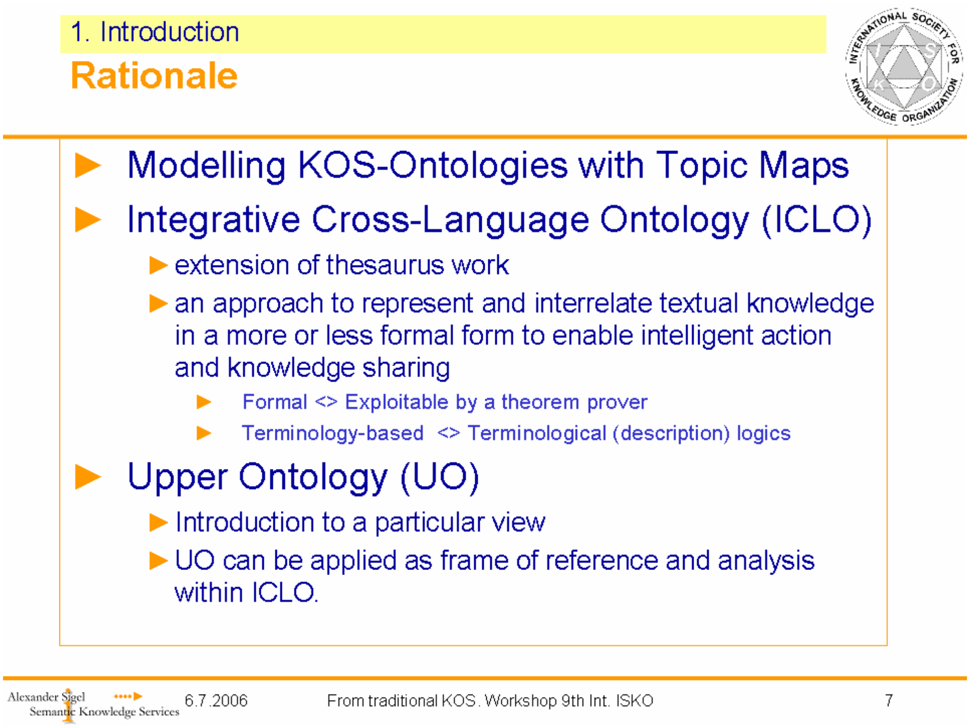


The second part will introduce the Integrative Cross-Language Ontology (ICLO). It is an approach to represent and inter-relate textual knowledge in a more or less formal form to enable intelligent action and knowledge sharing.
Formal is not used in the sense of a mathematical specification exploitable by a theorem prover. In particular, the internal semantics of many constructs are not further defined, but only identified, referenced, and described.
Terminology-based knowledge does not mean terminologic (description) logics, but refers to grounding knowledge, in particular propositions, on linguistic expressions encountered in real-world texts.
ICLO is one framework in which such leveraged KOS are crucial. The part on ICLO by Winfried Schmitz-Esser will elaborate the thinking behind the ICLO framework

The third part will introduce to a particular view on an upper ontology which can be applied as frame of reference and analysis within ICLO, beyond just using entries taken from an upper ontology for disambiguation purposes.

The **second** rationale is to communicate first insights on how to implement key parts of ICLO with the Topic Maps approach (see my separate paper: "First Idea Sketch on Modelling ICLO with Topic Maps"). This also serves as an illustration of how a semantic web technology (Topic Maps) can be used to model typical elements of a lightweight ontology. The main claim is that essential elements of the ICLO can be modelled with the Topic Maps paradigm, and that this is actually useful.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop “Introducing Terminology-based Ontologies” (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14



In sum, this first part provides an up-to-date basic understanding on possibilities for leveraging traditional KOS as lightweight ontologies to the semantic web. Selected issues and resources relevant to the move from traditional KOS towards lightweight ontologies on the web are sketched.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Problem statement

How can we leverage existing KOS as lightweight ontologies for semantic web applications?
How could new models and technical capabilities aid in the development of better KOS?

1. Introduction

Problem Statement

- How can we leverage existing KOS as lightweight ontologies for semantic web applications?
- How could new models and technical capabilities aid in the development of better KOS?

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 8

2. Some Characteristics and Shortcomings of Traditional KOS

2. Characteristics and Shortcomings of Traditional KOS

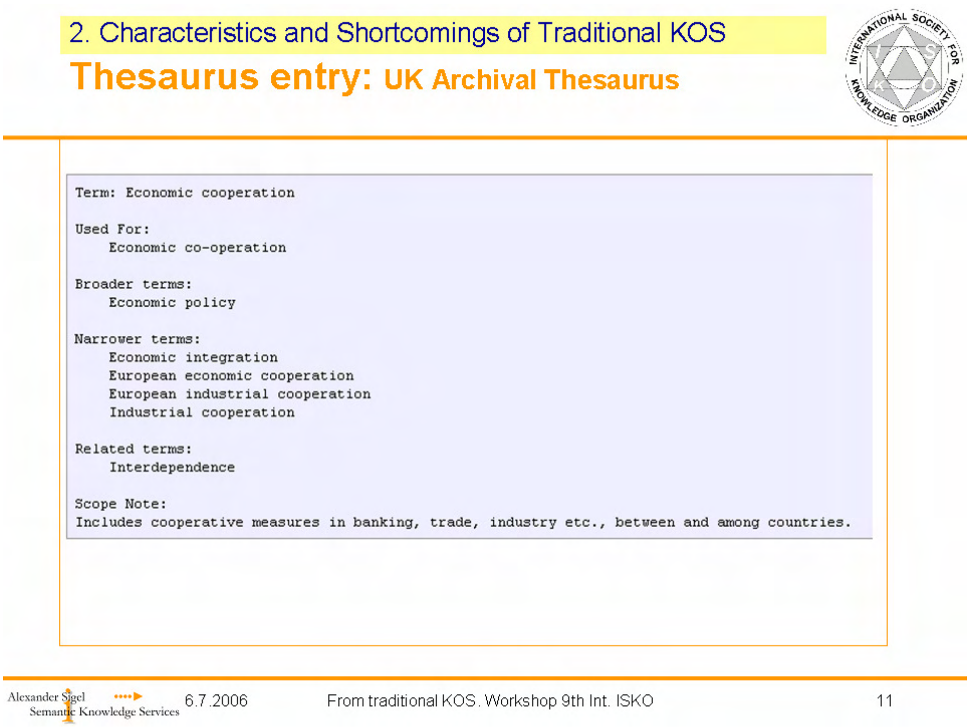
Characteristics

- Authority Files
- Glossaries
- Simple Taxonomies < Classifications
- Thesauri

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 10

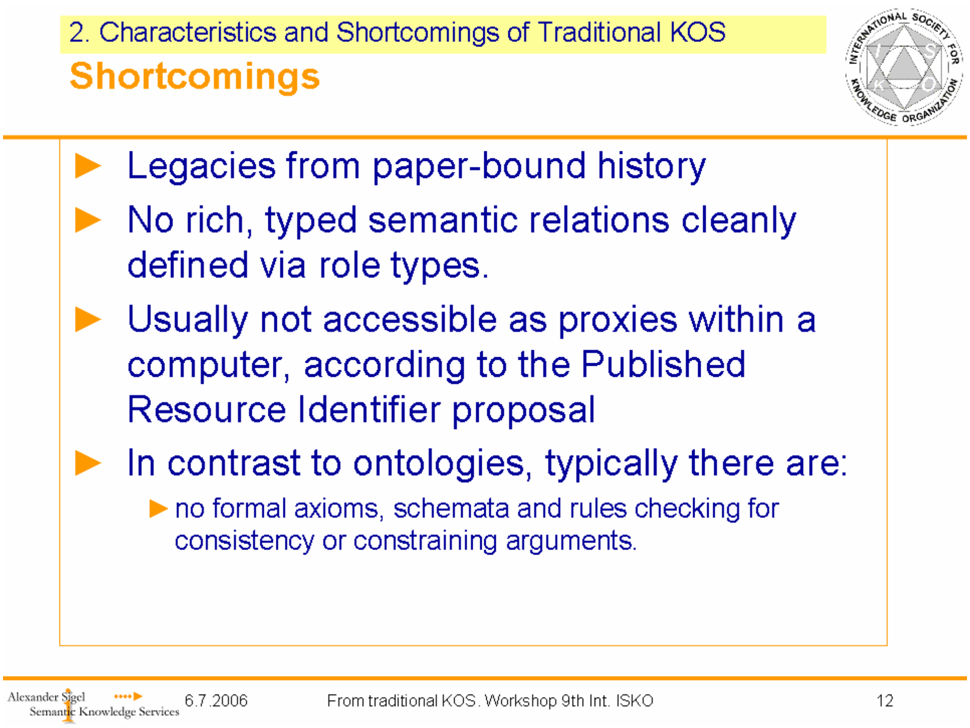
Authority files collect information useful in better identifying particular instances of common types (such as authors, corporate bodies, geographic entities, etc.), and, as a form of control,

provide a normative preferred name to be used in professional metadata creation (indexing, classifying). Typically, there are thousands of instances, but only a few types. Unlike taxonomies, almost no hierarchical relations are used. Unlike thesauri, the syndetic structure is quite poor. The main function of authority files is to provide enough identifying information such that humans can establish the identity of the entity at hand, to provide a unique identifier, and a preferred name.
Glossaries contain definitions for terms, and cross-reference entries. It would be helpful to facet classify glossary entries, but this is rarely done. Even better would be to follow the onomasiological principle (concept-to-term) (Riggs, 1996/1997), but this is almost completely neglected.
Simple classifications are *taxonomies*, i.e. (poly-)hierarchically ordered concepts with notations. More sophisticated *classifications* are not pre-coordinated, but freely facetted.
Simple *thesauri* render semantic relations between two concepts by an under-specified quasi-hierarchical broader-narrower relation, and a semantic relation of unspecified relatedness. Preferred labels (descriptors) are discerned from the access vocabulary (non-descriptors). It is possible to integrate both classifications and thesauri, in particular in faceted form, and to use them interchangeably (classifications stressing the notation, thesauri stressing the verbal labels).
Until recently, thesauri have been available in paper form, or have not fully used the capabilities of digital media. For a typical thesaurus entry not yet leveraged to semantic technologies, see the entry on "Economic cooperation" in the UK Archival thesaurus. But even if a thesaurus is already available in electronic format, as PDF, or in some XMLized thesaurus exchange format, in most cases, a user cannot further process and use it with semantic tools.



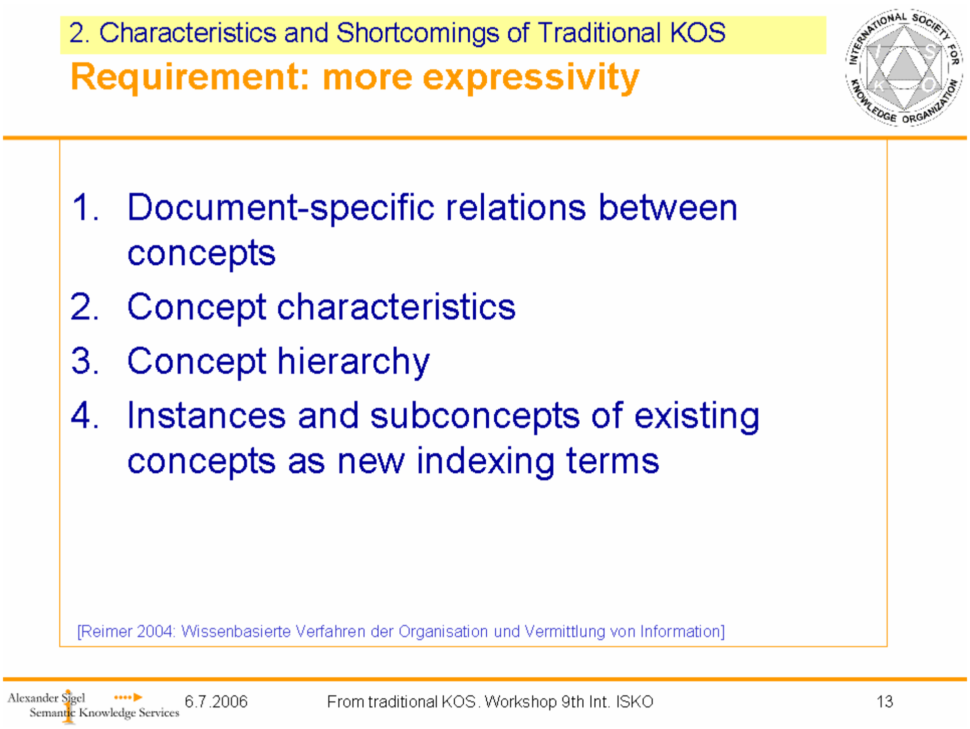
Absent from most traditional KOS are rich, typed semantic relations cleanly defined via role types. Most traditional KOS, though computer-readable, have still legacies from their history as being paper-bound. In addition, they are not yet accessible as proxies within a computer, according to the Published Resource Identifier proposal (Pepper, 2003, 2006a; Pepper &

Schwab, 2003). In contrast to ontologies, typically there are no formal axioms, schemata and rules checking for consistency or constraining arguments.

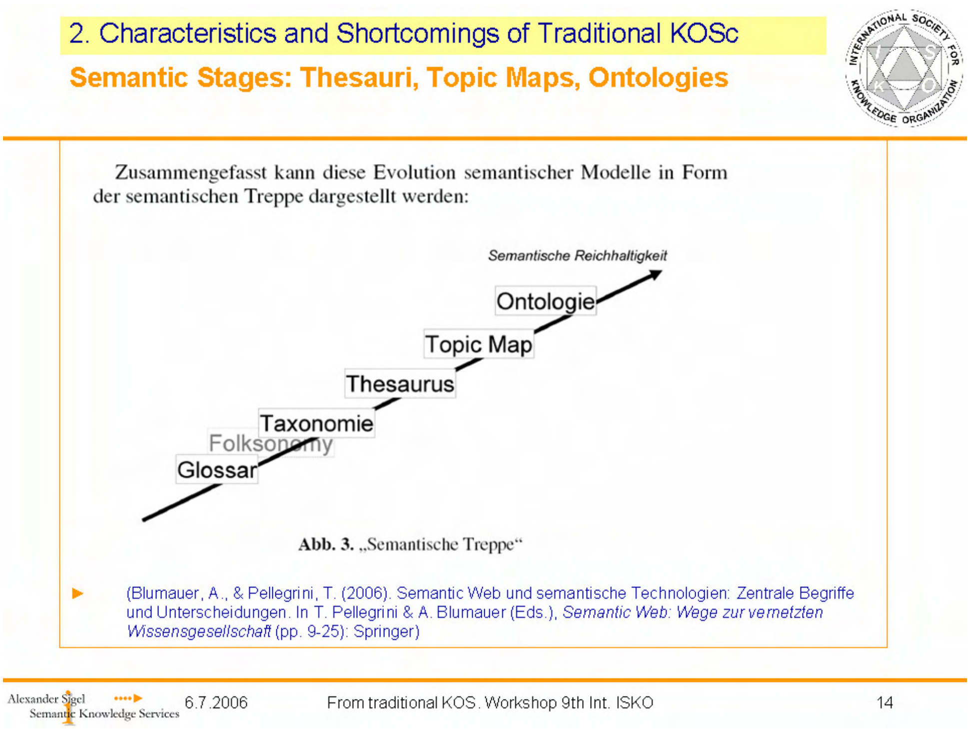


According to Reimer (2004), more expressivity is needed, in particular:

1. Document-specific relations between concepts
2. Concept characteristics
3. Concept hierarchy
4. Instances and subconcepts of existing concepts as new indexing terms



Semantic web authors have come up with a staircase-like, or staged arrangement, bringing glossaries, taxonomies, thesauri, topic maps and ontologies into a kind of partial ordering with increasing semantic richness (Blumauer & Pellegrini, 2006; Ullrich, Maier, & Angele, 2004).



The basic message is that full ontologies exhibit most semantic richness, while traditional KOS are low- to mid-level. Topic maps are somewhat higher than traditional KOS, but usally lower w.r.t. semantics than ontologies. Although this is message is fine, I want to point out that topic maps could be slided up and down the slope, since they can accommodate almost arbitrary semantics, from nearly nothing to full ontologies. This means, it is up to the knowledge engineer to model what is needed with the Topic Maps paradigm. Garshol (2004) explains from the viewpoint of Topic Maps how thesauri, Topic Maps, and ontologies relate to each other.

A second message is that traditional KOS can be leveraged to richer semantics, using topic maps, and that this is a viable and useful migration path. In sum, moving from taxonomies to thesauri adds semantic relations (the syndetic structure, moving from thesauri to topic maps adds (inter alia) typed semantic relations, attributes (inline occurrences), typed links to information resources (external occurrences), and sophisticated searching and displaying capabilities.

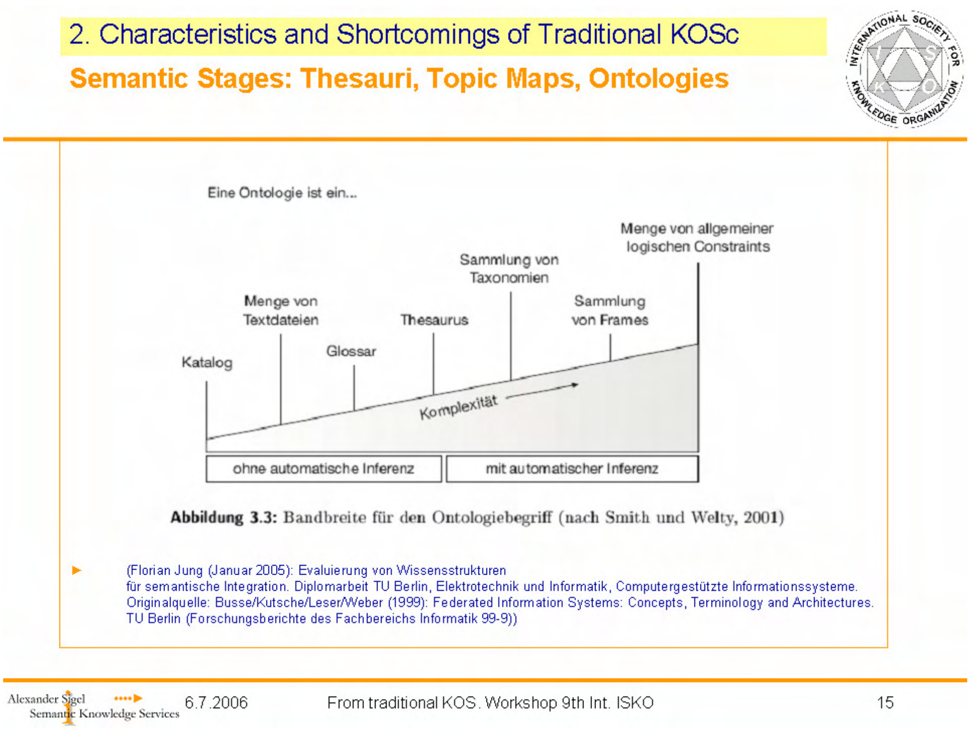
What knowledge organization background might be useful in the discussion of ontologies and semantic web technologies?

It is quite helpful, but not required, to have a solid background on foundations of knowledge organization. This includes inter alia:

- the intellectual foundation of information (Svenonius, 2000), including (FRBR, 1998)

- category theory (Barité, 2000), concept theory (Dahlberg, 1995), and indexing theory (Fugmann, 1993, 1999)
- proposals for relational systems (e.g. (Perreault, 1994; Rahmstorf, 1994), Rahmstorf with 41 binary relations, (Schmitz-Esser, 1999, 2006) with 13 relations, later extended to 23 relations, see also (Iyer, 1995)
- basics of thesaurus construction (Nielsen, 2003), domain analysis for thesaurus construction ((Hjørland, 2002), (Nielsen, 2000), (Hjørland & Albrechtsen, 1995) and ontology engineering (Endres-Niggemeyer, 2000)
- onomantics (concept-to-term) (Riggs, 1996/1997).

How can one migrate from a thesaurus to an ontology?
In a commercial whitepaper, (Ullrich, Maier, & Angele, 2004) compare taxonomies, thesauri, topic maps and ontologies.



Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

2. Characteristics and Shortcomings of Traditional KOS

1. Thesaurus

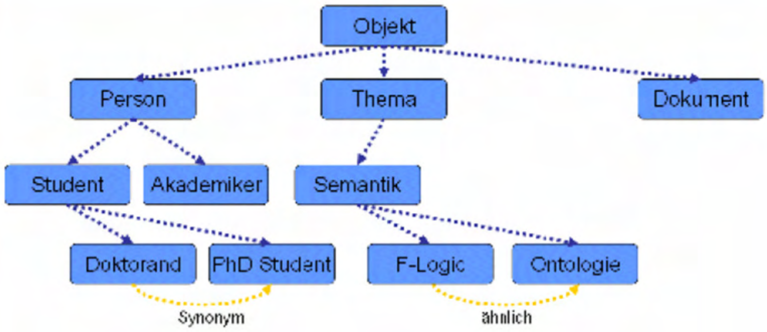


Abbildung 3: Beispiel eines Thesaurus

[Ullrich, Maier & Angele 2004]

2. Characteristics and Shortcomings of Traditional KOS

2. Topic Map

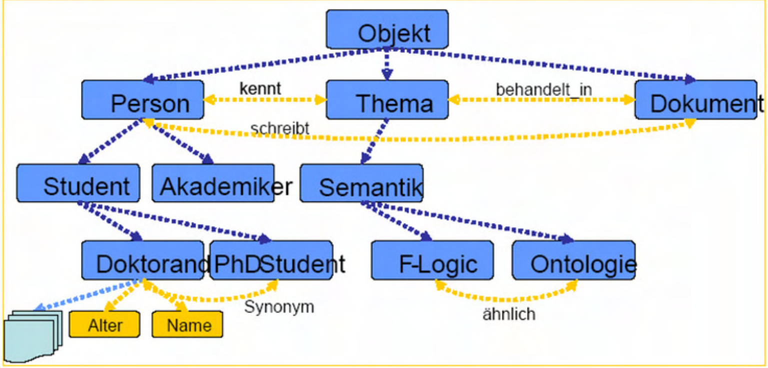


Abbildung 4: Beispiel einer Topic Map

[Ullrich, Maier & Angele 2004]

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

2. Characteristics and Shortcomings of Traditional KOS

3. Ontology

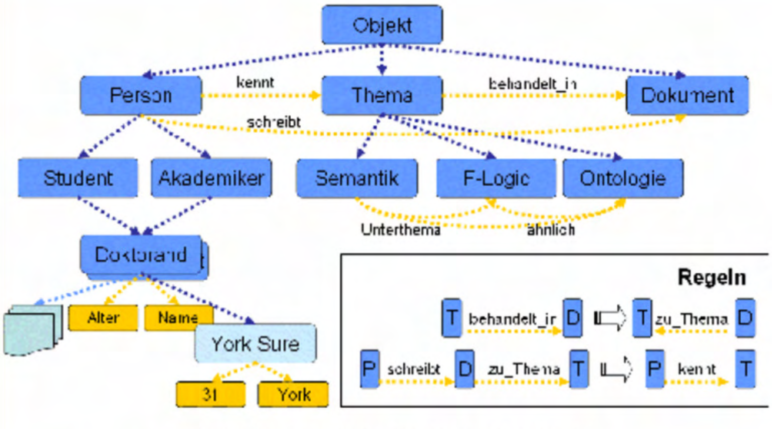
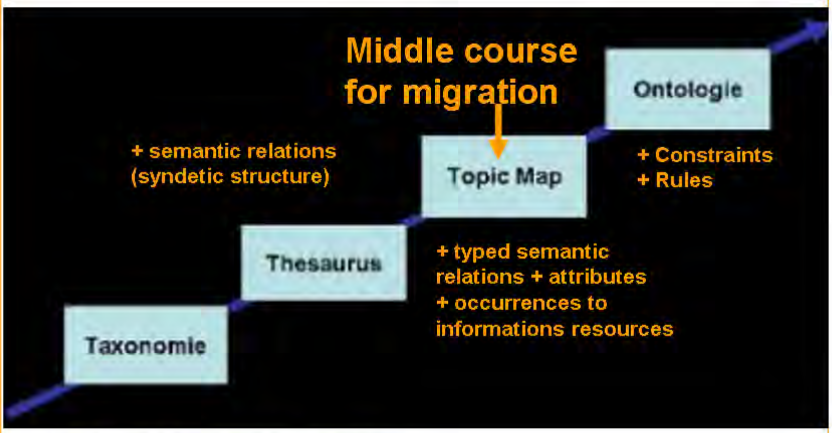


Abbildung 6: Beispiel einer Ontologie

[Ullrich, Maier & Angele 2004]

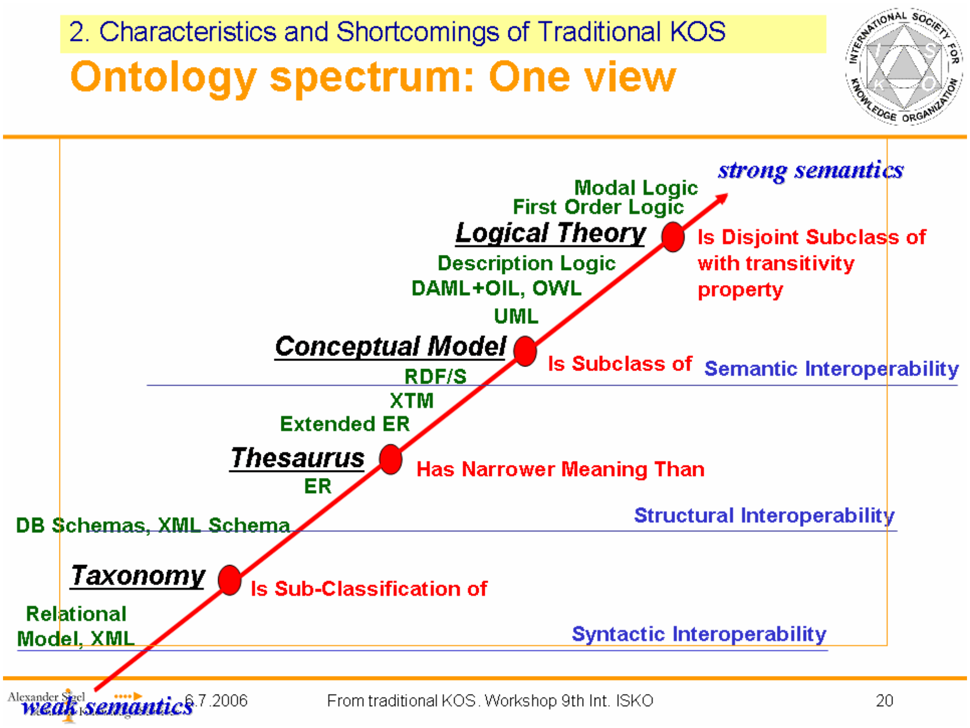
2. Characteristics and Shortcomings of Traditional KOS

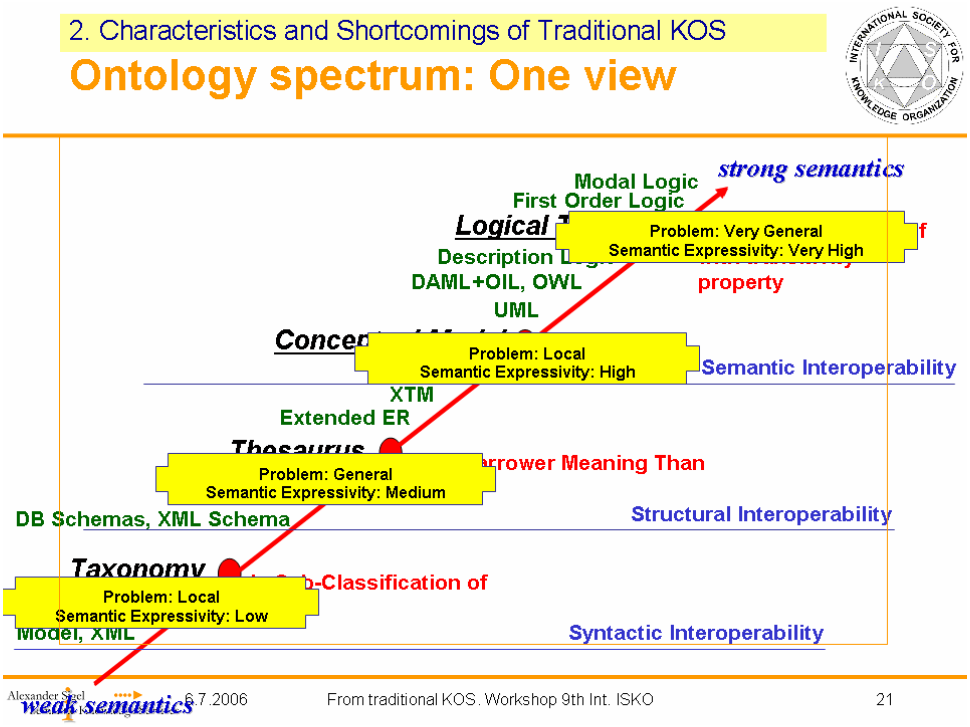
Evolution of semantic models towards higher expressivity



[Ullrich, Maier & Angele 2004] eigene Ergänzung

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop “Introducing Terminology-based Ontologies” (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14





Garshol (2004) provides a readable introduction into why Topic Maps are a more powerful solution than traditional thesauri.
(Amann, Fundulaki, & Scholl, 2000) report on an approach for building metadata schemas by integrating existing ontologies and thesauri. They use the result of this integration for RDF schema creation and metadata querying. A method for converting thesauri to RDF/OWL can be found in (van Assem, Malaisé, Miles, & Schreiber, 2006; van Assem, Menken, Schreiber, Wielemaker, & Wielinga, 2004).

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

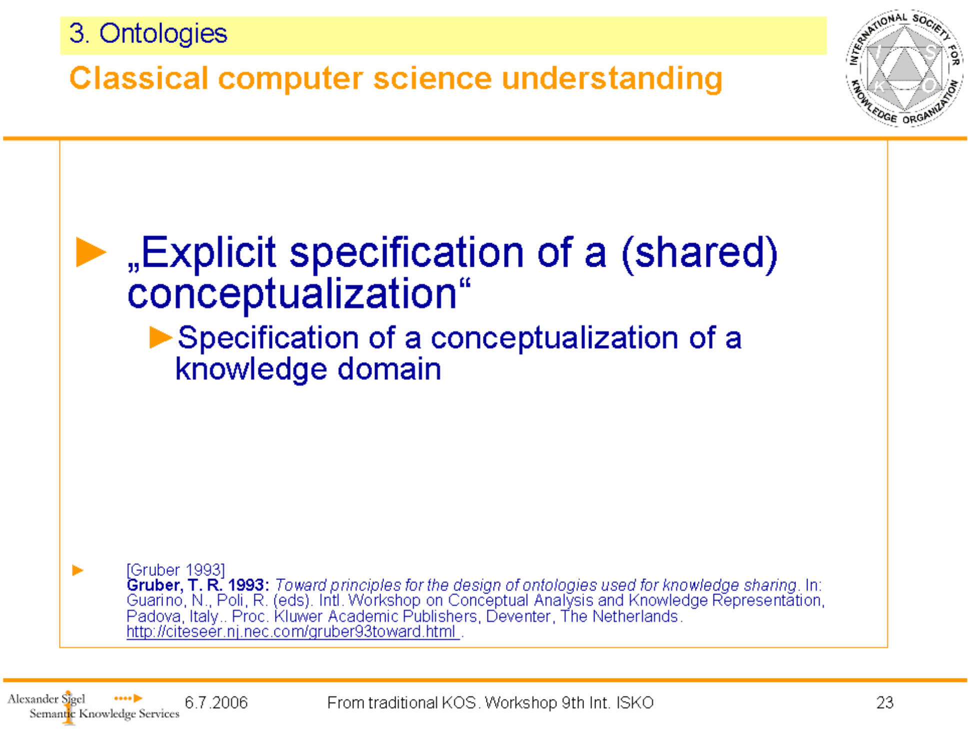
Within W3C's Semantic Web Best Practices and Deployment Working Group (SWBPD), the Porting Thesauri Task Force (PORT) is dedicated to support the deployment in RDF/OWL of thesaurus (and similar) structured vocabularies (PORT, n.d.).
One practical example of a thesaurus migrated to an ontology is the NCI thesaurus, another is the AGROVOC thesaurus:

- (Goldbeck et al., 2003) report on the NCI thesaurus migration, (Fischer, 2004) criticizes the methodology in the paper, and (Ceusters, Smith, & Goldberg, 2005) and (Kumar & Smith, 2005) find consistency problems
- (Soergel et al., 2004) report on how the AGROVOC was reengineered, (FAO, 2006) contains more information on FAO ontology resources. See also (Lauser, 2004; Lauser, Sini, Liang, Keizer, & Katz, 2006).

For practical experiences with faceted knowledge organization and ontologies, see (Tudhope, 2004) on the FACET project. (Paslaru Bontas, Tolksdorf, & Schrader, 2004) present an example of ontology-based knowledge organization in the domain of pathology.

3. Ontologies

According to the classical definition by Tom Gruber, an ontology as understood in computer science is an explicit specification of a (shared) conceptualization" (for a knowledge domain).



The concept of computer science ontologies, although it can be defined well, has been used rather vaguely, and not in a very consistent way.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

3. Ontologies

Warning!

- By some regarded as a rather vague concept, not consistenly used!

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 24

Some years ago, information science scholars have described and somehow criticized this upcoming field (Gilchrist, 2003; Soergel, 1999; Vickery, 1997), so there is no need going into more detail.

3. Ontologies

information science understanding
(critical examples)

- Gilchrist, A. (2003). Thesauri, Taxonomies and Ontologies – An Etymological Note. *Journal of Documentation, 59*(1), 7-18.
- Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science, 50*(12), 1119-1120.
- Vickery, B.C. (1997). Ontologies. *Journal of Information Science, 23*(4), 277-286.

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 25

What is an ontology (good for)?

From (Gruber, 1993) we know the now "classical" definition. (Jacob, 2003) introduces to ontologies from the viewpoint of information science. (Noy & McGuiness, 2001) give a simple and easy-to-read overview "for dummies". The same McGuiness elaborates on ontologies in (Fensel, Wahlster, Lieberman, & Hendler, 2003). In his guest editorial, (Welty, 2003) introduces into the state-of-the-art of ontology research. (Staab & Studer, 2005) is a very useful handbook on ontologies. Particularly relevant are the articles by Hahn & Schulz on building a very large ontology from medical thesauri; by Noy on tools for mapping and merging ontologies; and by Missikoff & Taglino on an ontology-based platform for semantic interoperability. For a comparative book review, see e.g. (Elzenheimer, Grollius, Heinemann, & Sternhuber, 2005). (Corazzon, 2006) provides a very thorough compilation of ontology resources from a philosophical viewpoint, but also ranging into formal ontologies. See in particular the graphic on this slide http://www.formalontology.it/table_onto_frames_file/slide0001.htm, where Sowa and Poli are related to logic in philosophy. Of course, you might also want to check the works of Poli on ontologies (Poli, 2002, 2003a, 2003b)

DELOS and NKOS

Although there has been considerable work in the context of the DELOS network of excellence in digital libraries, of NKOS (Networked Knowledge Organization Systems and Services), and of SKOS (Simple Knowledge Organisation System), its reception and adaption within the traditional knowledge organization community appears to be still low. In addition, work on Published Subjects and RDF/Topic Maps Interoperability is of high relevance to knowledge organization, but hitherto not present in the discourse.

Work comprises e.g. the workshop on "Building a meaningful web: From traditional Knowledge Organization Systems to new semantic tools" (NKOS, 2003), or the DELOS regional awareness event "Between Knowledge Organization and Semantic Web: Semantic Approaches in Digital Libraries" at Lund University in 2004 (DELOS, 2004), including the relevant presentations by (Lauser, 2004), (Miles, 2004) and (Tudhope, 2004). This resulted in a JoDI special issue on "new applications of Knowledge Organization Systems" (Tudhope & Koch, 2004), (Soergel et al., 2004), and will result in a NRHM special issue on "Knowledge Organization Systems and Services" (Tudhope & Nielsen, 2005). We owe the DELOS network of excellence in digital libraries also work on semantic interoperability in digital library systems (Patel, Koch, Doerr, & Tsinaraki, 2005). For current NKOS (Networked Knowledge Organization Systems and Services) activities, see the NKOS homepage (NKOS, 2005). A forthcoming paper on moving from a traditional thesaurus (AGROVOC) to OWL is (Lauser, Sini, Liang, Keizer, & Katz, 2006).
Further references on the relationship between KOS and ontologies can be found on slide 26.

3. Ontologies

Relationship between KOS and Ontologies

- CIDOC. (2004-05-03). CIDOC CRM Home page. Retrieved 2006-03-08, from http://cidoc.ics.forth.gr/
- Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine, 24*(3), 75-92.
- Fischer, D. H. (1998). *From Thesauri towards Ontologies?* Paper presented at the Structures and Relations in Knowledge Organization. 5th Int. ISKO Conference, 25-29th August 1998 Lille, France.
- Park, J., & Hunting, S. (2002). *XML Topic Maps: Creating and Using Topic Maps for the Web*. Boston: Addison Wesley.
- Poli, R. (1996). *Ontology for Knowledge Organization*. Paper presented at the Knowledge Organization and Change, 4th Int. ISKO Conference, 15-18th July 1996 Washington.
- Reimer, U. (2004). Wissensbasierte Verfahren der Organisation und Vermittlung von Information. In R. Kuhlen, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation* (5 ed., Vol. 1, pp. B 3, 155-166). München: Saur.
- Veltman, K. H. (2004). Challenges for the Semantic Web and Information Systems from Culture. *AIS SIGSEMIS Bulletin, 1*(1), 37-40.
- Veltman, K. H. (2004). Towards a Semantic Web for Culture, *Journal of Digital Information* (Vol. 4, pp. Article No. 255).

Frank Farance and Dan Gillman, serving on a standardization committee, in an Email 2006-04-10 to the Common Logic List, have posted the working definition of ontology of TC 37: Within the context of IT, an ontology is: „A concept system and its computational model". This means that taxonomies and KOS all could be ontologies, as long as they include their computational model.

3. Ontologies

Working definition TC 37

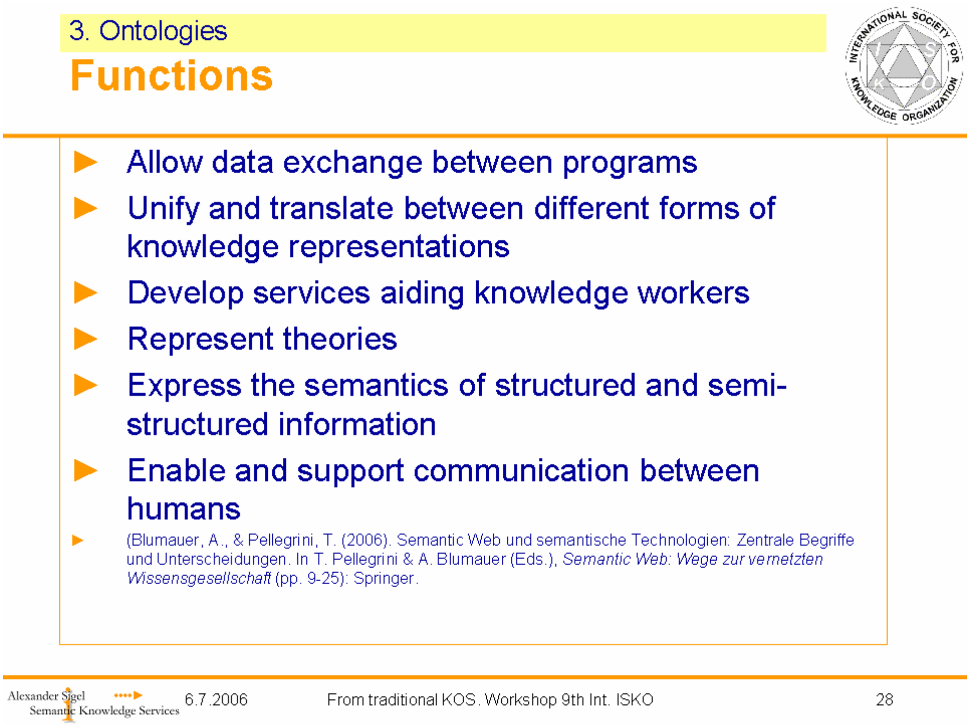
- Within the context of IT, an ontology is:
  - „A concept system and its computational model"
- Taxonomies and KOS all could be ontologies, as long as they include their computational model

- Frank Farance & Dan Gillman, Email 2006-04-10 to CL (Common Logic List)
- INCITIS/L8 (US metadata standard committee) and its corresponding ISO committee (ISO/IEC JTC1 SC 32 WG2 Metadata)

It is also useful defining ontologies according to their functions (Blumauer & Pellegrini, 2006):

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

1. Allow data exchange between programs
2. Unify and translate between different forms of knowledge representations
3. Develop services aiding knowledge workers
4. Represent theories
5. Express the semantics of structured and semi-structured information
6. Enable and support communication between humans



According to Sure, Ehrig and Studer (2006), an ontology consists of the four core elements:

1. Concepts („shared conceptualization") as representations of reality
2. Semantic Relations, linking concepts in an ontology with each other
3. Lexicon (Vocabulary), the designators (Symbols, Labels) for concepts and semantic relations, and
4. Rules (Statements in some logic about concepts and their relations)

Since we know concepts, relations and names, rules is what is new to knowledge organization.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

3. Ontologies

Four core elements

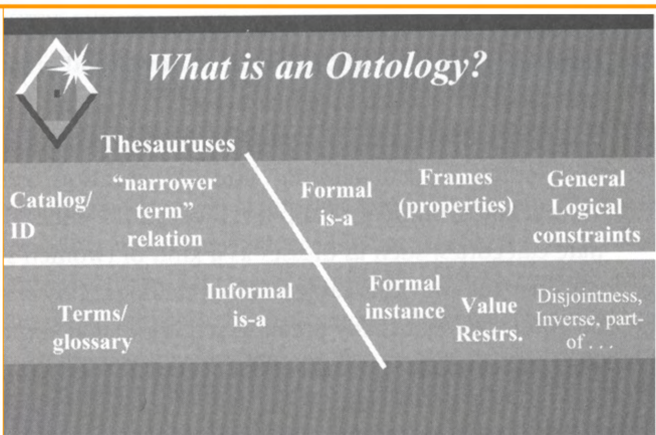
- **Concepts** („shared conceptualization") as representations of reality
  - e.g. person
- **Semantic Relations**
  - Link concepts in an ontology with each other
- **Lexicon (Vocabulary)**
  - Designators (Symbols, Labels) for concepts and semantic relations
- **Rules**
  (Statements in some logic about concepts and their relations)
  - Z.B. inverse relation, transitivity, …
- (Sure, Y., Ehrig, M., & Studer, R. (2006). *Automatische Wissensintegration mit Ontologien*. Paper presented at Modellierung 2006, Workshop Modellierung für Wissensmanagement, Innsbruck, AT, March 2006)

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 29

McGuiness (2003) (in (Fensel, Wahlster, Lieberman, & Hendler, 2003)) proposed that one can start speaking of an ontology once a formal is-a relationship is used.

3. Ontologies

Characteristics



Left of diagonal line: „Ontologie" because at least the first three necessary requirements are fulfilled (see next slide)

Right of diagonal line: additional resources for inferences

[McGuinness 2003], in [Fensel, Hendler, Liebermann & Wahlster 2003, 175]

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 30

For a simple, or lightweight ontology, necessary by definition are:

1. finite controlled (extensible) vocabulary
2. unambiguous interpretation of classes and term relationships
3. strict hierarchical subclass relationships between classes

This means that SKOS is a typical lightweight ontology.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

3. Ontologies

"Simple" ("lightweight") Ontologies

- **Necessary by definition:**
  - finite controlled (extensible) vocabulary
  - unambiguous interpretation of classes and term relationships
  - strict hierarchical subclass relationships between classes
    - SKOS is a typical lightweight ontology
- **Typical, but optional features:**
  - property specification on a per-class basis
  - individual inclusion in the ontology
  - value restriction specification on a per-class basis

[McGuinness 2003], in [Fensel, Hendler, Liebermann & Wahlster 2003]

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 31

3. Ontologies

"Simple" ("lightweight") Ontologies

- **Desirable, but neither necessary nor typical features:**
  - specification of disjoint classes
  - specification of arbitrary logical relationships between terms
  - distinguished relationships such as inverse and part-whole

[McGuinness 2003], in [Fensel, Hendler, Liebermann & Wahlster 2003]

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 32

If one abstracts from applied ontologies and goes up the ladder to the top concepts, one will reach the realm of upper ontologies, in which categories or types of most abstract nature are defined. These categories can be used to define applied ontologies. Typical examples for upper ontologies are SUO resp. SUMO (http://suo.ieee.org/; http://ontology.teknowledge.com/).

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

3. Ontologies

Upper Ontologies

- **Generic ontologies, or ontologies for describing very general knowledge,**
  - **e.g. for expressing knowledge about part-whole relationships (mereology, abstract categories/types),**
  - **on which more specific ontologies can be built**
- **Example: SU(M)O, ...**

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 33

See also the references to the work by Bateman, and Smith on slide 34:

3. Ontologies

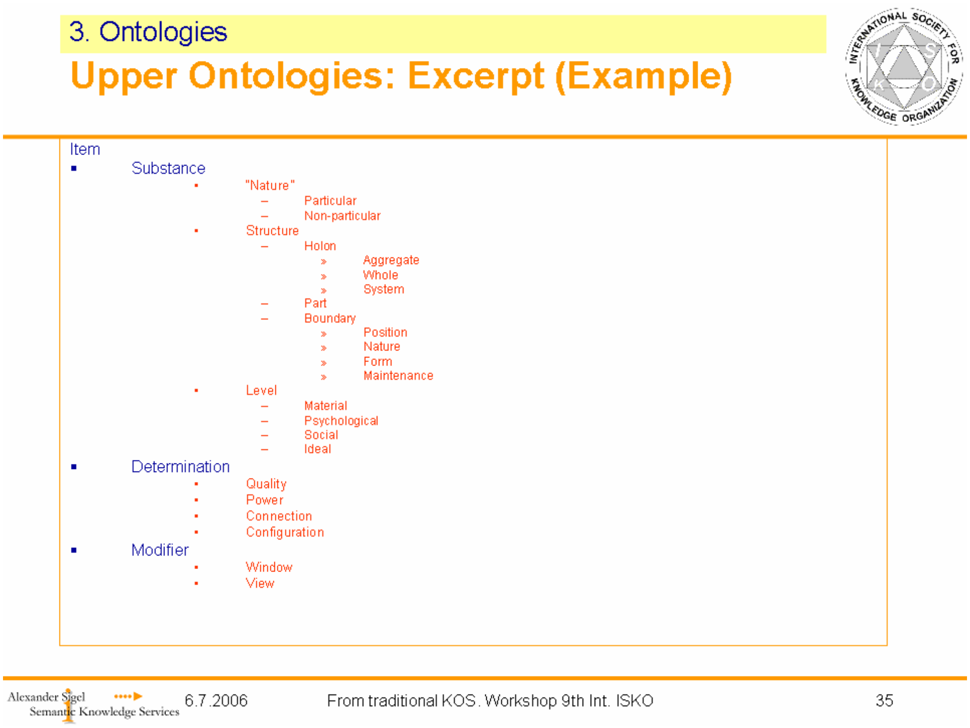
Upper Ontologies

- **Bateman, J. A. (2006):** *The Bremen Ontology Research Group.* http://www.fb10.uni-bremen.de/anglistik/ontology/ Last page update: 8 March 2006.
- **Bateman, J. A. (2006):** *The Generalized Upper Model.* http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/gum/index.htm Last Update: 17 March 2006.
- **Smith, B. (2006):** *Beginnings of a draft Upper Level Ontology.* http://ontolog.cim3.net/forum/uos-convene/2006-02/msg00059.html 26 Feb. 2006

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 34

On slide 35 an excerpt of the top levels of the upper ontology as proposed by Robert Poli is depicted.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14



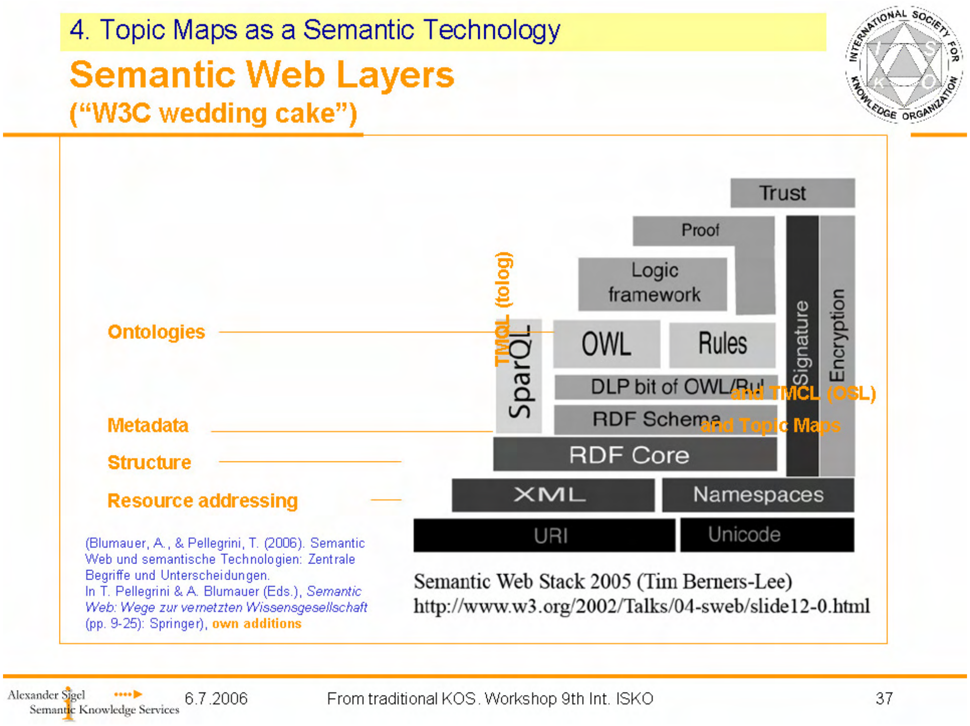
What are useful applications of ontologies?
Wherever knowledge workers share information, in particular in intranets and on the web (Stuckenschmidt & van Harmelen, 2005), like in distributed annotation (Handschuh & Staab, 2003), ontologies come into play, because there are categories, tags and KOS, and their social interrelation. Ontology-based knowledge management can be seen as a commercial application of ontology-based knowledge organization.

What applications have ontologies in knowledge management?
(Stephan Zelewski, 2002) introduces into knowledge management with ontologies. (Stephan Zelewski, Schütte, & Siedentopf, 1999) describe how to use ontologies to represent domain concepts for knowledge management. The importance of ontologies for knowledge management is stressed by (Studer, Oppermann, & Schnurr, 2001). (Davies, Fensel, & van Harmelen, 2002) elaborate on ontology-driven knowledge management. (Daconta, Obrst, & Smith, 2003) explore the future of knowledge management with semantic web technologies. Foundations and applications of ontologies for knowledge management and electronic commerce are the subject of (Fensel, 2003). (Schmaltz, 2004) critically discusses the value of semantic web technologies for knowledge management.

4. Topic Maps as a Semantic Technology

One definition of the the semantic web refers to the semantic web protocol stack (Berners-Lee 2002, 2005, (Blumauer & Pellegrini, 2006)). Since this stack usually does not include the Topic Maps paradigm, but this paradigm is part of the semantic web movement, I have taken the liberty to add where Topic Maps fit in the picture.

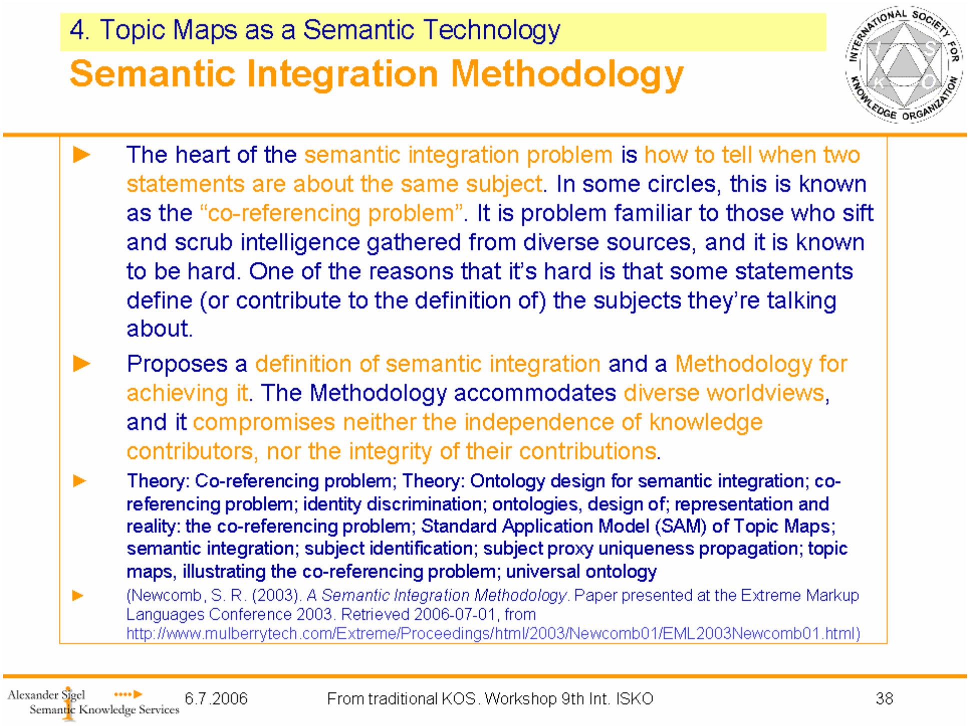
Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

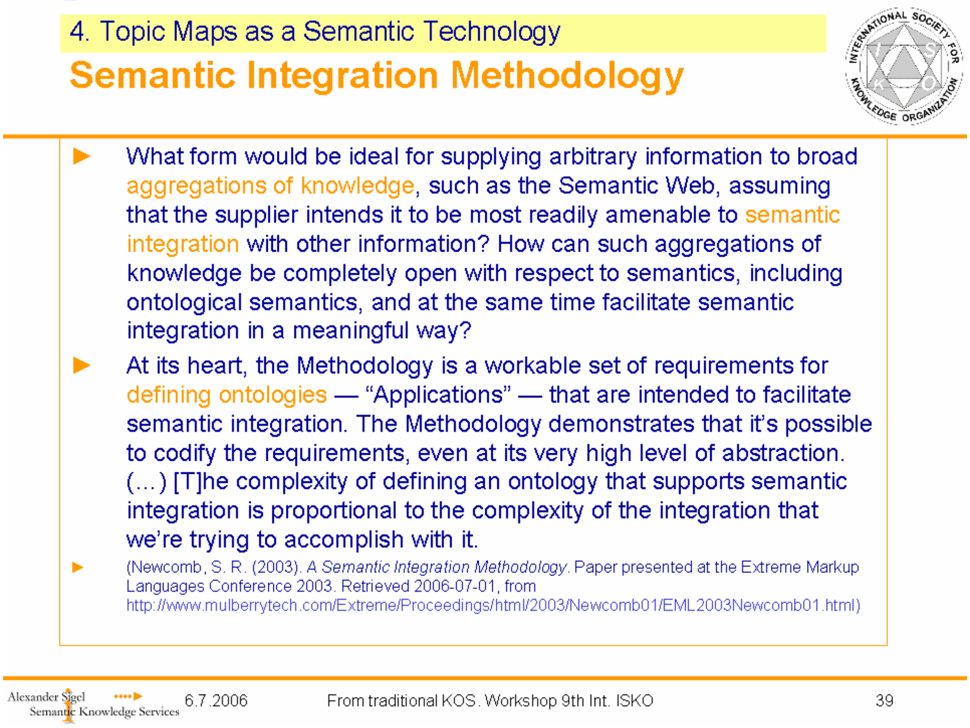


In order to address resources and identify subject proxies, we need URI and Unicode. To serialize our RDF and topic map fragments (e.g. topic maps according to XTM 2.0), we need XML as a structure. Where this "wedding cake" typically has RDF and RDF(s), one could also have topic maps and some RELAX schema, or a constraint language (TMCL). Instead of Sparql as a query language, TMQL (or e.g. tolog) could be used on the Topic Map side. Ontologies start, where rules come into play, e.g. consistency rules written in tolog.

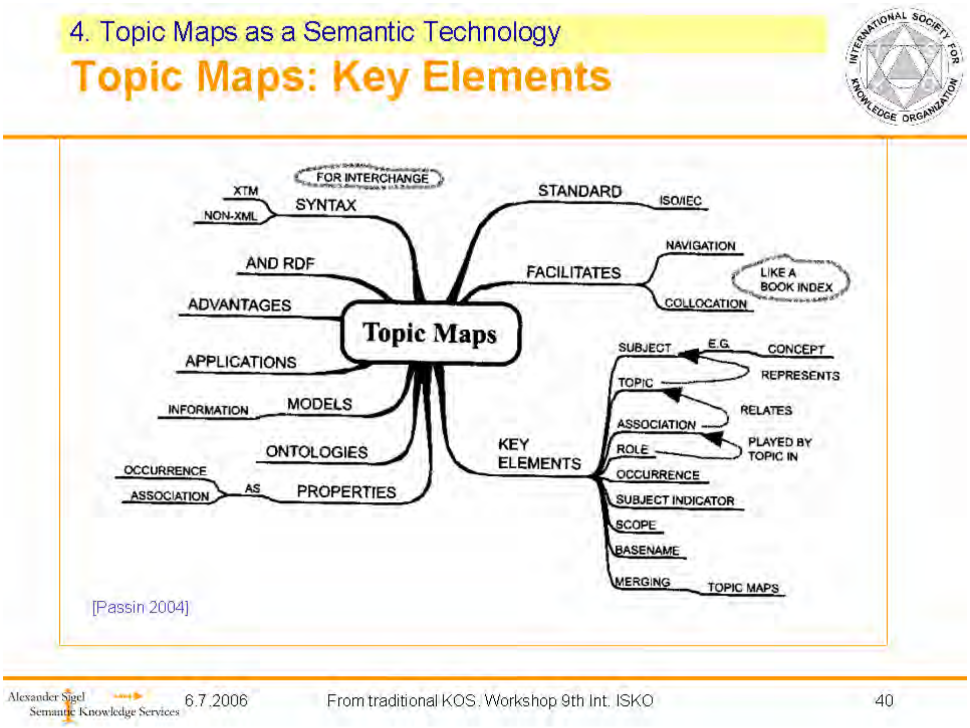
Topic Maps have been invented as a semantic integration methodology (Newcomb, 2003), solving the problem of how to tell that two statements are about the same subject. This gives us a means to virtually collocate subjects and resources relevant to these subjects.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14





In his book on the semantic web, Passin (2004), uses mind maps as chapter vignettes to communicate the contents of the chapters. On slide 40 the mind map for the topic map chapter is shown.



One can see that main concepts of the Topic Maps paradigm are:

1. **T**opics representing subjects, having several basenames
2. **A**ssociations (with roles and role players) (see e.g. (Ahmed & Moore, 2005))
3. **O**ccurrens (as typed links to information resources)

(the classical TAO of topic maps, see also (Garshol, 2006b)),
And of

- Subject Indicator, aiding humans and machines in the unambiguous identification and addressing of subjects, and the merging possible on the identity of subject indicators, and
- Scopes, discerning between views or validity restrictions.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop “Introducing Terminology-based Ontologies” (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

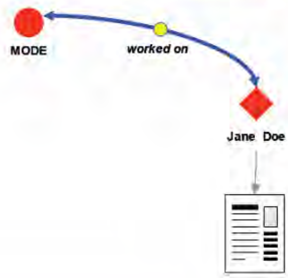
4. Topic Maps as a Semantic Technology

Topic Maps: Key Elements

ontopia

6. The TAO of Topic Maps

- **Topics**
  - represent things of interest
- **Associations**
  - represent relations between topics
- **Occurrences**
  - connect topics to information resources with relevant information

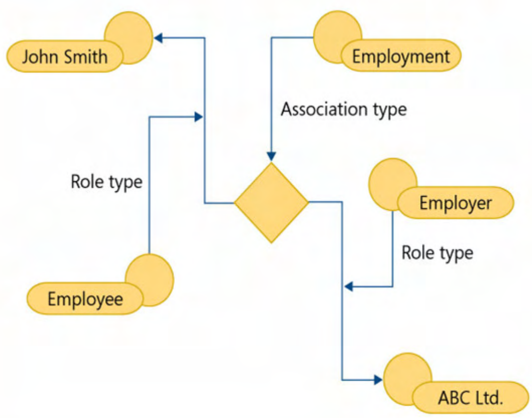


[Garshol 2006]

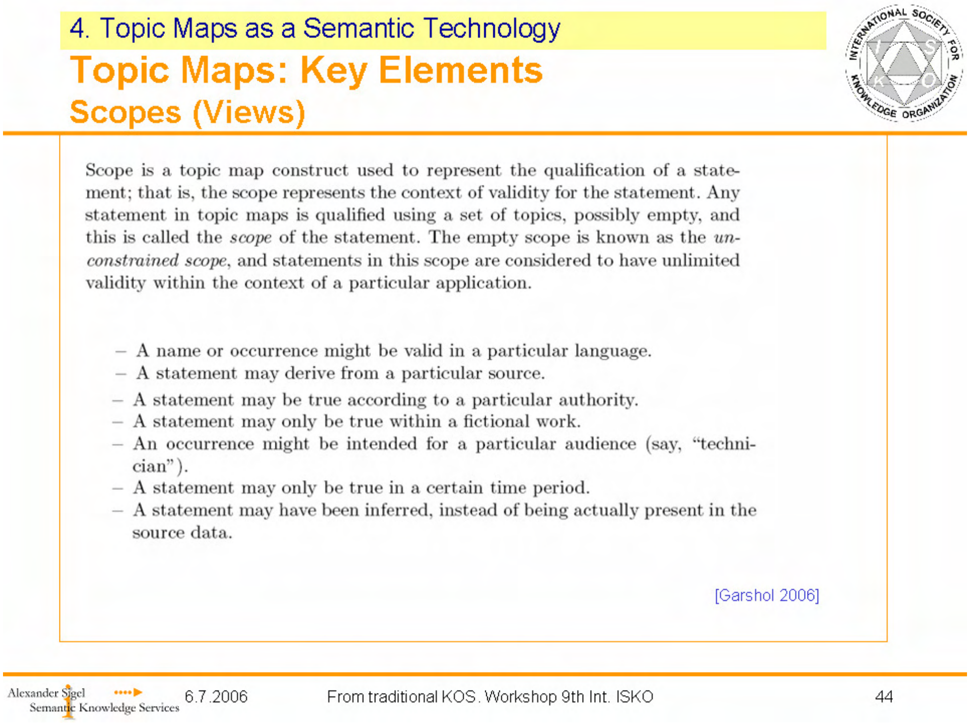
4. Topic Maps as a Semantic Technology

Topic Maps: Key Elements

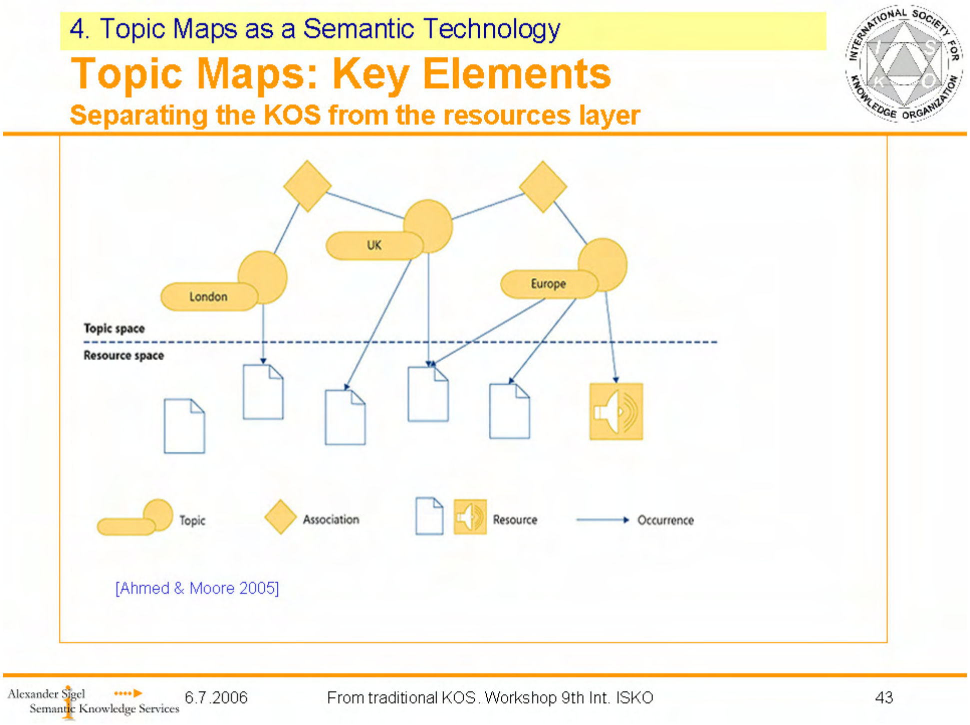
Associations



[Ahmed & Moore 2005]



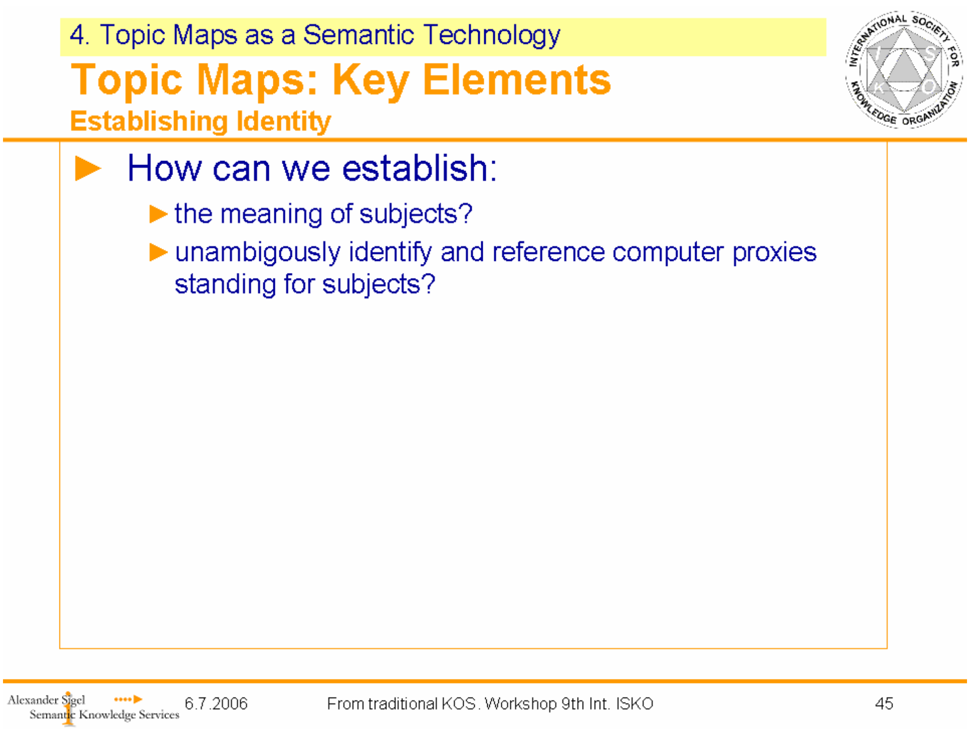
Interesting with topic maps is that the resource layer and the ontology layer are separated.



Published Subjects

The reader is referred to (Garshol, 2006b; Pepper, 2003, 2006a, 2006b; Pepper & Schwab, 2003; Sigel & Ahmed, 2007) for an introduction and explanation of the important concept of Published Resource Identifiers (PRIs) / Published Subject Identifier (PSIs) / Published Subjects.

In order to establish identity on the Semantic Web, an infrastructure based on Published Subjects has been proposed (Pepper, 2006a) which is apt both for RDF/OWL and Topic Maps (Pepper & Schwab, 2003). For each subject one likes to discourse (make statements) about (and which is not a web resource itself and such directly retrievable from the net), a subject proxy is published, i.e. a computer representation of the subject. The computer can discern between published subjects because each is associated with a unique published subject *identifier* (PSID) (a URI). This PSID must (should) resolve to a published subject *indicator* (PSI), which should contain human-interpretable metadata allowing a human to establish the identity of the subject. The PSI can contain machine-interpretable metadata. The metadata can be bound to KOS or ontologies. Recently, a new work item "Metadata for Published Subjects" has been proposed to ISO/IEC JTC 1/SC 34 (Naito, 2006). There have been first initiatives for registries of Published Subjects (see (Sigel, 2004) and the section on a PSI infrastructure for Topic Maps (Sigel, 2006), but Published Subjects are not yet in mainstream usage.



Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

4. Topic Maps as a Semantic Technology

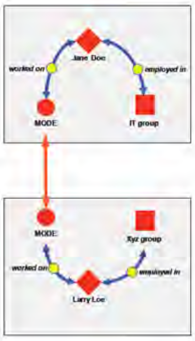
Topic Maps: Key Elements

Establishing Identity

ontopia

Topics have identity

- Topics can have globally unique identifiers
  - known as *PSIs*
- Two topics with the same PSI are the same
  - in a single topic map they would merge
- This makes it possible to merge Topic Maps automatically
  - assuming the necessary PSIs are provided
- This also makes it possible to connect information across Topic Maps



[Garshol 2006]

4. Topic Maps as a Semantic Technology

Published Subject

Published Subject
A published subject is any subject for which there exists at least one published subject indicator. A published subject indicator is a subject indicator published and maintained at an advertised location for the purpose of supporting topic map interchange and mergeability. A subject indicator is an information resource that is referred to from a topic map in an attempt to unambiguously identify the subject represented by a topic to a human being (Garshol & Moore, 2005; OASIS Published Subjects TC, 2003; Pepper, 2006).

Details in: Sigel, A., & Ahmed, K. (2006/2007). Topic-oriented portals. In A. Tatnall (Ed.), *Encyclopaedia of Portal Technology and Applications* (forthcoming): Idea Group

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

4. Topic Maps as a Semantic Technology

Published Resource Identifiers/Subjects

3. PUBLIC RESOURCE IDENTIFIERS

This section introduces a mechanism called Public Resource Identifiers and evaluates it on the basis of the foregoing requirements.

3.1 Overview of Public Resource Identifiers

The concept of Public Resource Identifiers has its roots in earlier concepts called Public Subjects and Published Subjects. Public Subjects originated within the Topic Maps community in late 1999 as the ISO/IEC 13250 [8] was nearing finalization.[1] Published Subjects is a refinement of Public Subjects developed in connection with the XML Topic Maps (XTM) specification in 2000-2001 [12], and refined by an OASIS Technical Committee in 2002-2003 [11]. The basic ideas are extremely simple and can be summarized as follows:

1) A *public resource identifier* (PRI) is an http: IRI[2] that has been minted for the express purpose of identifying some subject.

2) A PRI resolves to a human-interpretable document, known as a *public resource descriptor*, created explicitly to provide a compelling and unambiguous "description" of the subject so identified.

[Pepper 2006]

4. Topic Maps as a Semantic Technology

Published Resource Identifiers/Subjects

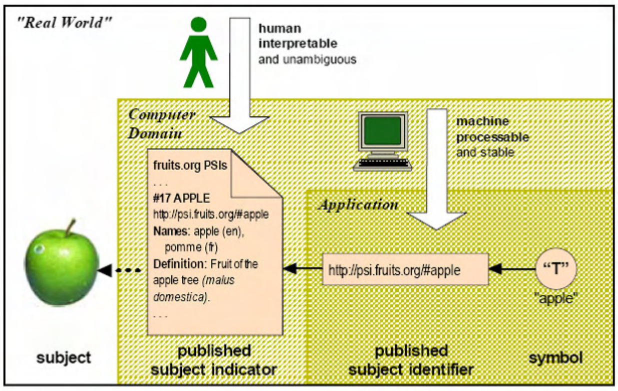


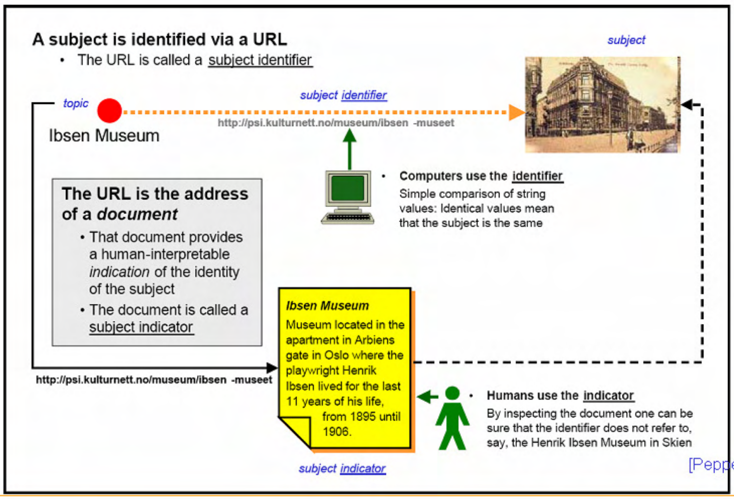
Figure 1: The roles of public resource identifier (PRI) and public resource descriptor (PRD) in identifying a subject

[Pepper 2006]

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

4. Topic Maps as a Semantic Technology

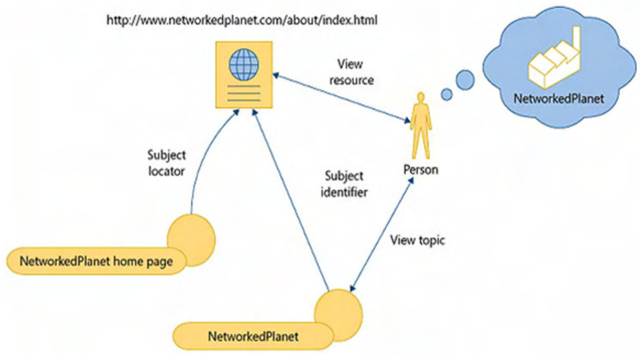
Published Resource Identifiers/Subjects



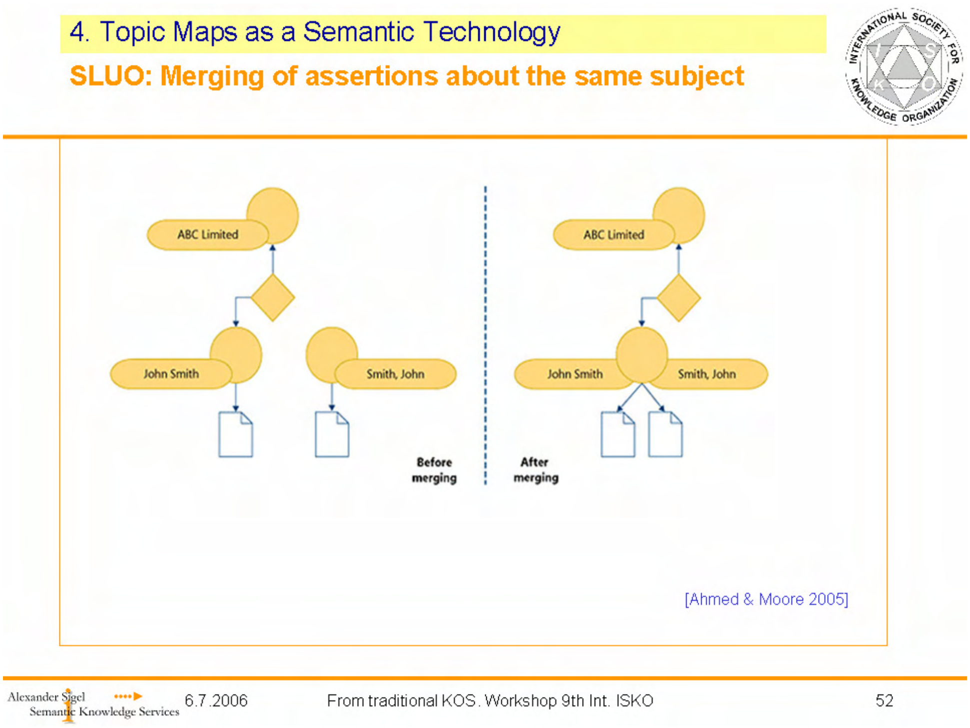
[Pepper 2006]

4. Topic Maps as a Semantic Technology

Published Resource Identifiers/Subjects



[Ahmed & Moore 2005]



How are Topic Maps and RDF related? (RDF/Topic Maps Interoperability)
RDF and Topic Maps can well co-exist and converted into each other (Garshol, 2003; Garshol & Naito, 2004; Naito, 2005). In general, RDF and Topic Maps are now quite interoperable, see the survey and recommendations resulting from active work on their interoperability (Pepper, Presutti, Vitali, Garshol, & Gessa, 2006; Pepper, Vitali, Garshol, Gessa, & Presutti, 2005, 2006). Therefore e.g. SKOS vocabularies can be used with Topic Maps.
No-one has to fear not being in the right semantic web camp, just because most people use RDF, and topic mappers are only a few. RDF is resource-centric, whereas Topic Maps is more assertion-centric. For knowledge organization purposes, Topic Maps are more natural than RDF since the modelling takes part on a more useful level.
According to (Garshol, 2005a, 2006a), RDF are triples (subject, property, object), whereas topic maps are quintuples (quints) (subject, property, object, **identity, scope**). This means that topic maps are semantically richer. Since both identity and scope are believed to be necessary in order to model knowledge, one should use topic maps. One can always convert down to RDF, using information.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

4. Topic Maps as a Semantic Technology

Topic Maps-RDF Interoperability: Model

- The Q model: RDF has tripels, Topic Maps has quints

Q represents Topic Maps as a set of five-tuples. The tuples can be thought of informally as a kind of extended RDF, with the following structure:
$(subject, property, identity, context, object)$
Here, *subject*, *property*, and *object* are as in RDF, *identity* is the identity of the statement, and *context* is the context in which the statement is considered true. In fact, the context is the identifier representing the set of topics making up the scope in a topic map. One difference with RDF, however, is that the values in the first four elements of a tuple can only be identifiers, which are propertyless objects used only as identifiers. Values, such as strings and URLs, are restricted to the last field.

- Topic Maps can be „downconverted" to RDF, losing information. With additional information, RDF can be „upconverted" to Topic Maps

In [Garshol05] a procedure for converting any TMDM instance into a Q instance is given, together with the reverse procedure, and also the same transformations for RDF models.

[Garshol05] Garshol, Lars Marius: *Q: A model for Topic Maps*. Proceedings of Extreme Markup 2005, IDEAlliance, August 1-5, Montréal, Canada. http://www.ontopia.net/topicmaps/materials/quads.html

[Garshol 2006]

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 53

4. Topic Maps as a Semantic Technology

Topic Maps-RDF Interoperability: WG

- **Guidelines for RDF/Topic Maps Interoperability**
- **W3C Editor's Draft 10 February 2006**
  http://www.ontopia.net/work/guidelines.html
- "The Resource Description Framework (RDF) is a model developed by the W3C for representing information about resources in the World Wide Web. Topic Maps is a model for knowledge integration developed by the ISO. This document provides guidelines for users who want to combine usage of the W3C's RDF/OWL family of specifications and the ISO's family of Topic Maps standards."
- "This document is aimed at anyone with an interest in the problem of RDF/Topic Maps interoperability and a willingness to acquire the necessary understanding of both formalisms. In particular it targets authors of topic maps and RDF documents, creators of tools for translating between RDF and Topic Maps, and people who seek assurance that data can be easily reused across the two paradigms."

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 55

How are ontologies and OWL (Web Ontology Language) related to Topic Maps?
The relationship between knowledge representation, ontological engineering and Topic Maps is detailed in Obrst & Liu in (Park & Hunting, 2002).
Topic Maps are hospitable to richer semantics, if needed, but we do not discuss this here. One early approach was by (Connolly, 2001), a quite different one was suggested by (Vatant, 2003, 2004). Recently, (Cregan, 2005) modelled TMDM (the Topic Maps Data Model) constructs in OWL-DL.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

4. Topic Maps as a Semantic Technology

Topic Maps und „Full“ Ontologies:

Ontology Engineering und Topic Maps

- A semantically rich and correct knowledge representation enhances the value of a topic map. Because an ontology is a knowledge representation, we assert that it plays an important role in topic map design.
- Whereas the topic map specification ensures syntactic interoperability, ontologies provide semantic interoperability.
- → Ontology-Driven Topic Map Approach: Topic Maps generated from ontologies

[Obrst & Liu 2003] Knowledge Representation, Ontological Engineering, and Topic Maps, in [Park & Hunting 2003, 103-148]

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 54

4. Topic Maps as a Semantic Technology

Topic Maps and „Full“ Ontologies

- Cregan, A. (2005, 2005-08-02). Building Topic Maps in OWL-DL. Extreme Markup Languages 2005 Retrieved 2006-03-08, from http://www.mulberrytech.com/Extreme/Proceedings/html/2005/Cregan01/EML2005Cregan01.html http://www.cse.unsw.edu.au/~annec/Building%20Topic%20Maps%20In%20OWL-DL%20(Extreme%20Aug2).pdf http://www.cse.unsw.edu.au/~annec/
- Vatant, B. (2004). Ontology-driven topic maps. Paper presented at the XML Europe 2004, 18-21 April, Amsterdam, NL.

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 60

What are useful references on Topic Maps and RDF?
For selected references to the Topic Map literature, see slides 56-59.
Useful introductions to the semantic web and semantic web technologies are (Passin, 2004), (Herman, 2006) and (Antoniou & van Harmelen, 2004). For a practical introduction to RDF, see e.g. (Powers, 2003) or (Hjelm, 2002). For an introduction to Topic Maps see e.g. selected chapters in (Park & Hunting, 2002).

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

4. Topic Maps as a Semantic Technology

Topic Maps: Introductory

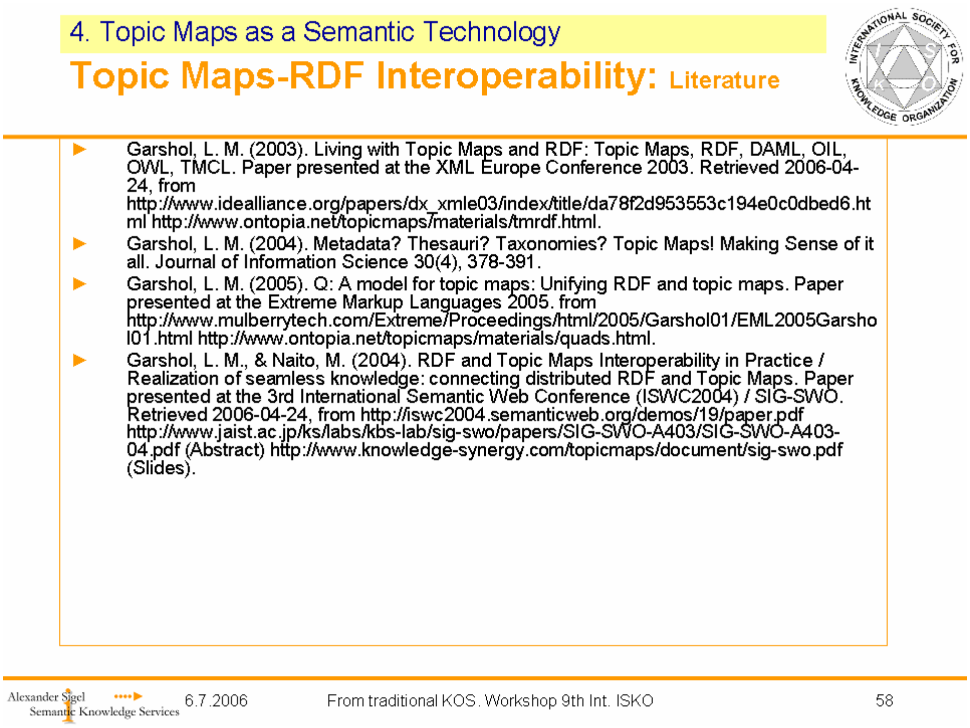
- Ahmed, K., & Moore, G. (2005). An Introduction to Topic Maps. *The Architecture Journal* 5. Retrieved 2006-04-24, from http://www.architecturejournal.net/2005/issue5/Jour5Intro/
- Ahmed, K., & Moore, G. (2006). Apply Topic Maps to Applications. *The Architecture Journal* 6. Retrieved 2006-04-24, from http://architecturejournal.net/2006/issue6/Jour6TopicMaps/
- Biezunski, M., Bryan, M., & Newcomb, S. R. (2002, 2002-05-19). ISO/IEC 13250 Topic Maps (Second Edition). Approved revised text. Retrieved 2006-04-24, from http://www.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-ed-v2.pdf http://www.jtc1sc34.org/repository/0322.htm
- Durusau, P., Newcomb, S. R., & Barta, R. (2006). Topic Maps Reference Model, 13250-5. Retrieved 2006-06-06, from http://www.jtc1sc34.org/repository/0710.pdf
- Garshol, L. M. (2006). *Topic Maps: Overview and Use Cases*. Paper presented at the Høykomseminar Leikanger. Retrieved 2006-04-24, from http://www.vestforsk.no/seminar/Ontopia.pdf (Slides only).
- Garshol, L. M., & Moore, G. (2005, 2005-12-16). Topic Maps — Data Model. Final Draft International Standard. Retrieved 2006-04-24, from http://www.isotopicmaps.org/sam/sam-model/
- Garshol, L. M., & Moore, G. (2006, 2006-05-02). Topic Maps — XML Syntax (Final Committee Draft, for review). *ISO 13250: Topic Maps* Retrieved 2006-05-04, from http://www.isotopicmaps.org/sam/sam-xtm/
- Maicher, L., Sigel, A., & Garshol, L. M. (2006). Topic Maps Research and Applications - TMRA 2005 "Charting the Topic Maps Research and Applications Landscape". In *Semantic Web Factbook 2005*.
- Park, J., & Hunting, S. (Eds.). (2002). *XML Topic Maps: Creating and Using Topic Maps for the Web*. Boston: Addison-Wesley Professional.

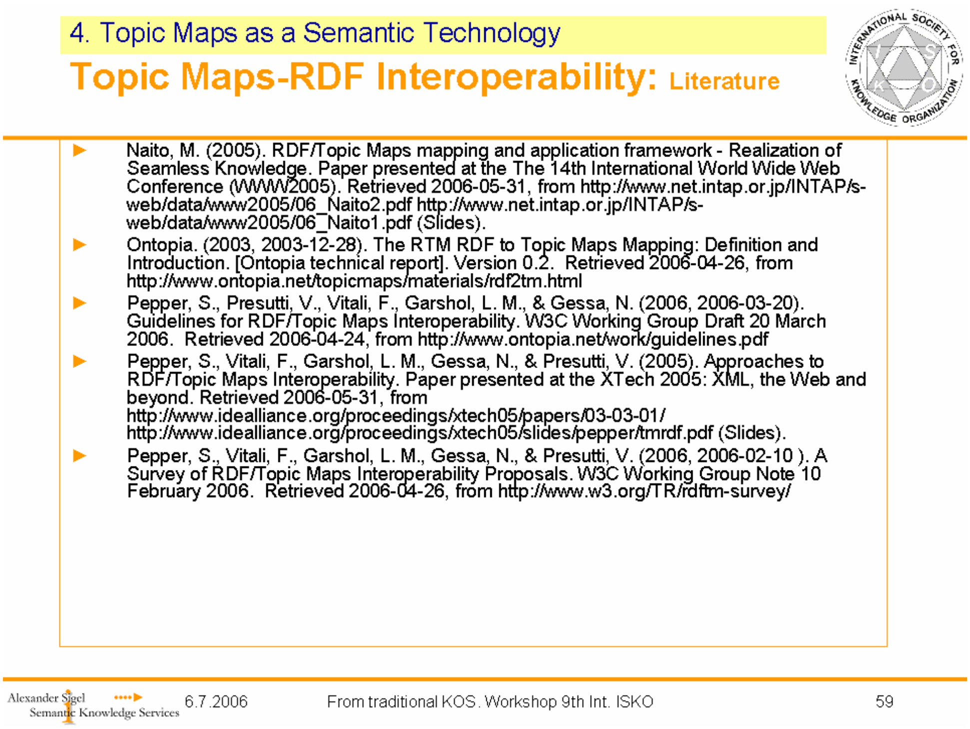
4. Topic Maps as a Semantic Technology

PRI/PSI-based Identity: Literature

- Maicher, L. (2004). *Subject Identification in Topic Maps in Theory and Practice*. Paper presented at the Berliner XML Tage 2004, Berlin, Germany.
- Maicher, L. (2006). Topic Map Exchange in the Absence of Shared Vocabularies. In L. Maicher & J. Park (Eds.), *Charting the Topic Maps Research and Applications Landscape, First International Workshop on Topic Maps Research and Applications, TMRA 2005, Leipzig, Germany, October 6-7, 2005, Revised Selected Papers* (Vol. LNAI 3873, pp. 77-92): Springer.
- Naito, M. (2006). Metadata for Published Subjects. In A. Sigel (Ed.) (pp. Proposal for a new work item (ISO/IEC JTC 1/SC 34)).
- Pepper, S. (2003, 2003-06-24). Published Subjects: Introduction and Basic Requirements. OASIS Published Subjects Technical Committee Recommendation. Retrieved 2006-04-26, from http://www.oasis-open.org/committees/download.php/3050/pubsubj-pt1-1.02-cs.pdf http://www.oasis-open.org/committees/tm-pubsubj/ http://www.oasis-open.org/committees/download.php/3052/pubsubj-pt1-1.02-cs.doc
- Pepper, S. (2006a). *The Case for Published Subjects*. Paper presented at the WWW2006 Workshop: Identity, Reference, and the Web (IRW2006). Retrieved 2006-06-06, from http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf (Final version) http://www.ibiblio.org/hhalpin/irw2006/spepper.pdf http://www.ibiblio.org/hhalpin/irw2006/spepper.html (Previous version, named: Towards the Semantic Superhighway. A Manifesto for Published Subjects) http://www.ontopia.net/topicmaps/materials/Public_Resource_Identifiers.ppt (Slides, named: The Case for Public Resource Identifiers: A Call for Action) http://www.ibiblio.org/hhalpin/irw2006/presentations/Public_Resource_Identifiers.ppt (Slides).
- Pepper, S. (2006b). *A Proposal for Public Resource Identifiers*. Paper presented at the TMRA 2006. Retrieved 2006-06-12.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14





How can ontologies and their representation in Topic Maps be employed in Knowledge Organization?

Early discussions of the relationship between knowledge organization and ontologies can be found in (Fischer, 1998; Poli, 1996; Soergel, 1999; Vickery, 1997). (Reimer, 2004) excellently introduces knowledge-based procedures for organizing and brokering information. He covers, albeit quite shortly, both traditional forms and RDF and Topic Maps. In sketching challenges for the semantic web from a culture viewpoint, (Veltman,

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

2004a; , 2004b) also discusses ontologies and knowledge organization. (Doerr, 2003) explains CIDOC CRM (Conceptual Reference Model) (CIDOC), an ontological approach for the semantic interoperability of metadata (in cultural heritage) which is currently elaborated as an international standard (ISO/DIS 21127).
For an in-depth review on Topic Maps in Knowledge Organization, see the chapter by Sigel in (Park & Hunting, 2002).

4. Leveraging Traditional KOS

Example: SKOS (Simple Knowledge Organization System)

I very much recommend the constantly updated semantic web tutorial by Herman (Herman, 2006), from which I have taken several slides.

SKOS can be considered as a quick fix to port traditional KOS for usage on the semantic web, without going into details of defining the internal semantics of concepts and relations. This is the reason why SKOS is needed also solutions like OWL exist for fuller semantics.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

5. Leveraging Traditional KOS

Why SKOS and OWL?

- OWL's precision not always necessary or even appropriate
  - *"OWL a sledge hammer / SKOS a nutcracker", or "OWL a Harley / SKOS a bike"*
  - *complement each other, can be used in combination to optimize cost/benefit*
- Role of SKOS is
  - *to bring the worlds of library classification and Web technology together*
  - *to be simple and undemanding enough in terms of cost and required expertise*
- A typical example: the **Glossary of project of W3C** stores all **terms in SKOS** (and extracted from W3C documents)

Herman, I. (2006, 2006-06-22). Tutorial on Semantic Web Technologies.
Retrieved 2006-06-22, from http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial/
Folie 150

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 64

Modern thesauri, also with SKOS, can be queried via webservices (see e.g. the CSA/NBII Biocomplexity Thesaurus, http://nbii-thesaurus.ornl.gov/thesaurus/).

5. Leveraging Traditional KOS

SKOS-Thesaurus with Webservice

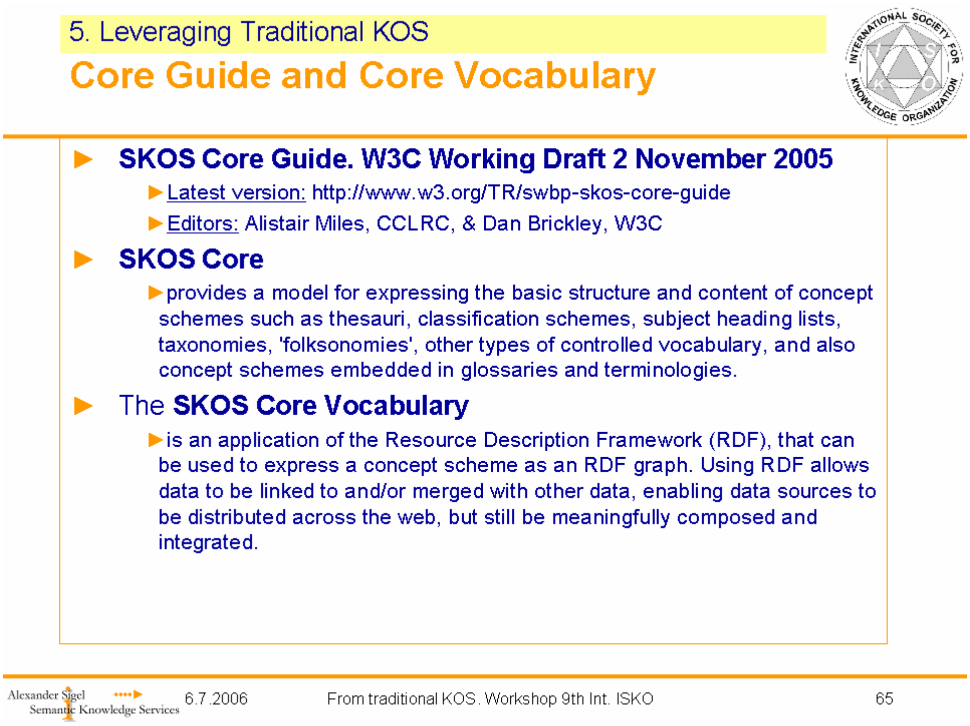
- CSA/NBII Biocomplexity Thesaurus
  - SOAP webservice
  - SKOS
- http://nbii-thesaurus.ornl.gov/thesaurus/

- Sall, Kenneth B. & Reck, Ronald P. (2005): Implementing a Government-wide Semantic Solution to Thesauri. XML 2005, Atlanta, GA, November 15, 2005.

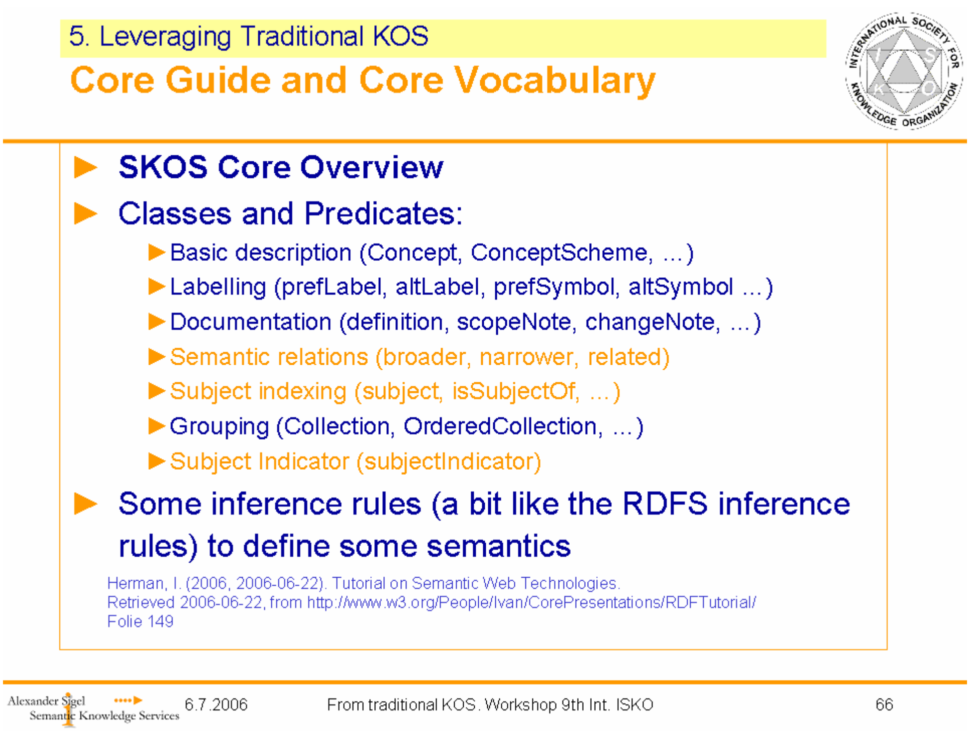
Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 63

SKOS: How to publish thesauri with RDF? (How can one represent a thesaurus with SKOS in RDF?)

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14



In the "Quick Guide to Publishing a Thesaurus on the Semantic Web" (Miles, 2005), it is described how "to express the content and structure of a thesaurus, and metadata about a thesaurus, in RDF", and the "Best Practice Recipes for Publishing RDF Vocabularies" (Miles & Swick, 2006) show how to publish "an RDFS or OWL vocabulary or ontology on the Web", using Apache. SKOS, the Simple Knowledge Organization System (Miles, 2006), is specified in (Miles & Brickley, 2005a) and (Miles & Brickley, 2005b) (see also (Miles, 2004) above).



Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

5. Leveraging Traditional KOS

Bringing a Thesaurus into the Semantic Web

- **"Quick Guide to Publishing a Thesaurus on the Semantic Web"**
- **"SKOS Mapping Vocabulary Specification"**

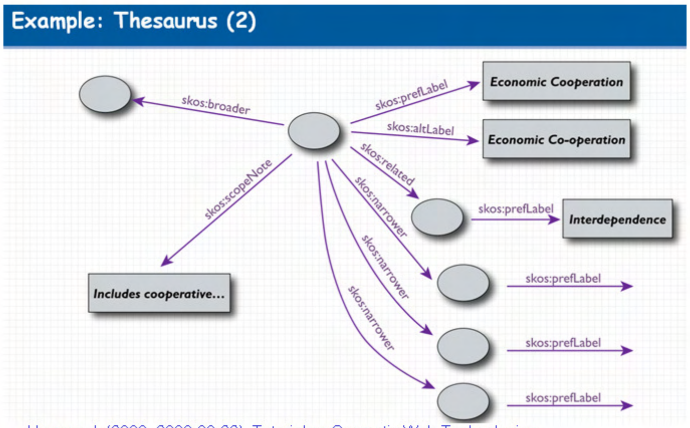
- van Assem, M., Malaisé, V., Miles, A., & Schreiber, G. (2006). A Method to Convert Thesauri to SKOS. In Y. Sure & J. Domingue (Eds.), *ESWC06* (Vol. LNCS 4011, pp. 95-109): Springer.

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 67

Slides 68 and 69 reconsider the traditional thesaurus entry of the UK archival thesaurus as SKOS thesaurus, in graphical and XML-serialized form.

5. Leveraging Traditional KOS

Thesaurus entry, graphically



Herman, I. (2006, 2006-06-22). Tutorial on Semantic Web Technologies. Retrieved 2006-06-22, from http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial/ Folie 148

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 68

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

5. Leveraging Traditional KOS

Thesaurus entry: RDF/XML

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:skos="http://www.w3.org/2004/02/skos/core#">

  <skos:Concept>
    <skos:prefLabel>Economic cooperation</skos:prefLabel>
    <skos:altLabel>Economic co-operation</skos:altLabel>
    <skos:scopeNote>Includes cooperative measures in banking, trade,
    industry etc., between and among countries.</skos:scopeNote>

    <skos:broader>
      <skos:Concept>
        <skos:prefLabel>Economic policy</skos:prefLabel>
      </skos:Concept>
    </skos:broader>

    <skos:related>
      <skos:Concept>
        <skos:prefLabel>Interdependence</skos:prefLabel>
      </skos:Concept>
    </skos:related>

    <skos:narrower>
      <skos:Concept>
        <skos:prefLabel>Economic integration</skos:prefLabel>
      </skos:Concept>
    </skos:narrower>

    <!-- ...more narrower terms omitted ... -->

  </skos:Concept>
</rdf:RDF>
```

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 69

How can one up-convert RDF vocabulary to Topic Maps, e.g. SKOS RDF thesauri?
How to convert RDF to Topic Maps and vice versa is conceptually described by Garshol (2003), and technically detailed in (RTMRDF, 2003). The necessary steps to up-convert SKOS from RDF to Topic Maps can be found in (Garshol, 2005b). (See slide 70).

5. Leveraging Traditional KOS

SKOS in Topic Maps: Demonstration

- Tutorial
  - Garshol, L. M. (2005, 2005-10-24). SKOS in Topic Maps [Blog entry]. http://www.garshol.priv.no/blog/10.html
- SKOS MAPPING
  - Lars Marius Garshol : RTM mapping file for SKOS from RDF to Topic Maps. http://www.garshol.priv.no/blog/mapping.rdff
  - Miles, A. (2004). SKOS: Simple Knowledge Organisation with the Semantic Web. DELOS 2004
  - Miles, A., Matthews, B., Beckett, D., Brickley, D., Wilson, M., & Rogers, N. (2005). SKOS: A language to describe simple knowledge structures for the web. Paper presented at the XTech 2005: XML, the Web and beyond.
- http://localhost:8080/omnigator/models/topicmap_complete.jsp?tm=MyThesaurus.xtm
  - **Creator:** Steve Pepper
  - **Description:** A Topic Maps ontology for monolingual thesauri based on SKOS (http://www.w3.org/TR/swbp-skos-core-guide/)
  - **Ontology Version:** 1.0
  - **Version:** $Id: MyThesaurus.xtm,v 1.2 2006/01/10 13:01:00 pepper Exp $

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 70

The question is: Why should I model my thesaurus with SKOS, which is RDF, and then leverage it to Topics Maps (this way I can't win additional information, because I do not add identity and scope information). However, you said that Topic Maps are "better" than RDF. Why not directly modelling the thesaurus as a topic map, and then converting down?

This is well possible and has done several times. It is highly recommended, and one should use thesaurus architectural patterns like those proposed by (Ahmed, 2003). However, sometimes it is more important to stay within a certain active community (here: SKOS). If you like to do so, you can also with Topic Maps, you do not have to use RDF, if you prefer Topic Maps, but you can interoperate with others.

Additional literature on leveraging traditional KOS can be found on slides 71 and 72.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

5. Leveraging Traditional KOS

Literature

- Miles, A., & Brickley, D. (2005a, 2005-11-02). SKOS Core Guide. W3C Working Draft. Retrieved 2006-03-08, from http://www.w3.org/TR/swbp-skos-core-guide
- Miles, A., & Brickley, D. (2005b, 2005-11-02). SKOS Core Vocabulary Specification. W3C Working Draft. Retrieved 2006-03-08, from http://www.w3.org/TR/swbp-skos-core-spec
- Miles, A., & Swick, R. (2006, 2006-01-18). Best Practice Recipes for Publishing RDF Vocabularies. Editor's Draft. Retrieved 2006-03-08, from http://www.w3.org/2001/sw/BestPractices/VM/http-examples/
- Paslaru Bontas, E., Tolksdorf, R., & Schrader, T. (2004). *Ontology-based Knowledge Organization in a Semantic Web for Pathology*. Paper presented at the KNOW 2004: Inquiring Knowledge Networks on the Web.
- PORT. (n.d.). Semantic Web Best Practices and Deployment Working Group. Porting Thesauri Task Force (PORT). Retrieved 2006-03-08, from http://www.w3.org/2004/03/thes-tf/mission
- Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., & Katz, S. (2004). Reengineering Thesauri for New Applications: the AGROVOC Example, *Journal of Digital Information* (Vol. 4, pp. Article No. 257).
- Stuckenschmidt, Heiner: RDF and Thesauri - A Perfect Match. http://www.sigsemis.org/columns/rdf/Tehsauri, last accessed 2006-03-15
- Tudhope, D. (2004). Semantic Terminology Services: experiences from the FACET project. *DELOS 2004* Retrieved 2006-03-08
- van Assem, M., Malaisé, V., Miles, A., & Schreiber, G. (2006). A Method to Convert Thesauri to SKOS. In Y. Sure & J. Domingue (Eds.), *ESWC06* (Vol. LNCS 4011, pp. 95-109): Springer.
- van Assem, M., Menken, M. R., Schreiber, G., Wielemaker, J., & Wielinga, B. (2004). *A Method for Converting Thesauri to RDF/OWL*. Paper presented at the 3rd Int. Semantic Web Conference (ISWC 2004), Nov. 2004, Hiroshima, Japan.

Alexander Sigel Semantic Knowledge Services 6.7.2006 From traditional KOS. Workshop 9th Int. ISKO 72

How can one represent FRBR expressions with RDF and Topic Maps?
(Not shown in the tutorial)
FRBR, the Functional Requirements for Bibliographic Records, are a model for work-oriented finding aids (FRBR, 1998). The inner structure can be modelled and exploited with semantic web solutions. (Jørgensen, 2004)[Jørgensen 2004] presented a RDF solution, and there is a Topic Map-based one (Sigel, 2005).

6. Conclusion

In sum, a major challenge will be the semantic interoperability of semantic web applications. Some years ago, the task was making KOS available on the web in HTML in digital form and providing registries for machine-readable KOS. Nowadays, KOS, adapted as lightweight ontologies, presented and interlinked on the web, together with a PSI-based information architecture, can aid in establishing part of the semantic foundation of the semantic web.

We have seen how Topic Maps can be part of the move from traditional KOS towards semantically richer ontologies. Check the learning objectives if I catched everything promised.

Citation: **Sigel, Alexander (2006): From traditional Knowledge Organization Systems (authority files, classifications, thesauri) towards ontologies on the web**, pp. 3-53, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Learning Objectives

1. understand what (computer science) ontologies are
2. understand how ontologies relate to traditional Knowledge Organization Systems (KOS)
3. gain insights into the potential of ontologies for the purpose of knowledge organization, retrieval and use
4. learn about migration paths from traditional KOS to ontologies expressed with semantic web technologies (RDF, OWL, Topic Maps)
5. see practical examples of how modern KOS with rich semantic relations can be represented and queried with current semantic tools
   – including conversion from RDF to Topic Maps (e.g. semantic relationships will be modelled in a Topic Map, and a thesaurus will be represented in RDF with SKOS and then up-converted to Topic Maps)

Learning Objectives

6. gain an overview (not an in-depth knowledge!) about ontological tools and their functions which are useful for modern knowledge organization
7. understand the difference between upper ontologies and domain ontologies, how this relates to concept definition and category theory, and what this means in practice
8. receive an update on the state-of-the-art in knowledge organization with ontologies
9. learn about the 10 best books and articles on this exciting topic
10. be prepared to understand the second and third part of this workshop and how they interrelate.

5. References

Ahmed, K. (2003). *Topic Map Design Patterns for Information Architecture*. Paper presented at the XML Conference 2003. Retrieved 2006-04-24, from http://www.idealliance.org/papers/dx_xml03/index/title/d2481ecb47625095ef1d920268.html http://www.idealliance.org/proceedings/xml03/slides/ahmed/ahmed.pdf (Slides).

Ahmed, K., & Moore, G. (2005). An Introduction to Topic Maps. *The Architecture Journal* 5. Retrieved 2006-04-24, from http://www.architecturejournal.net/2005/issue5/Jour5Intro/

Amann, B., Fundulaki, I., & Scholl, M. (2000). Integrating Ontologies and Thesauri for RDF schema creation and metadata querying. *International Journal of Digital Libraries, 3*(3), 221-236.

Antoniou, G., & van Harmelen, F. (2004). *A Semantic Web Primer*: MIT Press.

Barité, M. G. (2000). The notion of "category": its implications in subject analysis and in the construction and evaluation of indexing languages. *Knowledge Organization, 27*(1-2), 4-10.

Blumauer, A., & Pellegrini, T. (2006). Semantic Web und semantische Technologien: Zentrale Begriffe und Unterscheidungen. In T. Pellegrini & A. Blumauer (Eds.), *Semantic Web: Wege zur vernetzten Wissensgesellschaft* (pp. 9-25): Springer.

Ceusters, W., Smith, B., & Goldberg, L. (2005). A Terminological and Ontological Analysis of the NCI Thesaurus. *Methods of Information in Medicine, 44*, 498-507.

CIDOC. (2004-05-03). CIDOC CRM Home page. Retrieved 2006-03-08, from http://cidoc.ics.forth.gr/

Connolly, D. (2001). *Building Ontologies with RDF and DAML+OIL.* Paper presented at the WWW10, May 2001, Hong Kong.

Corazzon, R. (2006, 2006-04-03). Ontology: A resource guide for philosophers. Retrieved 2006-03-08, from http://www.formalontology.it/

Cregan, A. (2005, 2005-08-02). Building Topic Maps in OWL-DL. *Extreme Markup Languages 2005* Retrieved 2006-03-08, from http://www.mulberrytech.com/Extreme/Proceedings/html/2005/Cregan01/EML2005Cregan01.html http://www.cse.unsw.edu.au/~annec/Building%20Topic%20Maps%20In%20OWL-DL%20(Extreme%20Aug2).pdf http://www.cse.unsw.edu.au/~annec/

Daconta, M. C., Obrst, L. J., & Smith, K. T. (2003). *The Semantic Web: A guide to the future of XML, Web Services and Knowledge Management*: Wiley.

Dahlberg, I. (1995). Conceptual Structures and Systematization. *International Forum on Information and Documentation (IFID), 20*(3), 9-24.

Davies, J., Fensel, D., & van Harmelen, F. (2002). *Towards the Semantic Web: Ontology-Driven Knowledge Management*: Wiley.

DELOS. (2004, 2004-08-30). Between Knowledge Organization and Semantic Web: Semantic Approaches in Digital Libraries (DELOS Regional Awareness Event). Lund University Libraries, Sweden, June 23, 2004. Retrieved 2006-03-08, from http://www.delos.info/eventlist/LUB1.html

Doerr, M. (2003). The CIDOC Conceptual Reference Module: An Ontological Approach to Semantic Interoperability of Metadata. *AI Magazine, 24*(3), 75-92.

Elzenheimer, M., Grollius, T., Heinemann, E., & Sternhuber, J. (2005). Vergleichende Buchbesprechung: Ontologien. *Wirtschaftsinformatik, 47*(4), 298-309.

Endres-Niggemeyer, B. (2000). *Empirical methods for ontology engineering in bone marrow transplanation.* Paper presented at the 6. Tagung der Deutschen Sektion der

Internationalen Gesellschaft für Wissensorganisation (ISKO): Globalisierung und Wissensorganisation: Neue Aspekte für Wissen, Wissenschaft und Informationssysteme, 23.-25. September 1999, Hamburg.

FAO. (2006). Applied Ontologies in FAO. Retrieved 2006-03-08, from http://www.fao.org/aims/onto_domains.jsp http://www.fao.org/aims/aos.jsp

Fensel, D. (2003). *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce* (2 ed.). Berlin: Springer.

Fensel, D., Wahlster, W., Lieberman, H., & Hendler, J. (2003). *Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential*: MIT Press.

Fischer, D. H. (1998). *From Thesauri towards Ontologies?* Paper presented at the Structures and Relations in Knowledge Organization. 5th Int. ISKO Conference, 25-29th August 1998 Lille, France.

Fischer, D. H. (2004, 2004-02-18). Converting a Thesaurus to OWL: Notes on the Paper "The National Cancer Institute's Thésaurus and Ontology". Retrieved 2006-03-08, from http://www.ipsi.fraunhofer.de/contec/pubFulltexts/NCIReview18Feb04.pdf

FRBR. (1998). *Functional Requirements for Bibliographic Records. Final Report.* (Vol. 19). München: Saur.

Fugmann, R. (1993). *Subject Analysis and Indexing. Theoretical Foundation and Practical Advice*. Frankfurt/Main: Indeks.

Fugmann, R. (1999). *Inhaltserschließung durch Indexieren: Prinzipien und Praxis*. Frankfurt/Main: DGD.

Garshol, L. M. (2003). *Living with Topic Maps and RDF: Topic Maps, RDF, DAML, OIL, OWL, TMCL*. Paper presented at the XML Europe Conference 2003. Retrieved 2006-04-24, from http://www.idealliance.org/papers/dx_xmle03/index/title/da78f2d953553c194e0c0dbed6.html http://www.ontopia.net/topicmaps/materials/tmrdf.html.

Garshol, L. M. (2004). Metadata? Thesauri? Taxonomies? Topic Maps! Making Sense of it all. *Journal of Information Science 30*(4), 378-391.

Garshol, L. M. (2005a). *Q: A model for topic maps: Unifying RDF and topic maps*. Paper presented at the Extreme Markup Languages 2005. Retrieved 2006-07-13, from http://www.mulberrytech.com/Extreme/Proceedings/html/2005/Garshol01/EML2005Garshol01.html http://www.ontopia.net/topicmaps/materials/quads.html.

Garshol, L. M. (2005b, 2005-10-24). SKOS in Topic Maps. (Blog entry by Lars Marius Garshol). Retrieved 2006-03-08, from http://www.garshol.priv.no/blog/10.html

Garshol, L. M. (2006a). *tolog – a topic maps query language.* Paper presented at the Charting the Topic Map Research and Applications Landscape. TMRA'05 – International Workshop on Topic Map Research and Applications, October 6-7, 2005, Leipzig.

Garshol, L. M. (2006b). *Topic Maps: Overview and Use Cases*. Paper presented at the Høykomseminar Leikanger. Retrieved 2006-04-24, from http://www.vestforsk.no/seminar/Ontopia.pdf (Slides only).

Garshol, L. M., & Naito, M. (2004). *RDF and Topic Maps Interoperability in Practice / Realization of seamless knowledge: connecting distributed RDF and Topic Maps*. Paper presented at the 3rd International Semantic Web Conference (ISWC2004) / SIG-SWO. Retrieved 2006-04-24, from http://iswc2004.semanticweb.org/demos/19/paper.pdf http://www.jaist.ac.jp/ks/labs/kbs-lab/sig-swo/papers/SIG-SWO-A403/SIG-SWO-A403-04.pdf (Abstract) http://www.knowledge-synergy.com/topicmaps/document/sig-swo.pdf (Slides).

Gilchrist, A. (2003). Thesauri, Taxonomies and Ontologies – An Etymological Note. *Journal of Documentation, 59*(1), 7-18.

Goldbeck, J., Fragoso, G., Hartel, F., Hendler, J., Parsia, B., & Oberthaler, J. (2003). The National Cancer Institute's Thésaurus and Ontology. *Journal of Web Semantics, 1*(1).

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge Acquisition, 5*(2), 199-220.

Handschuh, S., & Staab, S. (2003). *Annotation for the Semantic Web*. Amsterdam: IOS Press.

Herman, I. (2006, 2006-06-22). Tutorial on Semantic Web Technologies. Retrieved 2006-06-22, from http://www.w3.org/People/Ivan/CorePresentations/RDFTutorial/

Hjelm, J. (2002). *Creating the semantic Web with RDF*. New York: Wiley.

Hjørland, B. (2002). Domain Analysis in Information Science: Eleven Approaches - Traditional as Well as Innovative. *Journal of Documentation*, 422-462.

Hjørland, B., & Albrechtsen, H. (1995). Toward a New Horizon in Information Science: Domain Analysis. *Journal of the American Society for Information Science, 46* (6), 400-425.

Iyer, H. (1995). *Classificatory Structures: Concepts, Relations and Representation*. Frankfurt/Main: Indeks.

Jacob, E. K. (2003). Ontologies and the Semantic Web. *Bulletin of the American Society for Information Science and Technology, 29*(4 April/May, Special Issue on Semantic Web), 19-22.

Jørgensen, P. H. (2004). *ONE Shared Authority Control with RDF and FRBR.* Paper presented at the Interoperability: New Challenges and Solutions. 28th Library Systems Seminar, June 9-11th, 2004, Trondheim, Norway.

Kumar, A., & Smith, B. (2005). *Oncology Ontology in the NCI Thesaurus.* Paper presented at the Artificial Intelligence in Medicine Europe (AIME 2005), 23–27 July 2005, Aberdeen.

Lauser, B. (2004). From thesauri to rich ontology: The AGROVOC case. FAO Ontology Portal Prototypes: Fishery. Building a rich ontology from AGROVOC. *DELOS 2004* Retrieved 2006-03-08, from
http://www.delos.info/eventlist/LUB1/FAO_Delos_20041.ppt
http://www.delos.info/eventlist/LUB1/Boris_Lauser/ontology_portals_fishery_20040205.ppt

Lauser, B., Sini, M., Liang, A., Keizer, J., & Katz, S. (2006). *From AGROVOC to the Agricultural Ontology Service / Concept Server. An OWL model for creating ontologies in the agricultural domain*. Paper presented at the 5th European Networked Knowledge Organization Systems (NKOS) Workshop, at the 10th ECDL Conference. Retrieved 2006-07-13, from http://www.ukoln.ac.uk/nkos/nkos2006/.

Miles, A. (2004). SKOS: Simple Knowledge Organisation with the Semantic Web. *DELOS 2004* Retrieved 2006-03-08, from
http://www.delos.info/eventlist/LUB1/Alistair_Miles/Lund_presentation.ppt

Miles, A. (2005, 2005-05-17). Quick Guide to Publishing a Thesaurus on the Semantic Web. W3C Working Draft. Retrieved 2006-03-08, from http://www.w3.org/TR/swbp-thesaurus-pubguide

Miles, A. (2006, n.d.). Simple Knowledge Organisation System (SKOS). Retrieved 2006-03-08, from http://www.w3.org/2004/02/skos/

Miles, A., & Brickley, D. (2005a, 2005-11-02). SKOS Core Guide. W3C Working Draft. Retrieved 2006-03-08, from http://www.w3.org/TR/swbp-skos-core-guide

Miles, A., & Brickley, D. (2005b, 2005-11-02). SKOS Core Vocabulary Specification. W3C Working Draft. Retrieved 2006-03-08, from http://www.w3.org/TR/swbp-skos-core-spec

Miles, A., & Swick, R. (2006, 2006-01-18). Best Practice Recipes for Publishing RDF Vocabularies. Editor's Draft. Retrieved 2006-03-08, from http://www.w3.org/2001/sw/BestPractices/VM/http-examples/

Naito, M. (2005). *RDF/Topic Maps mapping and application framework - Realization of Seamless Knowledge*. Paper presented at the The 14th International World Wide Web Conference (WWW2005). Retrieved 2006-05-31, from http://www.net.intap.or.jp/INTAP/s-web/data/www2005/06_Naito2.pdf http://www.net.intap.or.jp/INTAP/s-web/data/www2005/06_Naito1.pdf (Slides).

Naito, M. (2006). Metadata for Published Subjects. In A. Sigel (Ed.) (pp. Proposal for a new work item (ISO/IEC JTC 1/SC 34)).

Newcomb, S. R. (2003). *A Semantic Integration Methodology*. Paper presented at the Extreme Markup Languages Conference 2003. Retrieved 2006-07-01, from http://www.mulberrytech.com/Extreme/Proceedings/html/2003/Newcomb01/EML2003Newcomb01.html.

Nielsen, M. L. (2000). *Domain analysis, an important part of thesaurus construction.* Paper presented at the 11th ASIS&T SIG/CR Classification Research Workshop, November 12th, 2000, Chicago, IL.

Nielsen, M. L. (2003). Thesaurus construction. Key issues and selected readings. *Cataloging & Classification Quarterly, 37*(3-4), 57-74.

NKOS. (2003, 2003-05-16). NKOS 2003: Building a meaningful web: From traditional Knowledge Organization Systems to new semantic tools. The 6th NKOS (Networked Knowledge Organization Systems and Services) workshop at ACM-IEEE Joint Conference on Digital Libraries (JCDL). JCDL 2003, Houston, Texas. Retrieved 2006-03-08, from http://nkos.slis.kent.edu/DL03workshop.htm

NKOS. (2005, 2005-11-11). Homepage of NKOS (Networked Knowledge Organization Systems and Services). Retrieved 2006-03-08, from http://nkos.slis.kent.edu/

Noy, N. F., & McGuiness, D. L. (2001). *Ontology Development 101: A Guide to Creating Your First Ontology*. Stanford: Stanford Knowledge Systems Laboratory and Stanford Medical Informatics.

Park, J., & Hunting, S. (2002). *XML Topic Maps: Creating and Using Topic Maps for the Web*. Boston: Addison Wesley.

Paslaru Bontas, E., Tolksdorf, R., & Schrader, T. (2004). *Ontology-based Knowledge Organization in a Semantic Web for Pathology.* Paper presented at the KNOW 2004: Inquiring Knowledge Networks on the Web.

Passin, T. B. (2004). *Explorer's Guide to the Semantic Web*: Manning.

Patel, M., Koch, T., Doerr, M., & Tsinaraki, C. (2005, 2005-06-29). Semantic Interoperability in Digital Library Systems. DELOS: A Network of Excellence on Digital Libraries. Deliverable D5.3.1 within Task 3: Semantic Interoperability, WP5: Knowledge Extraction and Semantic Interoperability. Retrieved 2006-03-08, from http://www.lub.lu.se/tk/publ/SI-in-DLs.doc

Pepper, S. (2003, 2003-06-24). Published Subjects: Introduction and Basic Requirements. OASIS Published Subjects Technical Committee Recommendation. Retrieved 2006-04-26, from http://www.oasis-open.org/committees/download.php/3050/pubsubj-pt1-1.02-cs.pdf http://www.oasis-open.org/committees/tm-pubsubj/ http://www.oasis-open.org/committees/download.php/3052/pubsubj-pt1-1.02-cs.doc

Pepper, S. (2006a). *The Case for Published Subjects*. Paper presented at the WWW2006 Workshop: Identity, Reference, and the Web (IRW2006). Retrieved 2006-06-06, from http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf (Final version) http://www.ibiblio.org/hhalpin/irw2006/spepper.pdf http://www.ibiblio.org/hhalpin/irw2006/spepper.html (Previous version, named: Towards the Semantic Superhighway. A Manifesto for Published Subjects) http://www.ontopia.net/topicmaps/materials/Public_Resource_Identifiers.ppt (Slides, named: The Case for Public Resource Identifiers: A Call for Action) http://www.ibiblio.org/hhalpin/irw2006/presentations/Public_Resource_Identifiers.ppt (Slides).

Pepper, S. (2006b). *A Proposal for Public Resource Identifiers*. Paper presented at the TMRA 2006. Retrieved 2006-06-12.

Pepper, S., Presutti, V., Vitali, F., Garshol, L. M., & Gessa, N. (2006, 2006-03-20). Guidelines for RDF/Topic Maps Interoperability. W3C Working Group Draft 20 March 2006. Retrieved 2006-04-24, from http://www.ontopia.net/work/guidelines.pdf

Pepper, S., & Schwab, S. (2003). *Curing the Web's Identity Crisis: Subject Indicators for RDF*. Paper presented at the XML Conference 2003. Retrieved 2006-04-24, from http://www.idealliance.org/papers/dx_xml03/html/abstract/05-01-05.html http://www.ontopia.net/topicmaps/materials/identitycrisis.html.

Pepper, S., Vitali, F., Garshol, L. M., Gessa, N., & Presutti, V. (2005). *Approaches to RDF/Topic Maps Interoperability*. Paper presented at the XTech 2005: XML, the Web and beyond. Retrieved 2006-05-31, from http://www.idealliance.org/proceedings/xtech05/papers/03-03-01/ http://www.idealliance.org/proceedings/xtech05/slides/pepper/tmrdf.pdf (Slides).

Pepper, S., Vitali, F., Garshol, L. M., Gessa, N., & Presutti, V. (2006, 2006-02-10). A Survey of RDF/Topic Maps Interoperability Proposals. W3C Working Group Note 10 February 2006. Retrieved 2006-04-26, from http://www.w3.org/TR/rdftm-survey/

Perreault, J. M. (1994). Categories and Relators: A new Schema. *Knowledge Organization, 21*(4), 189-198.

Poli, R. (1996). *Ontology for Knowledge Organization.* Paper presented at the Knowledge Organization and Change, 4th Int. ISKO Conference, 15-18th July 1996 Washington.

Poli, R. (2002). Ontological Methodology. *International Journal of Human-Computer Studies, 56*(6), 639-664.

Poli, R. (2003a). Descriptive, Formal and Formalized Ontologies. In D. Fisette (Ed.), *Husserl's Logical Investigations Reconsidered* (pp. 193-210): Kluwer.

Poli, R. (2003b). Ontology, Step by Step. In H. Fujita & P. Johannesson (Eds.), *New Trends in Software Methodologies, Tools and Techniques. Procs. Lyee_W03* (pp. 60-70): IOS Press.

Poli, R. (2006). *Upper Ontologies Hold It Together.* Paper presented at the Ninth International ISKO Conference 2006, Vienna, Austria.

PORT. (n.d.). Semantic Web Best Practices and Deployment Working Group. Porting Thesauri Task Force (PORT). Retrieved 2006-03-08, from http://www.w3.org/2004/03/thes-tf/mission

Powers, S. (2003). *Practical RDF*: O'Reilly.

Rahmstorf, G. (1994). *A New Thesaurus Structure for Semantic Retrieval.* Paper presented at the 47th FID General Assembly: Finding New Values and Uses of Information, Tokyo, October 6-8, 1994.

Reimer, U. (2004). Wissensbasierte Verfahren der Organisation und Vermittlung von Information. In R. Kuhlen, T. Seeger & D. Strauch (Eds.), *Grundlagen der praktischen Information und Dokumentation* (5 ed., Vol. 1, pp. B 3, 155-166). München: Saur.

Riggs, F. W. (1996/1997). Onomantics and Terminology. Parts 1-4. *Knowledge Organization*, 23(21), 25-34, 23(23), 156-168, 123(154) 216-224, 124(151), 158-117.

RTMRDF. (2003, 2003-12-28). The RTM RDF to topic maps mapping. Definition and introduction. Technical report, v0.2. Retrieved 2006-03-08, from http://www.ontopia.net/topicmaps/materials/rdf2tm.html

Schmaltz, R. (2004). Semantic Web Technologien für das Wissensmanagement. *Arbeitsbericht Nr. 1/2004* Retrieved 2006-03-08, from http://www.wi2.wiso.uni-goettingen.de/getfile?DateiID=466

Schmitz-Esser, W. (1999). Thesaurus and Beyond: An Advanced Formula for Linguistics Engineering and Information Retrieval. *Knowledge Organization, 26*(1), 10-22.

Schmitz-Esser, W. (2006). *An Introduction to Integrative Cross-Language Ontology (ICLO).* Paper presented at the Ninth International ISKO Conference 2006, Vienna, Austria.

Sigel, A. (2004, 2006). *Organisation verteilten Wissens mit semantischen Wissensnetzen und der Aggregation semantischer Wissensdienste am Beispiel Digitale Bibliotheken/Kulturelles Erbe.* Paper presented at the 9. Tagung der Deutschen ISKO (Wissensorganisation 2004). Wissensorganisation und gesellschaftliche Verantwortung. Informationstheoretische, ökonomische und ethische Aspekte, Duisburg.

Sigel, A. (2005, 2005-10-13). FRBR and XTM. Retrieved 2006-03-08, from http://kpeer.wim.uni-koeln.de/~sigel/Projects/FRBR_and_XTM.html

Sigel, A. (2006). Report on the Open Space Sessions. In L. Maicher & J. Park (Eds.), *Charting the Topic Maps Research and Applications Landscape. First International Workshop on Topic Map Research and Applications, TMRA 2005, Leipzig, Germany, October 6-7, 2005, Revised Selected Papers* (Vol. 3873, pp. 271-280): Springer.

Sigel, A., & Ahmed, K. (2007). Topic-Oriented Portals. In A. Tatnall (Ed.), *Encyclopaedia of Portal Technology and Applications* (pp. forthcoming): Idea Group.

Soergel, D. (1999). The rise of ontologies or the reinvention of classification. *Journal of the American Society for Information Science, 50*(12), 1119-1120.

Soergel, D., Lauser, B., Liang, A., Fisseha, F., Keizer, J., & Katz, S. (2004). Reengineering thesauri for new applications: the AGROVOC example. *Journal of Digital Information, 4*(4), Article No. 257.

Staab, S., & Studer, R. (2005). *Handbook on Ontologies. Springer*: Springer.

Stuckenschmidt, H., & van Harmelen, F. (2005). *Information Sharing on the Web*: Springer.

Studer, R., Oppermann, H., & Schnurr, H.-P. (2001, September 2001). Die Bedeutung von Ontologien für das Wissensmanagement. *Whitepaper ontoprise – semantics for the web, September 2001* Retrieved 2006-03-08, from http://www.ontoprise.de/documents/Bedeutung_von_Ontologien_fuer_WM.pdf

Sure, Y., Ehrig, M., & Studer, R. (2006). *Automatische Wissensintegration mit Ontologien.* Paper presented at the Modellierung 2006, Workshop Modellierung für Wissensmanagement, Innsbruck, AT, March 2006.

Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*: MIT Press.

Tudhope, D. (2004). Semantic Terminology Services: experiences from the FACET project. *DELOS 2004* Retrieved 2006-03-08, from http://www.delos.info/eventlist/LUB1/Doug_Tudhope/DELOS04-Tudhope.ppt

Tudhope, D., & Koch, T. (2004). New applications of knowledge organization systems: introduction to a special issue. *Journal of Digital Information, 4*(4), Article No. 286.

Tudhope, D., & Nielsen, M. L. (2005, 2005-11-02). CFP for Special Issue of New Review of Hypermedia and Multimedia (NRHM) on Knowledge Organization Systems and Services. Retrieved 2006-03-08, from http://lists.w3.org/Archives/Public/public-esw-thes/2005Nov/0018.html

Ullrich, M., Maier, A., & Angele, J. (2004). Taxonomie, Thesaurus, Topic Map, Ontologie – ein Vergleich V.1.4 (ontoprise whitepaper). Retrieved 2006-03-08, from http://www.ullri.ch/ullri_ch/download/Ontologien/ttto13.pdf

van Assem, M., Malaisé, V., Miles, A., & Schreiber, G. (2006). A Method to Convert Thesauri to SKOS. In Y. Sure & J. Domingue (Eds.), *ESWC06* (Vol. LNCS 4011, pp. 95-109): Springer.

van Assem, M., Menken, M. R., Schreiber, G., Wielemaker, J., & Wielinga, B. (2004). *A Method for Converting Thesauri to RDF/OWL.* Paper presented at the 3rd Int. Semantic Web Conference (ISWC 2004), Nov. 2004, Hiroshima, Japan.

Vatant, B. (2003, 2003-08-01). Cooking for the Semantic Web. OWL and Topic Map Pudding. Retrieved 2006-03-08, from http://www.mondeca.com/owl/owltm.htm

Vatant, B. (2004). *Ontology-driven topic maps.* Paper presented at the XML Europe 2004, 18-21 April, Amsterdam, NL.

Veltman, K. H. (2004a). Challenges for the Semantic Web and Information Systems from Culture. *AIS SIGSEMIS Bulletin, 1*(1), 37-40.

Veltman, K. H. (2004b). Towards a Semantic Web for Culture. *Journal of Digital Information*, *4* 4. Retrieved 2004-03-15, from http://jodi.tamu.edu/Articles/v04/i04/Veltman/veltman.pdf

Vickery, B. C. (1997). Ontologies. *Journal of Information Science, 23*(4), 277-286.

Welty, C. (2003). Ontology research. *AI magazine, 24*(3), 11-12.

Zelewski, S. (2002). Wissensmanagement mit Ontologien: Eine einführende Darstellung. *Arbeitsbericht Nr. 15* Retrieved 2006-03-08, from http://www.kowien.uni-essen.de/publikationen/ArbeitsberichtWissensmanagementmitOntologien.pdf http://www.kowien.uni-essen.de/publikationen.cfm

Zelewski, S., Schütte, R., & Siedentopf, J. (1999, 2001). *Ontologien zur Repräsentation von Domänen.* Paper presented at the Wissen in Unternehmen - Konzepte, Maßnahmen, Methoden. Tagung der Wissenschaftlichen Kommission "Wissenschaftstheorie" des Verbandes der Hochschullehrer für Betriebswirtschaft e.V., 18.-19.06.1999, Berlin.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Introduction to Integrative Cross-Language Ontology (ICLO)

Formalizing[1] and interrelating textual knowledge to enable intelligent action and knowledge sharing

Winfried Schmitz-Esser
http://www.schmitz-esser.de
University of Applied Sciences, Hamburg

Abstract: A framework for a general, integrative cross-language ontology (ICLO) is presented, and its underlying conditions are explained and discussed. ICLO combines some selected, mainly IT and AI driven features from last-word ontological thought and practice with aspects and requests from philosophy, linguistics and information science in a fresh, unconventional way. The ontology works on (1) fragments of terminology selected and transferred from natural language source texts, and (2) on multiple, defined semantic inter-relations between the meanings of such fragments. The ontology is multilingual, and to express it with Topic Maps initial work is under way. The ICLO is designed to be complemented and enhanced by light, linguistic software for particular languages. Basics of ontological cross-language modelling, conditions for construction and maintenance, and the most salient points in application are addressed, such as cross-language text mining on heterogeneous sources, and autonomous knowledge generation by virtue of the ontology proper. A common, natural language platform is offered for integration and reuse of knowledge organized in other ontologies and related knowledge order works and systems such as classifications, taxonomies, and thesauri.

Part 1. Outline of ICLO

What the Integrative Cross-Language Ontology is

The Integrative Cross-Language Ontology (ICLO) is an ontology well in the sense of Gruber **[1]** as a shared conceptualisation, but with the difference that it includes instances defined as instantiations of their semantically related concepts as well **[2]**.

The ICLO is conceived as a model; a model of the knowledge about the real world or, more realistically, some little part or parts of it. The model is constructed by means of natural language (NL)[2] terms and term sequences that represent its respective prototype. The "I" in the ICLO stands for "integrative" in two respects:

- Main elements of both domain and upper ontologies are synergistically combined.
- A common, natural language-based platform is offered for the integration of knowledge found organized in existing, isolated ontologies and in many other established Knowledge Organization (KO) resources, including structured term lists, classifications, taxonomies, and thesauri.

---

[1] Formalization is a process of representing knowledge in terms of categorical (linguistic) expression. POLI, R. 2003: *Descriptive* **[19]** 185, so as to conform with requirements of symbol processing. Not contradictory to this is JAENECKE, P.1996: *Elementary Principles* **[20]** 91, who sees formalization as "process of representing knowledge in a formal language in such a way that syntax and semantics are identical".
[2] For abbreviations and use of terms see legend: "Acronyms and use of terms" at the end of this paper.

Also, the ICLO is multi-lingual in principle which means that, on the languages implemented, input and output functions feature true semantic cross-language capabilities.

The ICLO presented extends on the author’s earlier work on multi-lingual ontologies (Schmitz-Esser 1999, 2001, 2003, 2005) for which practical feasibility has been shown and practical experience was earned in open, real-world applications such as EXPO 2000, **[3], [4]** SERUBA **[5], [6], [22], [23], [24],** as well as in some more specialized, proprietary realizations. Further advancements are owed to joint work conducted with Roberto Poli, and to Alexander Sigel (**[7]**, and as documented in his contribution to this working paper].

ICLO rationale for a quick reader

Using just some single words (as we usually do it on the Internet) is not enough to efficiently perceive or communicate the semantics of a textual message. It seems as if we had forgotten how in our childhood we learned to understand the world, namely by establishing relations between what the words mean. Mother told us that a frog ***is*** *an animal able to live on land and in water, that we* ***descend from*** *our ancestors, that too much fast food* ***harms*** *health,* ***may cause*** *obesity,* ***may hamper*** *participation in social life, etc.*

In a machine significantly upgraded from the search engine on the Internet, knowledge of this sort could then be used to automatically prompt questions like: ***Name other hazards threatening young people’s health,*** *or:* ***What may be obstacles to participation in social life?***

Let us call such a machine a Knowledge Machine. Performing like a search engine for information retrieval (IR) would be just one of its functions, the much more important one being knowledge generation (KG). The machine would be capable of exploiting knowledge inferred according to logic and reasoning. It could then give direct, understandable answers to the above type of questions, could help to find out different expressions for the same meaning, also in other languages, read and understand unknown texts, detect or hint to hitherto unknown knowledge in such texts, ease perception and learning in a suggestive, playful way, guide information seekers through the world of thought and knowledge, and still more.

The Knowledge Machine would be based on an appropriate knowledge model, an ontology, where the knowledge is represented in linguistic terms, but independently of a specific language. This, then, enables 1. that knowledge organized in different ways and different languages could be related to, or integrated in, the ontology, and 2. that services derived from the ontology’s functions can be rendered in all languages implemented alike.

The paper presented outlines the conditions required so as to make this ontology feasible. Technical questions such as appropriate mapping, description languages, and all other software engineering needed for the approach shall be left for later consideration.

Main conditions required:

- *A general construction principle is needed that enables linguistic representation of semantic relations existing between objects of thought and reality, - from basic formalized unit ("argument") to basic formalized phrase ("statement"). Two different arguments (A1, A2), linked by at least one semantic relation (R) build a statement (S). The statement states what the knowledge is. All statements together form the semantic knowledge net. No argument is accepted to stand alone. For details, see rules of construction **RC1 – RC3.***

- *More complex statements may be built by bracketing, and several simple or bracketed statements may be linked by joint IDs to form representations of still higher expressiveness. **RC4 – RC5.***

- *Human constructors (peers, curators) assuming editorial responsibility would extract the knowledge from validated source texts and formalize it for the ontology. Transferring the authoritative wording encountered with the authors of the texts would be their main guideline. All choices and sources involved in this process must be thoroughly documented. **RC6 – RC7.***

- *An argument either stands for a category (concept, universal, notion) or a case (instance of category). The denominator of the instance is the (individual) name. **RC8.***

- *Automatic processes of logic and reasoning as required for KG cannot be expected to justly apply on floppy or changing semantics. On the other hand, committed word use is cursed. So, the choice of good representatives of the meaning in a given natural language (NL) is reduced. Expressions only which represent the object in a clear, univocal, and exhaustive way can be admitted. One such expression for each object is stipulated as selected representative (proxy) in the system (DESC), the others ranging as equivalent expressions (AAE). Third type on a respective Equivalence Chain of Expressions (ECE) are expressions with equal, but still other meaning (MULTI). To attain efficient term control in construction, a KWIC register is required for each language implemented. Both categories and instances have their respective DESC, AAE and MULTI. **RC9 – RC11.***

- *Stipulating instances requires utmost care and often also some additional research since rights and interests of the named parties are at stake. A typology of instances is proposed as an example. **RC12 – RC13.***

- *To share and evaluate knowledge organized the same or other ways in third systems such as word lists, taxonomies, thesauri, defined types of linking must be foreseen. **RC14**.*

- *To enable that in unknown texts a content of relevance can be automatically detected and captured, or that machine-generated output adapted to a question can be produced, it is necessary to establish additional reference to some categories on higher abstract level. Special qualifiers derived from an Upper Ontology (UOQ) are foreseen to be attributed to each single argument. Instances will by definition carry the same UOQ as their category counterparts. **RC15**.*

- *Intra-language semantic relations such as synonyms must be distinguished from their "essential", pan-language counterparts. Each language implemented in the ICLO will require an equivalence chain of expressions (ECE) of its own. The same applies for expressions used for semantic relations (ECRE).* ***RC16, RC19, RC20.***

- *Complying with all these demands, the ICLO argument ends up with eight different elements.* ***RC18****.*

- *Now, what about the semantic relations in the starring role? First of all, to match with reality, the ontology must allow any relation to be modelled in at least four different modes. Some chemical substance* ***is*** *[assertive], or* ***may be*** *[modal], or* ***is intended to be*** *[intention], or* ***is not*** *[negation] beneficial to health.* ***RC21****.*

- *An almost endless variety of semantic relationships between objects can be discerned, but relatively few, general, robust relations actually play the predominant role in normal communication. The charming thing about these few is that what they mean needs not much explanation. This happy circumstance can be exploited for knowledge modelling. And where finer granularity has to be dealt with, semantic relating and group building may be the answer. But which are those happy few? Based on more than a decade's research, a canon ("Arsenal") of 23 basic semantic relations, grouped in five main sections, is proposed. A canonized format for other, more refined relationships is indicated.* ***RC22.***

- *Stating knowledge for a knowledge machine wouldn't be worth the entry without reference to its validity and truth in time. This calls for provision of corresponding timelines.* ***RC23.***

- *Well, - "needing not much explanation" (above) does not mean needing no explanation at all. Here comes the full canon of definitions for all 23 semantic relations of the "Arsenal".* ***RC24.***

- *Light grouping of semantic relations, and putting them in a loose hierarchical order of the BROADER/NARROWER type, promises yields formerly out of reach.* ***RC25.***

- *Finally, the ontology will have to know and learn how to cope with the immense, dynamic way of wording authors use to choose in their texts of the different languages implemented on the ontology.* ***RC26****.*

The promise of ontologies

The concept of ontology as a description and explanation of what "is" reaches back to the philosophers of antiquity. Today's computerized ontologies are considered as key technologies in coping with the ever more pressing task of organizing the knowledge of our times. Funded on principled Knowledge Organization[3], their promise, in last instance, is the achievement of easy and trustful access to all available knowledge, and its unhampered reuse and exchange, on world-wide level.

As to the present, we are still way off such bold expectations.

[3] SIGEL, A. 2003: *Topic Maps* **[18]** 389

Present-day shortcomings and trade-offs

Of course, after all, the W3 with its many inroads on henceforth uncharted territories has become a reality. Modern ontologies have emerged in great number[4]. But wherever advances in this and related types of modern knowledge organization works and systems have been achieved, this was on very small segments of world knowledge only, or for very special tasks, and with heavy trade-offs in many respects such as general usability, and world-wide knowledge sharing.

In the face of an explosion-like growth, and of chaotic dispersion of knowledge and knowledge sources, an ever more urgent demand is being felt for their efficient organization, for reliability, and especially so for responsible intellectual and social control.

Divergent views and attitudes have come up with respect to this secular phenomenon. Yet, no general consensus has been reached on what this "information bang" could mean for the society and its sustainable development. In the face of an extreme complexity of concepts, systems, actors, facts, events, and of changing views and terminologies, holistic, ontological views understandably are difficult to be gained. Looking at the details, however, explanations and policies flow abundantly. The focus is on single aspects and most urgent shortcomings, such as:

1. Machines and programs needed that are capable of coping with much higher complexity than to-day, and at much higher levels of expressiveness in representing principled knowledge in terms of the different languages (in multilingual environments);
2. More transparency of knowledge is required that is open for public access, along with adequate means of how knowledge and knowledge sources can be gathered in a more predictable manner, preferably according to defined, openly accessible knowledge structures;
3. More, better, preferably standardized, bridges and links to the vast, heterogeneous worlds of established knowledge organization (library authority files, classifications, thesauri, etc.), in order to reach workable levels of interoperability, if not compatibility;
4. Moreover, the call is out for knowledge sources to be rendered more reliable and accountable; for consistency of treatment that must be established on these sources, including warranties on their diachronic stability and fair documentation, and ease of ways and means of accessing them at any time on the W3;
5. The goal is a general facility of responsible, computer-assisted, openly accessible generation of knowledge from warranted sources. Such generation must be sanctioned by, and "safe" for, the society, and directly and ubiquitously applicable for any entitled user (man and machine) in processes of interoperability.

Ontologies could substantially help to bring such achievements about, but have to be adapted, refined and improved for the purpose.

ICLO: A fresh, unconventional approach

The Integrative Cross-Language Ontology (ICLO) is a fresh, unconventional approach in this direction, hopefully that some of the aforementioned shortcomings can be tackled to a certain, after all, better defined degree. New, still further refined and enhanced methods and techniques may evolve from the approach as more practical experience is gained on real-world applications.

[4] For details see Alexander Sigel's first part of this working paper

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

The core of the problem: Capturing the representational function of natural language (NL) independently of a particular language. How, and in how far, can this be achieved?

For an ontology of our times the ideal condition is that the knowledge it contains has been/can be expressed by means of natural language expressions which unmistakably represent the content (the meaning, the semantics). The expressions then could be processed as symbols of the meaning, logic and calculus could be applied on the symbols, and new, hitherto unknown knowledge that everybody would understand could be generated.

However, natural language, by nature, is not of this simple kind. Most natural expressions are based on polysems, metaphors, metonymies and other highly complex semantic structures (such as anaphora) which all differ from language to language, and culture to culture. For instance, looking at close distance, in a given language there are hardly any two words standing exclusively for exactly the same, unmistakable meaning.

Nevertheless, expressions do exist that determine a distinct meaning in an un-confoundable (univocal) way, independently of any sort of context, and dispensing with all further need for special commitment in term use. We have to find, select and isolate these expressions in a knowledgeable way, and to take provision in our systems that we can deal with them.

First of all, we have to prepare for considerably longer strings of words than we are used to in our current search systems, and we would have to give up the idea that all what at best can be used for computerized processes is a nominal phrase. In natural language texts nominality is seldom encountered. In fact, the "nominality craze" in KO, which in last instance was a result of early restrictions in organizational and technical governance, has alienated us from the much more active way of expression we encounter in un-"enhanced", natural language texts. There, for normal, it's the verbs that govern the meaning.

> *Having semantic relations at our disposal in stating knowledge, what holds us back to express that a "looming labour shortage in the economy" may CAUSE the "lowering of interest rates", that "diversifying the portfolio" is INSTRUMENTAL to "minimize risk exposure", that Achilles' way of "conflict solving" on the plains of Troy WAS "choosing a short life of honour over a long, ordinary life"?*

So, for representation of a given object of thought in a computer, any NL expression may be a candidate provided it denominates the meaning in a clear, univocal, shortest possible but exhaustive way.

- To be solidly based on real-world demand, expressions of this quality should be manifest in terms of the natural language (NL) of a language community. They would have to be detected in validated source texts by a skilled, knowledgeable human resource (peers, curators) who transfer the respective wording and stipulate it for use and reuse in the ontology. Since univocal, context-free expressions are but a fragment of natural language expressions, provision must be taken that expressions with almost equivalent, or multiple meaning, metaphors, metonymies and other more complex semantic structures can be dealt with in a subtle way as the ontology is refined and grows.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

- It is clear, of course, that the representational function must be performed by means of *some* language, whichever it may be[5].
- Two concomitant problems arise thereof: (1) What are the right words, since in all languages by far not every term or term sequence is apt of carrying out the main function which is sole, full and univocal representation, and (2) binding this function exclusively to NL terms of one (privileged) language is not enough. The axiomatic framework of the ontology must allow univocal linguistic representation of a given concept or instance of concept in NL terms of any modern language.
- It is assumed that this in principle is possible for any item that now and in the future might be up for representation[6]. Of course, this is the ideal. That such conditions can be exhaustively fulfilled cannot seriously be claimed. The axiomatic framework of the ontology therefore must allow for a controlled, but dynamic and open knowledge model, but it is obvious that solutions feasible under these conditions will always hit still relatively narrow frontiers of expressiveness.
- In how far under fixed axiomatic conditions a given level of expressiveness can be maintained also in other languages may indicate wide fields for future model enhancement and scientific research.
- Only if the two problems above are coped with in a satisfactory way, automatic processes of propositional (sentential) logic and extended logic can be applied to the knowledge isolated and inferred in the ontology.
- New knowledge then can be generated either automatically or synergistically by man and machine. The ideal is that such formalized knowledge, one day, in last instance and in principle, would be open and apt for use, reuse, merging, segmentation for anyone entitled, be it man or machine, everywhere on the globe.

But how to formalize linguistic representations (terms, term sequences) of knowledge (content, meaning, message, etc.) in unique and univocal units, or fragments, of NL utterances of – potentially – *any* given natural language, i. e. independently of the conditions of a *particular* language?

This is the question [8].

To express a certain meaning in a given language, the corresponding words may be different and used in different ways. It is virtually unpredictable how we find them applied in single, specific cases. This does not contradict the observation that a surprisingly high degree of consensus is found to exist in all language communities with respect to how a certain piece of knowledge is currently addressed.

This is where in practical applications statistics come in. The thesis is defended that ontologies can be constructed on the assumption that wherever in a source of rank (validated source) some fragment of knowledge is addressed by some author in terms of NL, the way he/she does it is not arbitrary but has value of its own. Experience **[3], [4], [5], [6]** shows that to construct an ontology like the ICLO, a suitable breeding environment can be established where (1) a knowledgeable, pluri-lingual, well trained human resource, (2) the textual knowledge resource, and (3) the IT resource are organized to work together in such a way that the statistical factor comes to a bearing as to the current choice of words and word strings for target contents of relevance.

---

[5] "… independent of a specific language but not independent of linguistic expression in language". SIGEL, A. 2003: *Topic Maps* **[18]** 405.
[6] SCHMITZ_ESSER, W. 1999: *Thesaurus and Beyond* **[9]** 11

The Human Resource is of prime importance

The importance of the human resource in constructing, maintaining, and editing the ontology can hardly be overstated. A most demanding, high-level task (source evaluation, knowledge extraction, transfer, coding and related research) requires high-level domain and language specialists (peers, curators). Efficient, ergonomic machine-assistance must be offered to staff of this rank. Investment needed for brainware and software will be considerable, and exhibition of trustful, credible long-term sustainability of operation will be the prime condition for wider acceptance and use.

The editor has to assume classic editors' responsibilities. The ICLO is not for open solutions, but various different business models based on distributed construction and knowledge sharing seem possible, including solutions of the Wikipedia type. Main problems cannot be discussed in this paper, such as copyright clearance (needed for proper documentation), rigor of encoding and update (for reliability), consistency of choice (for equilibrated performance), etc.

Onomasiologic approach and transfer of expressions

Expertly trained, human intellect, assisted by suitable linguistic programs and tools would be capable of selecting and stipulating sole, univocal representations in at least one language which then are apt to be formalized. Multi-language projects can be realized by bilingual manpower and language pair overlap. In all languages, as prime condition, two basic principles have to be applied in the process:

- An onomasiologic approach
- Transfer of expressions, as known from current practice in translation

Onomasiologic approach says that for a distinct meaning the current expressions in NL must be sought (not the other way round), and transfer basically says that in a cross-language environment the same approach must apply also for the target language.

Experience, however, shows that in practical applications rigorous enforcement of the principle is not possible. Exceptions must be admitted, be it only for the sake of a tolerable workflow in ontology construction and update. But concomitant bias can be corrected in some degree by further inference under same statistical conditions. This way, the ontology could be dynamically widened and deepened **[6].**

Admittedly, there are many factors in this approach which in practice are critical and/or difficult to activate. They will be dealt with in detail further down in this paper along with the specific provisions for the ICLO model, including references to practical experience gained by the author on first real-world applications.

For this introduction, suffice it to know that the proposed procedure implies:

1. The ICLO is not static.
2. The ICLO does not pretend to equal NL as to expressiveness and exactitude of articulation, but spells considerable advance over state-of-the-art expressiveness achieved with present day ontological engineering technology.
3. The ICLO essentially depends on constant update, correction and expansion, is dynamic in principle and therefore needs a visible, credible pledge for sustained continuity of operation.

4. The ICLO cannot be realized without the best of machine assistance, but at any point of time remains under human intellectual control and responsibility.
5. The ICLO's formalized part of the knowledge can generate new knowledge by logic and reasoning, provides terminology control and reasonable degree of consistency.
6. The ICLO one day may be mature, but never will be complete.

State-of-the-art: Particular use of the language enables solutions, but poses constraints
One of the main reasons for the unsatisfactory, early state in ontological engineering is a widespread neglect of natural language and its underlying complex semantics and corresponding linguistic structures and rules. Typically, to fix an intended meaning, one or two words are considered sufficient. For the sake of simplicity, they are stripped from their original context and used, like cans, to hold and transport some meaning. The meaning is then ascribed for exclusive use in a distinct domain, or for a specific task, according to the goal of the respective ontology.

In many cases, the definition of such "commitment" is kept apart of the ontology proper, as a pre-supposition set *ex ante*, without semantic, axiomatic grounding and no corresponding bearing on the reasoning mechanism inside.

In part, the terms are used within the structure of the ontology grossly the same way as in the respective language communities. This then raises questions of how a distinct community of ontology users can be defined vis-à-vis other communities, how homogeneous, stable as in time and composition the community is, how consensuality and rigor of encoding among the members of the community can be maintained, etc., which, in turn, calls for explanations of what a given, stipulated term or term sequence means, or at least for some recommendation on how a stipulated term or term sequence shall be used the "correct" way, etc., etc.

Committed use of terms abounds in thesauri and classifications, and also in modern ontologies; we find "authoritative" jargons and idiosyncracies, restrictions of the "valid in domain only" type, all sorts of "annotations", "definitions", "Used For" rules and the like.

The usual naïve, positivistic approach no doubt favours un-reflected, simplified, abridged or even idiosyncratic use of the words, all the more so as in the face of an otherwise un-mastered linguistic complexity only a committed practice enables feasibility of at least some (however narrowly defined) goal which is the formalization of the stipulated meaning.

Formalization of this sort then perfectly complies with what computerized systems require for processes of logic and reasoning on the ascribed, intended meaning. Excellent, most useful results can be achieved on the method. The words are taken as symbols in the sense of classic AI, and the Physical Symbol Systems Hypothesis (PSSH)[7] can be applied[8].

But in dealing with knowledge from NL texts we are on different ground. Here is no ontology designer who tells us what he decided the term to mean, or the domain to encompass. We are on the much more complex ground of textual knowledge and knowledge sources of common life, of human thought and culture from the most divergent, more or less amorphous language communities encountered in real life.

[7] NEWELL, A., SIMON, H.A. 1976: *Computer Science* **[37]** saying that symbol processing "has the necessary and sufficient means for general intelligent action".
[8] Fensel, D. 2002: *Language standardization* **[29].**

One remarkable, reverse side of committed practice in word use is that, in striving for ever greater exactness and precision in operational environments of this kind, practical needs force the ontology constructor into ever more "precise" stipulation, which constantly widens the distance between committed term use and normal term use. The drive is known from classical thesaurus construction. Ontologies, with their committed, stipulated terminologies, make no difference.

The tendency has two consequences:

1. The ontology's potential for purposes of text mining is jeopardized.
2. In interoperable environments, the much needed consensuality on such committed use is difficult to achieve.

The result is known: An archipelago of organized but isolated term-based knowledge resources, many of them of the highest quality, but with little or no linguistic, social and technical grounding in textual public discourse. In some way, rather a tendency of drifting further apart is observed, - a drift towards further specialization and seclusion.

Summarizing, it can be stated that

- "No one can be expected (let alone be forced) to use a certain controlled vocabulary according to given rules."[9]
- However, committed term use is state-of-the-art technology in present day ontologies.
- Admittedly, committed term use enables a high degree of formalization, and a corresponding high level of application of logic and automatic reasoning;
- But the commitment itself, and with it the semantics of the terminology as used in NL environments, cannot be made subject of logic and reasoning by virtue of the ontology;
- Committed term use is characteristic trait of domain ontologies and task ontologies;
- This entails constraints of many sorts, especially so with respect to consensuality among constructors and users, reuse and sharing, in interoperable applications, but also in the dedicated ones.

And linguistic ontologies?

Crystallized in the concept of "ontology" we find the universal idea of an explanation of what we find in this world "is about". Applied on phenomena of natural language and multilingualism, "linguistic ontologies" have emerged. The term suggests that distinction could be made with respect to "domain ontologies" and "task ontologies", and perhaps even that linguistics could be regarded as separate from semantics.

But this is confusing. "Evidently, domains exist. But what are the frontiers? What is a domain's specific language as opposed to the specific languages of other domains, and how does this relate to those parts of some general language authors would use anyway in their NL texts?[10] In ontologies (which should be representations of the world), admittedly, it is hard to separate linguistics from semantics, and if a distinction from "domain", or "task", or anything else was necessary, all what seems justified would be to define "linguistic ontologies" as dealing with light, formal linguistics, i. e. with lexical and grammatical units and structures (words, syntax, spelling) irrespective of specific domains (i. e. being general), with respect to

[9] SIGEL, A. 2003: *Topic Maps* **[18]** 411
[10] SCHMITZ-ESSER. W. 2005: *Ontology-based understanding*, 35

a specific NL only (i.e. on strictly monolingual terms). Following this logic, but also for reasons of practical feasibility in a multi-lingual set-up, supplementing the ontology with just light, formal linguistics would be enough. This, then, would be some piece of lingware, one pack for each language implemented, not a separate ontology.

In contrast, the few known "linguistic" ontological endeavors like SENSUS (Knight and Luk, 1994), MIKROKOSMOS (Mahesh, 1996), CYC Ontology (Lenat, 1995)[11] all exhibit one characteristic feature: semantics are interwoven with linguistics, whereby in pluri-lingual offerings the linguistic structure is drawn from English as "source" language. This spells a handicap in many respects.

Looking at linguistic representation of the concept, Gerhard Rahmstorf, the proposer of Concepto[12], can be seen as being closest to what the ICLO embodies. This comes not as a surprise because his work on nominal phrases and semantic relations[13] as well as his later contributions to the discussions in the German Classification and Thesaurus Committee[14] triggered and much inspired work on new forms of linguistic representation, - from advanced multi-lingual thesaurus to ICLO. Driven by practical needs that came up in multilingual project work, I subsequently widened his early idea of concept representation over pure nominality, introduced the onomasiologic principle of construction, included instantiations in the model, and took provision for equal handling of relations for both concepts and cases (instantiations) alike.[15]

These days, a "general linguistically motivated ontology for interfacing between … computational components …and natural language technology components for natural dialog" is being developed by John Bateman's Bremen Ontology Research Group[16]. Their goal is "to construct a set of interrelated general common sense ontology modules suitable for all areas of representation, but focussing particularly on spatial representations and information relevant for robotic movement in space"[17]. Although the linguistic features in this project (some reaching back to the CYC development[18]) differ much from those of the ICLO approach, the mainstream of the underlying philosophy can be shared.

Position of the ICLO

In the sense explained above, the Integrative Cross-Language Ontology – ICLO is neither "domain" nor "linguistic". It is general, semantic, and cross-language. To address a distinct meaning, the ICLO features the expressions that are used by authors of authoritative texts in the different languages but inasmuch as linguistics is concerned counts only on linguistic software in the indicated, much more restricted sense.

For every language implemented on the ontology, a different linguistic software or supplementary ontology of this type is required. If no corresponding software is available for a specific language implemented (which is no far-fetched assumption), the ICLO nevertheless would work well as to its semantic definitions, logic, reasoning and consistency except on

---

[11] see the review by HOVY, E. 2002: *Comparing* **[34]**
[12] RAHMSTORF, G. 2000: *Begriffsdarstellung* **[35]**
[13] RAHMSTORF, G. 1983: *Die semantischen Relationen* **[36]**
[14] for background and results see SCHMITZ_ESSER, W. 1999: *Thesaurus and Beyond* **[9]**
[15] SCHMITZ-ESSER, W. 2005: *More information* **[38]**
[16] a member of ECOR, the European Centre for Ontological Research. BATEMAN, J. A.2006: *Bremen* **[30]**1.
[17] ibid
[18] CyCorp 2006: *CYC* **[32]**

functions of text mining which then would be less efficient. Seen under the constraint above, the condition of the ICLO in fact is that of a true general and multilingual ontology.

Situated between the two poles of present day domain/task commitment and an un-mastered linguistic complexity, the ICLO addresses some of the most acute shortcomings in ontological engineering. It shows a natural, language-based way of inter-relating the different ontological archipelagos that have emerged. It helps to foster communication and interoperability on matters of general, principled knowledge organization.

NL is the only possible platform for world-wide sharing of knowledge

So, what are the conditions for world-wide sharing of textual knowledge in a multilingual world?

1. Solutions, if there will be some at all, are only imaginable on the basis of natural language (NL);
2. It is that kind of NL which in the different communities is used in written documents for the settling of their respective affairs;
3. In an integrative ontology, distinction between NL as used in special domains and NL in general must be possible (with a view to segmentation, merging) but should not be a condition for efficient operation. Frontiers between what should be "domain" and what is "general" will always be fuzzy;
4. What to the community "affairs of importance" are in this context is found reflected and manifest in the community's respective textual messages. Ideally, these must be regarded as the only potential authentic knowledge sources with respect to both semantic content and their related NL expressions. They have to be validated and exploited in a skilful, knowledgeable way. It must be claimed that this whole (mainly intellect driven) process must be thoroughly documented for reasons of transparence, clearness of expression, and editorial responsibility;
5. If a piece of knowledge to be expressed in the ontology needs NL and NL is not all univocal, but univocal expression is a prerequisite for the formalization of the knowledge and thus for logic manipulation and automatic reasoning, then only such NL expressions contained in the validated source texts can be admitted to formalization which unmistakably represent the knowledge (the meaning).
6. For every single piece of knowledge, this requires as many representations of equal quality as there are languages implemented in the ontology.

Logic, mathematics, and automatic reasoning

Thus, logic and automatic reasoning can only be applied on the univocal part of linguistic representation. But even then the range of functions applicable to arguments, relations and statements would have to be restricted, or newly defined, e. g. with respect to potential logical functions resulting from rather soft or equivocal meaning of some semantic relations such as the following:

> *ENHANCING may mean enhancing what IS_BENEFICIAL_TO as well as enhancing what IS_DETRIMENTAL_TO).*

To judge this, and to perform the selection, an expertly trained intellectual human resource as well as ergonomic, high-level machine assistance (MA) are required. In a given operation, the respective MA will recur on the entire, instantly updated knowledge accumulated in the ontology including the respective results of steady internal processing and updating. This includes the full range of spell and consistency checks, propositional logic, and parts, at least, of first-order logic – FOL.

How is high level representational quality achieved? The role of the Ontological Descriptor (OntoDesc)

The core of the ICLO, the Basic Semantic Reference Structure – BSRS, works (reasons) on abstract knowledge which is modelled as formalized semantics of their respective NL expressions. The crucial condition is that admitted to this function are only those NL expressions that possess true proxy capabilities. Selected to function in the model as sole and exclusive denominators, they "represent" the meaning of their respective NL expressions, in a clear and univocal way. They work independently from any other context that otherwise may be needed to explain or to identify a distinct meaning.

In the present paper, NL expressions (propositions) chosen to function as knowledge proxies of a concept in a particular language be called "ontological descriptors" (OntoDesc), or just DESC.

> Remark: The trope of a "descriptor" is drawn from classic thesaurus construction, although its functionality and formation rules are slightly different. Even in classic thesauri, the descriptor does not really "describe" the meaning; however, designed to function as a preferential expression, the descriptor in a thesaurus stands for the intended meaning, and insofar it is justified to consider the descriptor as the "proxy" for the meaning.

The role of equivalent Additional Access Expressions (AAE)

Expressions with equal or equivalent meaning can be attributed to the DESC to function as Additional Access Expressions – AAE. They provide improved access to the ICLO, or to NL texts browsed in searches. The underlying multilingual structure is shown in ***Fig 1a,*** and a corresponding example on a given concept is presented in ***Fig 1b.***

The role of expressions with multiple meaning (MULTI)

Expressions with still other meanings can also be attributed to this DESC, but must be attributed also to the corresponding other DESCs, accordingly. We call them MULTI - Expressions with Multiple Meaning. These are polysems in a broad sense. The ICLO provides a suitable mechanism that exploits the ontology's knowledge of what a MULTI is, featuring two functions:

1. Expressions of the multiple meaning type (MULTI) are automatically suppressed in Boolean searches. In normal search situations, i. e. when striving for precision, a MULTI spoils the search results. But when the search is out for recall, a polysem sometimes may be a last resort. To cope with such queries, the ICLO, enables MULTIs to be included in the search strings, but this then must be triggered on purpose.
2. Expressions with multiple meaning widen the view and stir imagination. So, MULTIs enable the ontology to provide search word assistance, guidance, orientation, and other useful services, like playful learning. They help in constructing what artist LOWRY BURGESS for the Hannover World EXPO 2000 visitor system proposed as "Thought Space Travel" [3].

On lower levels of systems definition (classic information retrieval), given DESCs, together with their respective AAEs, are used for browsing/querying unknown texts sources, irrespective of their origin (e.g. controlled/heterogeneous), and quality (e. g. raw, enhanced, warranted). Query quality is considerably enhanced (as to both recall *and* precision). Boolean searches on MULTIs are possible on purpose.

Seen from an ICLO user's perspective, DESCs and all equivalent NL expressions "known" by the ontology as AAE are accepted for:

- Access to, and use and reuse of, the totality of the BSRS knowledge, including its generated knowledge parts. MULTIs can be used on purpose upon request;
- The cross-language facility which means: Query in language A, answer in language B or any other language implemented on the ICLO.

The underlying structure, now with MULTIs included, is shown in ***Fig 2a***, and a corresponding example on a given concept is presented in ***Fig 2b***.

ICLO: Defining the meaning by relations, not annotations
In the BSRS of the ICLO, the OntoDescs function as arguments – Arg - under the PSSH assumption. No other than equally formalized definitions of the argument are admitted. This enables the ICLO to dispense with how-to-use instructions, descriptive definitions and footnotes of the usual, non-formalized (and hardly formalizeable) kind. In the ICLO, seven out of nine qualifiers of an argument will be made up as formalized definitions right from the beginning instead.

In the ICLO, the meaning of an expression (argument) will be defined by semantic relations which in valid source texts are found to exist between relevant objects of thought. They are extracted by the peers or curators, and their occurrence, the wording, source and other background, along with the hows and whys of the choices performed will be thoroughly documented.

Many different semantic relations may occur, but research on semantic relations and practical experience in applying them on broader scale have shown that remarkable precision and expressiveness can be achieved even on very small sets of the most frequently used "standard" relations **[9].** A four-language semantic web for an Internet user guidance system (SERUBA "Lex4", 15,000 nodes) actually did well with less than a dozen of different semantic relations **[5].**

For the ICLO, in the present paper, a canon of 23 different semantic relations, structured in five groups, is tentatively proposed and defined. They form what in this paper is called the "Arsenal". Users are free to define and use other, more specialized semantic relations in formal conformance with those of the Arsenal as need for them arises.

Fig. 3 shows how the different semantic relations are linked to the NL proxies of the meaning which are found united on the Equivalence Chain of Descriptors – ECD. Each ECD in its turn has an ID which in the model presented either is a Meta Language Identification Number – MLIN symbolizing a category (concept, topic), or the Meta Name Identification Number – MNIN as a symbol for the instance of a category (concept, topic) defined. Except this distinction, both categories and instantiations of categories are dealt with on symbol level exactly the same way. The result is the ICLO multi-lingual semantic net. ***Fig. 4***

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

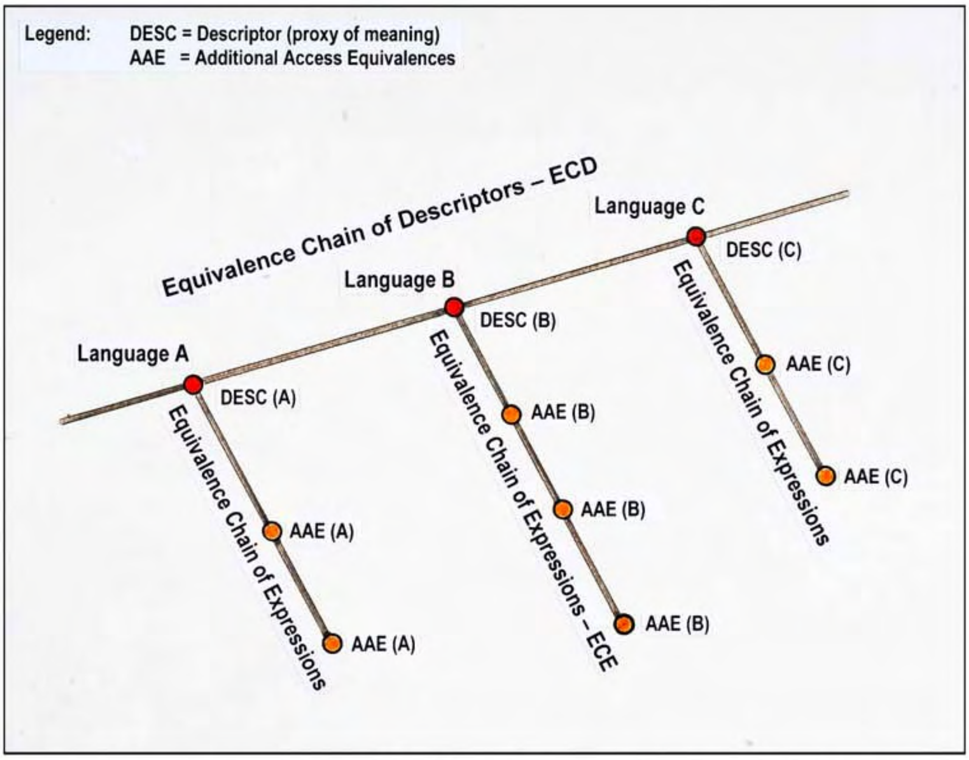


Fig. 1a. Structure of mono- and multilingual equivalences of a proposition (argument)

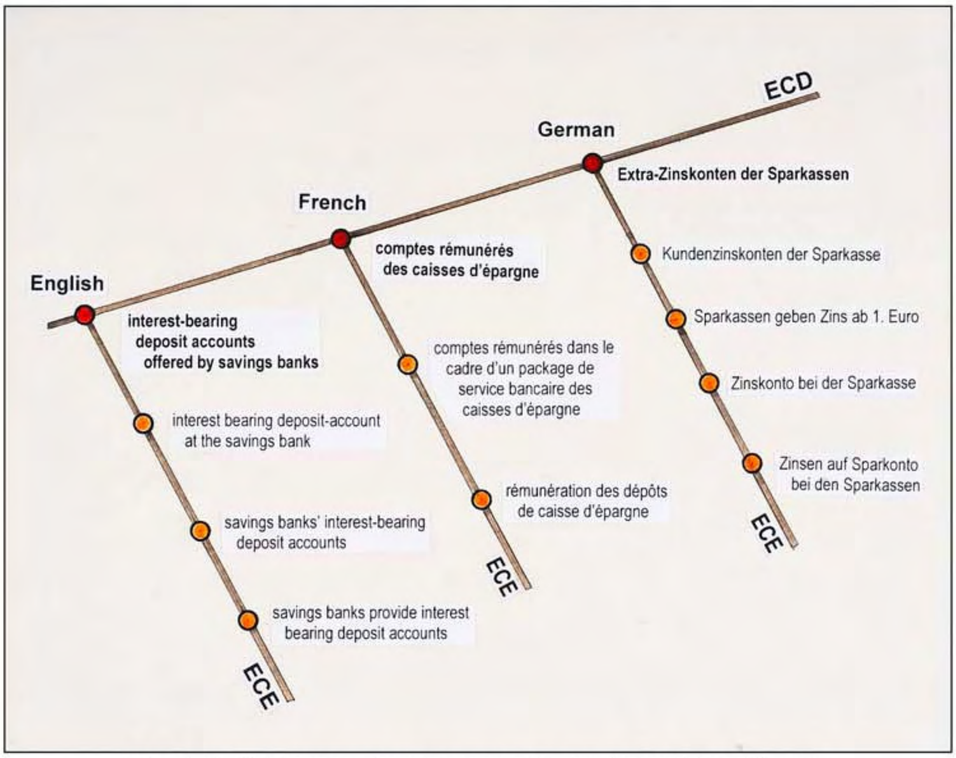


Fig. 1b. Equivalence structure of argument: Savings banks` interest-bearing deposit accounts in three languages

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

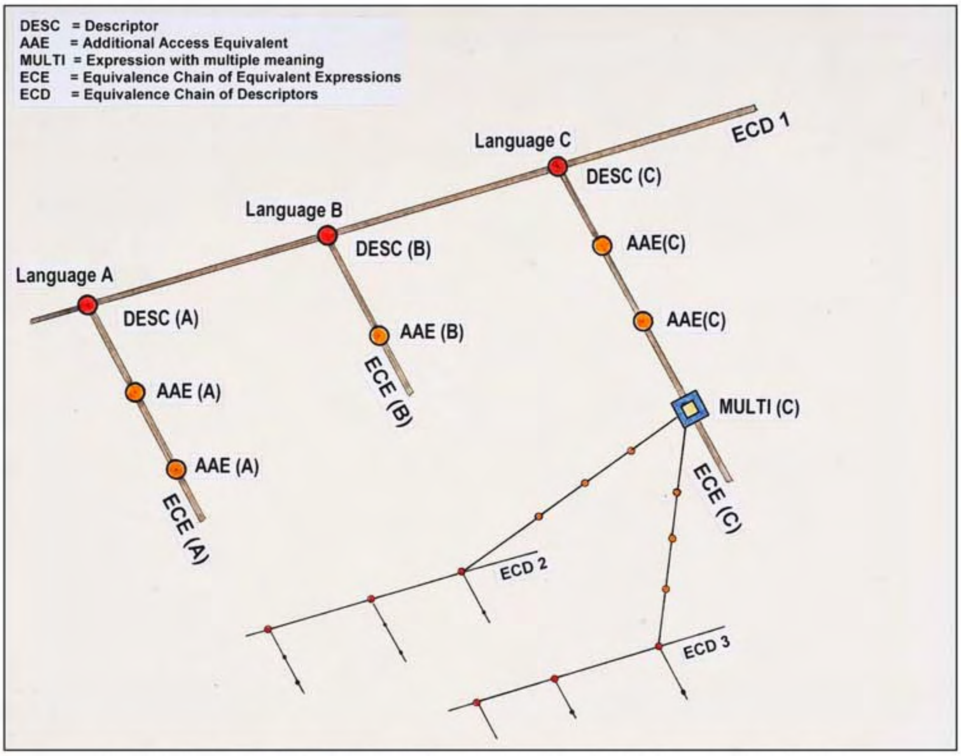


Fig. 2a. An expression with more than one meaning (MULTI) is member of all its corresponding Equivalence Chains

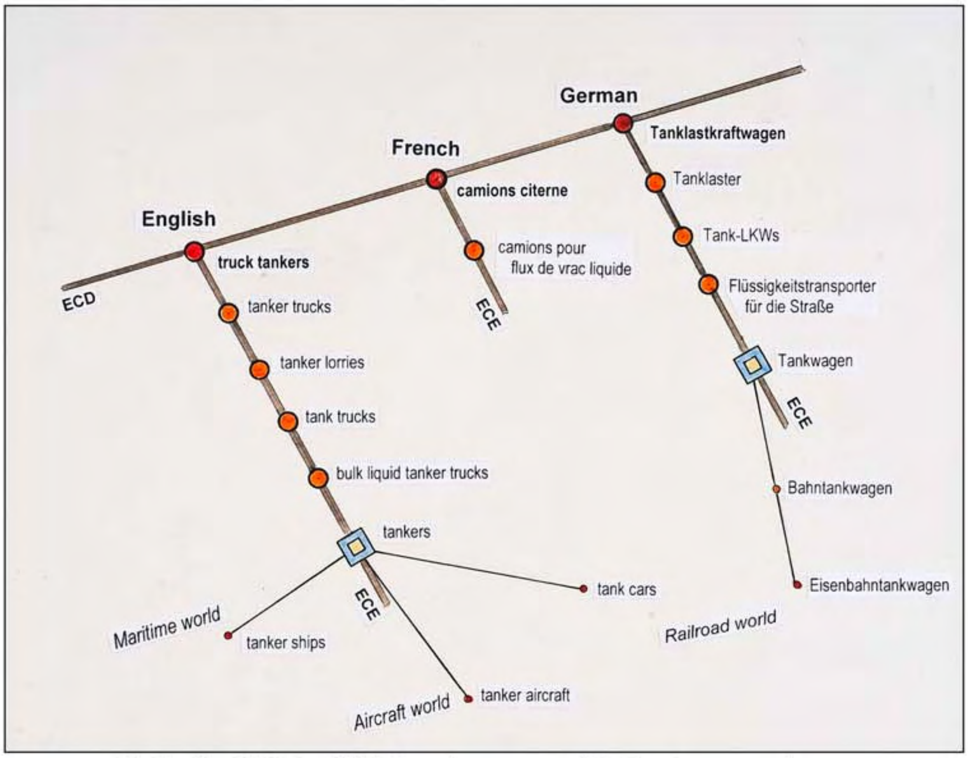


Fig. 2b. Like AAE, the MULTI for a given concept differ from language to language

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

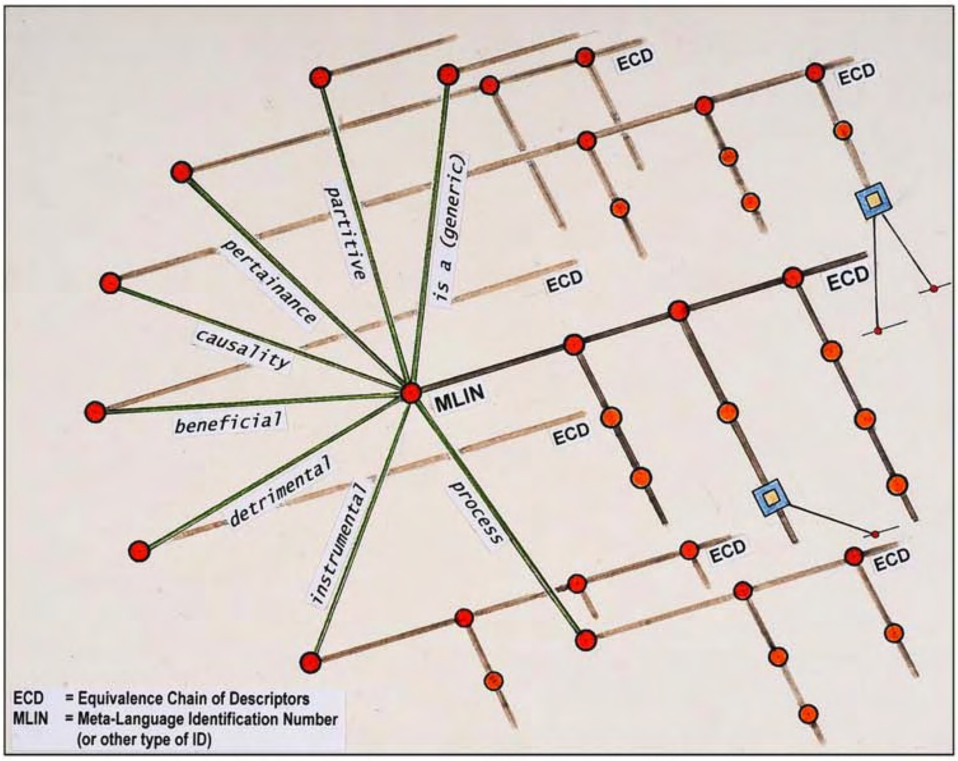


Fig. 3. The MLIN stands for its respective ECD. Semantic relations of the „essential", pan-language type link other MLIN to form the multi-lingual semantic net.

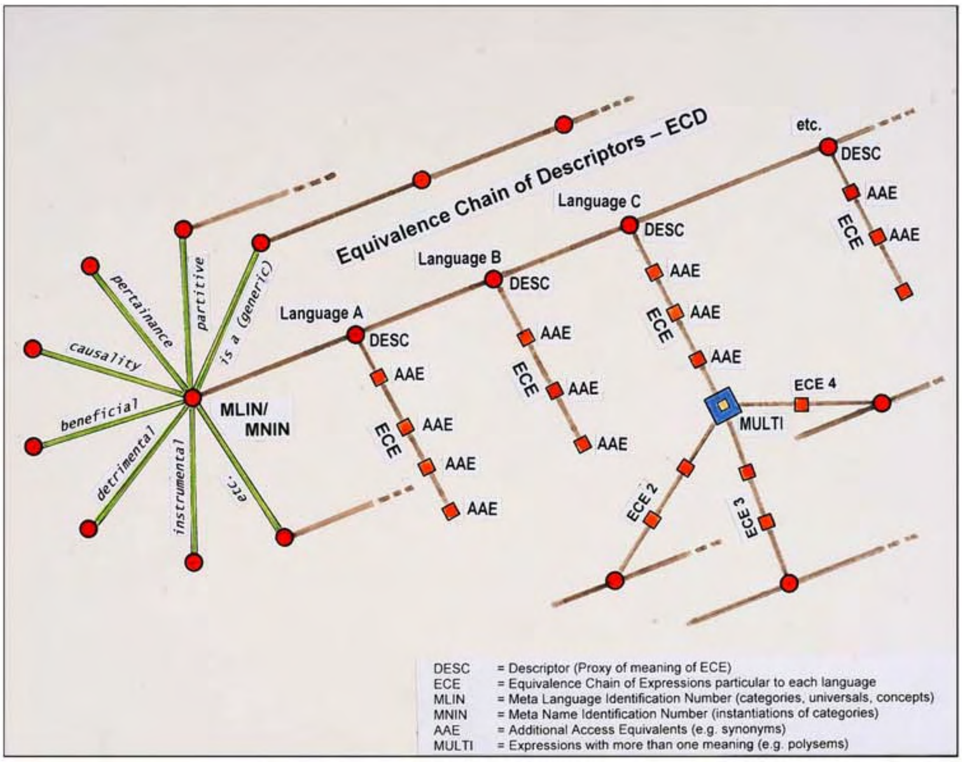


Fig. 4. The Multilingual Semantic Net

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Advantages and other salient points

Defining the semantics of the propositions (arguments) by means of formally standardized semantic relations has many advantages over present day proposition description, above all:

1. Only the meaning of a few dozens of semantic relations needs to be defined, known and applied by users, peers and constructors, not the meaning of thousands and thousands of expressions traditionally stipulated in term lists, classifications, thesauri and similar works, possibly in more than one language, for use on distinct single topics, and with no chance of ever getting formalized in a universally acceptable way.
2. Such relatively few relations, in turn, once defined, can be grouped in logical order to form hyper-structures of relational categories. This largely improves flexibility to react to lesser expressive needs, renders stipulation/codification more foreseeable while meeting mnemonic demands to the benefit of occasional, or non-expert, users.
3. Only such relations need to be stipulated that actually occur as of relevance in valid source texts.
4. Consensuality on all stages of use, reuse, construction and maintenance can be considerably improved. This is important in many respects, especially so as a prerequisite for warranting the consistency of the statements built by means of the semantic relations.
5. Defining what things "are about" by means of semantic relations truly mirrors views and relevance reasons that otherwise are difficult to express, but are of prime importance for reuse and sharing. A long-standing problem of present-time KO works and systems can thus be solved in a most natural way.
6. Shortcomings and deficiencies in KO works arising from amalgamation of existing terminologies such as reported on the NCI Thesaurus by CEUSTERS, W. et al. (2005): *A terminological* **[10]** could be referenced to a corresponding ICLO for analysis in the light of real world NL practice the ICLO would mirror. This would considerably extend the actual platform available for correction and all sorts of other improvements, including re-engineering.
7. For re-engineering, see the interesting AGROVOC Thesaurus case, where some semantic relationships are applied that also are contained in the ICLO Arsenal **[21].**

Multilingualism, barrier for the interchange of knowledge

In multi-lingual environments, committed term use, a comparatively humble requirement, is even more difficult to achieve. Lexical translation seldom renders the desired same meaning and rather seems to engender further commitment and further restrictions in the respective target languages. Commitment of term use no doubt tends to foster monolingual KO applications – among which English since long is the *de facto* language of convenience. Most of to-day's known ontologies are made up in English.

Seen from a broader perspective, the price for this is high. Commitment spells a particularly hard-shelled obstacle for effective, inter-lingual interchange of knowledge and KO works and devices on international and global level. Yields, e.g., obtained in the many English language ontologies from the formalization of English terms or term sequences cannot be earned on Spanish or French textual knowledge sources. The extent and magnitude of the ensuing bias is difficult to estimate but can hardly be overstated.

The formalization of knowledge as key to new, fresh, and consolidated knowledge has become a major driving force behind many of the recent achievements in IT, AI and information science, as well as for science and R&D in general. Advanced text mining is such an achievement. Here, the language barrier has not been overcome, yet. It tenaciously persists.

Knowledge contained in billions of textual sources written in other languages than English (e.g. patient records, dissertations, media content analyses), tendentiously remains undetected – for lack of suitable formalized cross-language knowledge instruments. As a result, the corresponding wealth of knowledge, written in languages other than that of the ontology, has practically no chance to gain any bearing on the actual course of science and cognition.

General, non-committed, multi-lingual ontologies like the ICLO could bring the change about. ***Figs. 5a and b*** show how an argument drawn from everyday life is dealt with in the multi-lingual environment to tell a whole story.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

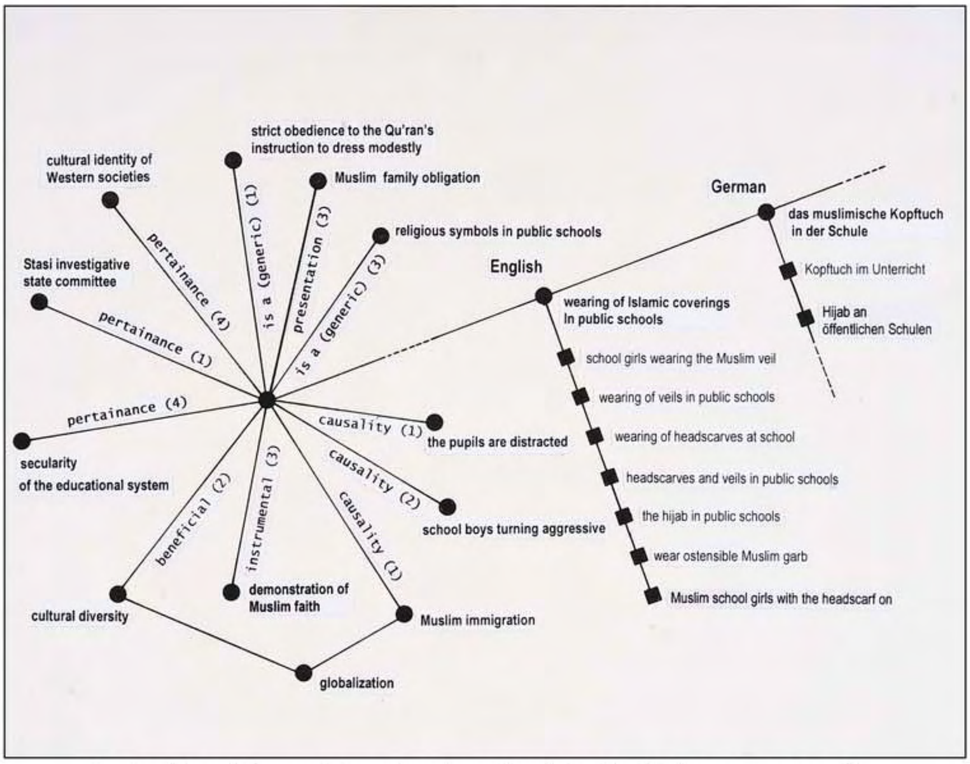


Fig. 5a. The ontology explains a French problem to the English language community

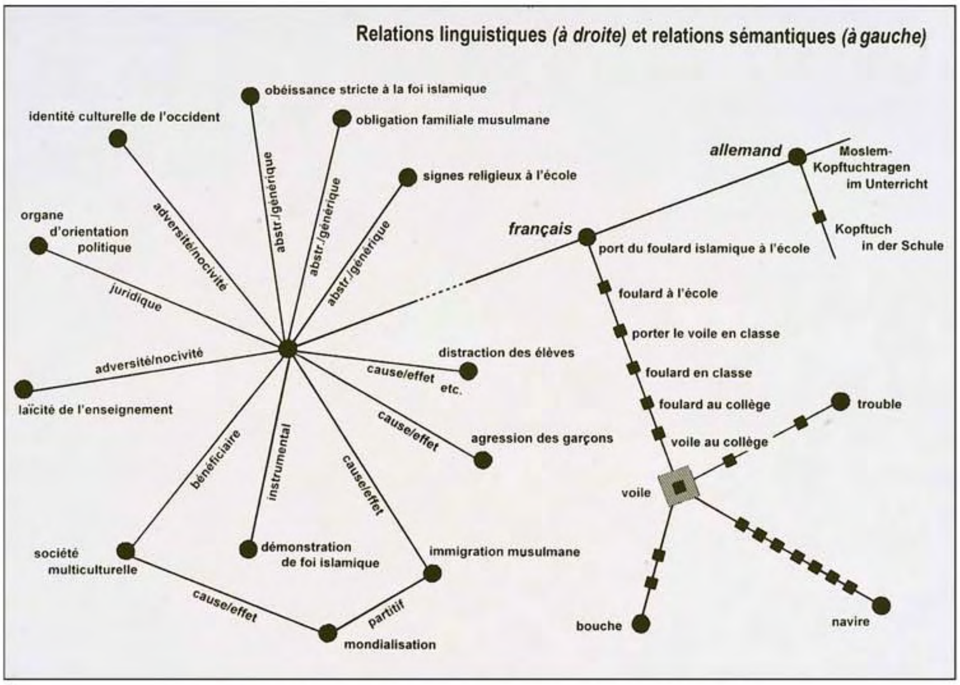


Fig. 5b. The French problem in French o-tone and perception

Summary of Outline

The ICLO is designed as a terminological model of the world's reality or of parts of it. Its terminology is drawn from authors' perceptions of the modelled (concepts and instances of concepts) according to the onomasiological principle (from thought to symbol). The model is based on NL, though not that of a particular language. The inner core of the ICLO, the Basic Semantic Reference Structure - BSRS, functions under classical PSSH. This enables logic and reasoning to be applied, and it allows its cross-language capability to work without intermediary language along with corresponding advantages.

The ICLO dispenses with the usual non-formalized descriptions and definitions of propositions. Semantics, instead, are modelled by means of just a rough two dozens of well-defined types of universal semantic relationships that are found to exist between concepts and their instances. Options for more specialized relationships exist. They can be made up formally the same way as the canonical ones of the "Arsenal".

The ICLO starts out with an aim to enable world-wide sharing of formalized knowledge on a broad NL basis, with NL input from validated sources, and with warranted, human-understandable output. It is designed as a means to open up the World's existing wealth of knowledge and knowledge sources that has grown in decades and centuries up to the present day and is found organized in a multitude of different forms and formats. It is hoped that the ICLO will facilitate world-wide cross-language access to these sources for reuse and sharing, less restricted by domain or other traditional barriers. Especially so, it should help in defining the semantic coordinates of existing KO works such as classifications, taxonomies, thesauri, domain and task ontologies, on the ICLO's general semantic net.

The assumption is defended that potentially any modern language can be implemented on the ICLO. However, under the best imaginable conditions, the ICLO will never be complete, never be perfect. It embodies just a new, fresh assault on a goal that hitherto seemed totally out of reach.

Not discussed here:
A number of further questions arising from the ICLO proposition presented cannot be discussed in this paper. Worth an extra study would be each of them: Organizational set-up of the ontology, human resource and training required, ontology languages, IT models and platforms, phases of construction and implementation, editorial responsibility and authors' rights, life cycle (of data, semantic relations, application-related meta-data, etc.), warranty of service, social and political impact, just to name what I find on the way, - and to stop here.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

PART 2

ICLO rules and principles of construction (RC)
Rationale in detail and organizational frame

RC 1

In the ICLO, knowledge is modelled by means of statements that are formalized. Each formalized statement S consists of two arguments A1, A2 which are linked by means of a Relational Operator R

S = R1(A1, A2).

This way, a binary (dyadic) relationship is expressed.

The idea that any two objects, if of interest, can be supposed to be linked by some corresponding semantic relation, leads to what in this paper is called a *statement* of knowledge. The statement then can be formalized in a universal way. This basic supposition determines the internal structure, organizational frame, and axiomatics, of the ICLO **[2]** 36.

Rationale: It is the relations between the objects that make us understand the world. Why? Because asking for a relation is the most natural way to fetch an understandable answer. Every relation coming along with an unknown object can be used as a question for this and other objects of equal nature, and objects retrieved can be semantically identified by means of one or more different relations with which we find them semantically linked.

There are different ways to express such relationships. A recent guideline proposes a combined RDF and Topic Maps expression in which the arguments assume roles ***[25].*** *In the following outline, the binary relationship is presented in a more NL-like form, for shortness, and ease of understanding:*

A1 – R1 – A2
short circuit – Causes = IS/ARE_CAUSE_OF – blackout

whereby the inverse view (IS/ARE_CAUSED_BY) be always tacitly included ***(Fig. 6)****. In the ontologic model presented, the way R1 is expressed in a given NL be called the Rface. Standing for something that "cause****s****" let us name the "active" face.*

Presentation of the relationships R used in the present working paper such as
IS/ARE_CAUSE_OF / IS/ARE_CAUSED_BY
are artificial, symbolic constructs needed to address them internally, not meant for use in searches on NL text surfaces. Each of them stands for a bunch of different NL expressions with equal or near-equal meaning known to occur in texts of the respective languages implemented on the ontology, and used for knowledge inference (e.g. in NL text browsing), and in the generation of ontology output in NL terms ***(Fig. 7).***

Also for reasons of simplicity, we dispense with conforming singular/plural forms of the different relational expressions in this paper ***(Fig. 8).*** *Putting it the right way, in the right language, is supposed to be performed by corresponding lingware in future applications*

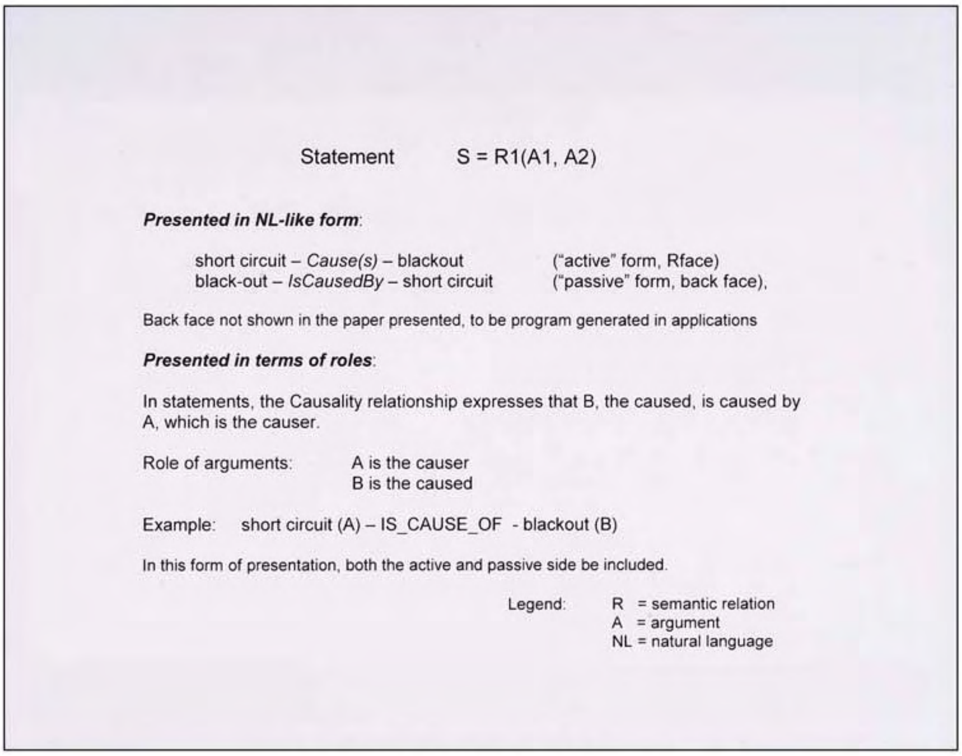


Fig. 6. Forms of presentation of Statement

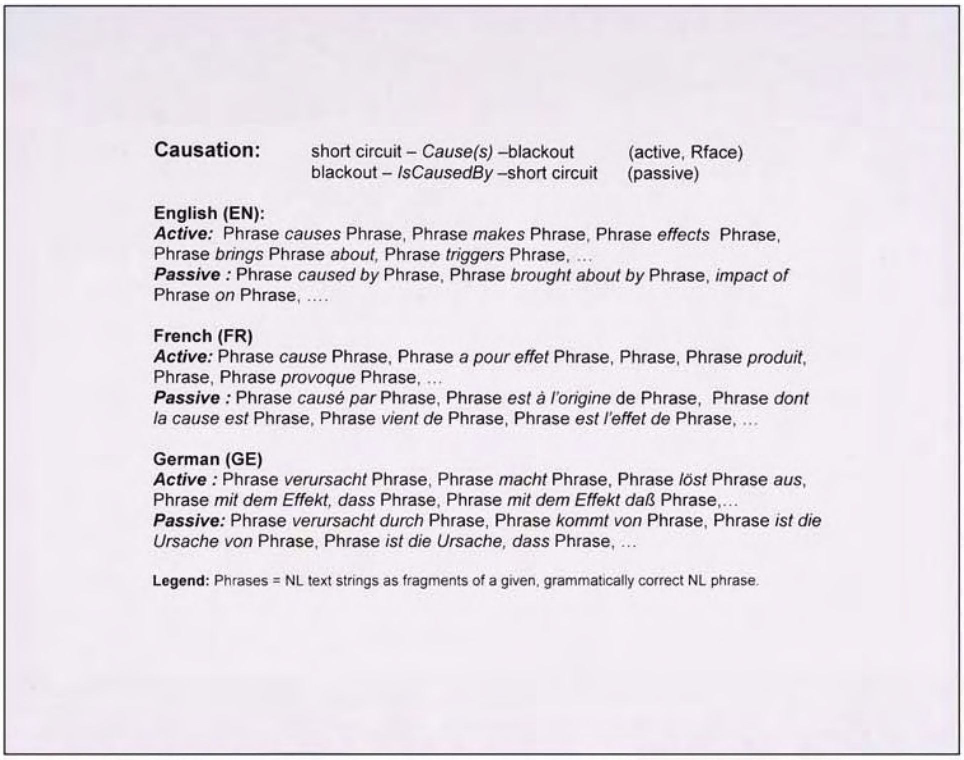


Fig. 7.How „Causation" potentially may appear embedded in an English, French, and German NL sentence (simplified).

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

EN
sing: short circuit – *Is/AreCauseOf* - blackout
plural: government bail-outs – *Is/AreInstrumentalFor* – rescue of private corporate entities

FR
sing : court-circuit – *Is/AreCauseOf* – panne d'électricité
plural : facteurs de croissance IGF – *enhance* – procès métastatique

GE
sing: Kurzschluss – *Is/AreCause of* – Stromausfall
plural: Truppenübungen – *Is/AreDetrimentalTo* - Naturbiotope

- *"Is/AreCauseOf"* as EN language Rface of the above relations is recommended for internal use on all languages implemented in the ontology. This will simplify coding and reading, and add to consensuality among the different user/language communities.
- But if required for reasons of political correctness, the Rfaces may likewise be presented for use in another national language implemented in the ICLO.

Fig. 8. Use of EN language Rfaces is recommended

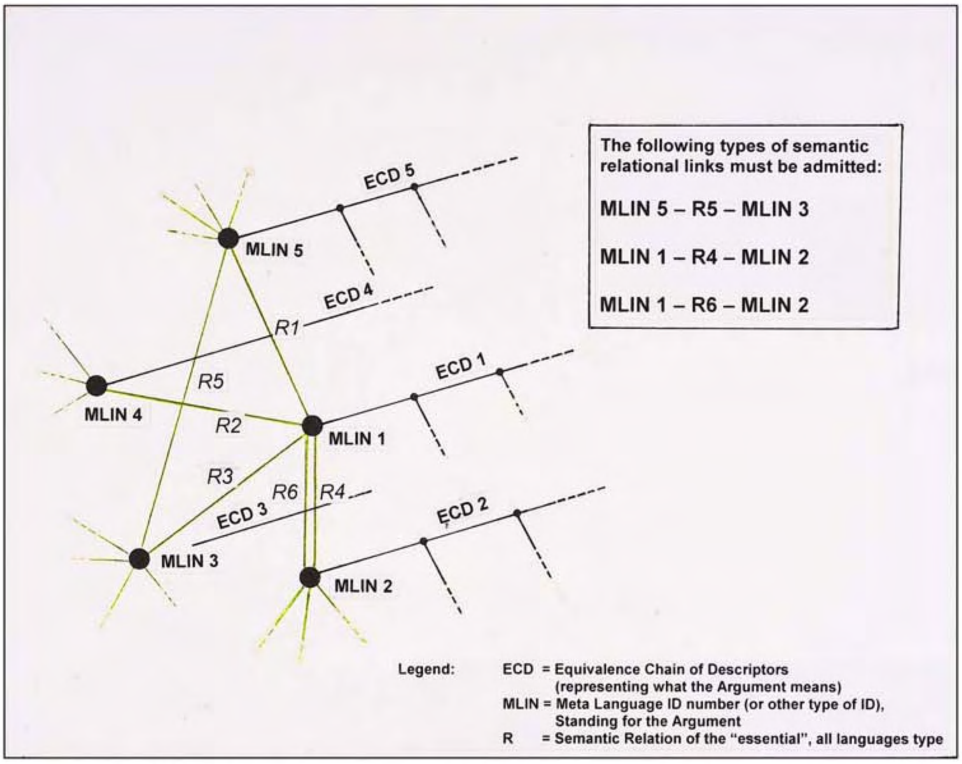


Fig. 9. „At least one other argument and a relation". No argument is allowed to stand alone.

RC 2

Semantic relations are directional and statements built with them insofar can be regarded as vectors. Some of them may be bi-directional.

That an object can be modelled as playing an active or a passive role in a given relation is widely accepted practice in ontological engineering; mathematical conclusions can be drawn on the basis of the underlying axiom and insofar this spells progress. But to model reality this is not enough. Objects sometimes assume both roles, the active and the passive one, in a given relation. If A1 talks to (IS_MESSENGER_TO) A2 while A2 talks to (IS_MESSENGER_TO) A1, then their relation IS_IN_CROSSTALK_WITH each other, thus reciprocal. In

progesterone – IS_IN_CROSSTALK_WITH – insulin-like growth factors

both arguments play the role of a talker as well as a listener. The ICLO must cope with reality, thus also with bi-directional relations.

Reciprocal relations are accepted and treated as reciprocal pair of their respective mono-directional relations.

RC 3

Each argument in the ontology is related to at least one other argument by means of one semantic relation out of the authoritative canon of admitted, formalized semantic relations (the "Arsenal"). The result is a semantic net (Fig. 9).

Scopenotes, bylines, comments, explanations, how-to-apply instructions for term-based arguments are standard instruments in KO up to the present day. They may be necessary and valuable as sources of knowledge in their proper contexts but are of no use in formalized, semantic ontological modelling.

- Since, in the ICLO, the meaning of an argument is defined by its relations with at least one other argument, arguments without semantic link are considered erratic and have no place in the ontology.
- The onomasiologic principle requires that any relationship between any two arguments be based on some content of relevance encountered in a valid NL source text.
- A relation found to be false or obsolete has to be dropped for this reason whereby the related argument, if orphaned by the drop, is for drop as well.
- Consequently, bracketed statements affected by a drop have to be revised.

Rationale:

In the ICLO, the defined, semantic relation to any one other argument plays an important part in asserting and elucidating the semantics of the argument. Other than bylines, annotations and comments, semantic relations can be used directly in logic and other ontological engineering functions.

Since the onomasiologic principle requires that the relationship between any two arguments must be based on, and extracted from, a well documented utterance in a valid NL source text, a correspondingly high validity of the model can be expected.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 4

Statements requiring a more complex structure may be built by brackets to form nested statements. Each binary statement and each bracketed statement is given a separate Statement ID - StID.

StID 4995 {[StID 9299 progesterone – IS_IN_CROSSTALK_WITH – insulin-like growth factors]- IS_IN - breast cancer cells}

Rationale:
Main pillar of expressiveness.

- One main handicap in modern ontological modelling is restricted expressiveness resulting from removal, or gross neglect, of linguistic features in most cases.
- Few terminology-based ontologies have been built so far on the basis of grammatical units bigger than words.
- To cope with real world terminology encountered in heterogeneous NL source and target texts, much higher expressiveness in an ontology is required. This implies that the ontology must be drawn up to deal with much higher complexity.
- Tremendous demand is felt for solutions on expressions with higher complexity.
- A trade-off with robustness, however, remains inevitable for reasons of practical feasibility. Not all relations are based on just two arguments. In this first ICLO proposal, ternary, and still finer tuned, non-reducible relations remain unconsidered.
- Ontologies should model shared understanding. "Since understanding is open-ended, it is a requirement that we can also represent incomplete or partial knowledge".[19].

With RC4, the ICLO is in a position to better conform with this requirement. The mechanism based on steadily growing knowledge on unwanted term worlds which the ontology will acquire while growing (see RC12, MULTIs) can be expected to also produce the rules for passing from single word expressions to more complex ones.

RC 5

Different binary or nested binary statements can be linked by a Joint Identification number (or other type of ID) – JtID to form an even more complex Chain of Statements – CSt.

Equally deplored is a lack of means to put the different statements together in a more constructivist, thus more predictable way, independently of the different languages, and on universal level. Encoding should come closer to universally acceptable synthesization and construction.

The solution is to have an open number of binary or nested binary statements linked together. In the ICLO, this could be done by means of a common Joint ID number (or other ID) – JtID which then gives an even more complex Chain of Statements – CSt.

JtID 2247	StID 5643	[European Commission, Brussels – IS_OPPOSED_TO[20] – sectorial aid]
JtID 2247	StID 2111	[government bail-outs for corporate entities – IS – sectorial aid]
JtID 2247	StID 8823	{StID 1777 government bail-outs for corporate entities - IS_INSTRUMENT_IN – [StID2222 Ministère des Finances, Paris – IS_SUPPORTER_OF – Alstom S.A]}

This way, even more complex Chains of Statements – CSt can be built.

[19] SIGEL, A.2003: *Topic Maps* **[18]** 399
[20] the internal way of expression from which the user surface will exhibit the respective Rface plain language equivalent in the different languages: "oppose(s)" (EN), "s'oppose(nt) à" (FR), "ist(sind) gegen" (GE).

Rationale, comment, and extensions:
The Chain of Statements – CSt is a unique and formidable instrument, that, in the above case, and along with other statements like

Ministère des Finances, Paris – IS_LEGAL_BODY_OF – French government
French government – IS_LEGAL_PART_OF – European Union
European Commission – IS_LEGAL_PART_OF – European Union

would enable the ontology to draw conclusions such as that the French support to Alstom "IS_INSTANCE_OF" sectorial aid, that because of this, the European Commission must, or at least should be[21] expected to be "IS_OPPOSED_TO" - the grant of a government bail-out to Alstom.

Implemented over the years, exceptions from a political rule thus could be traced and spotted. Possible consequences relating to the real world modelled, could then be highlighted or used for further reasoning, e. g. exceptions and trespasses of the rules, or dangers that might be implied in JtID2247, like future tacit nationalization of Alstom,. Other cases of sectorial aid given by the French government or by any another country could be traced, etc., etc.

Provided the semantic net is adequately filled, NL prompts could be given to questions like: "What alternative instruments are there that EU official economic policy disposes of to help endangered industries?", "What is the difference between sectorial and horizontal aid?", "What are the advantages of horizontal aid over sectorial aid, as in terms of beneficial effects that can be expected?", etc., etc.

RC 6

The construction of the ICLO should be performed according to the onomasiologic principle the strictest possible way. This applies not only to the verbal expressions sought in authors' texts for distinct ideas, but also for the underlying relations.

Only relationships encountered in valid, real world NL source texts should ideally be chosen as prototypes for being modelled in the ICLO.

Sometimes, however, this is not possible. The ontological construct then needs explicit inference of general knowledge. This calls for source text validation, and raises the question of responsible choice and documentation.

A NL text considered as a valid source is supposed to also reveal and render the right words that express a meaning of relevance for the purpose of the ontology. This complies with, and emanates from, the onomasiologic principle. In the ICLO, it is the onomasiologic principle of construction that is of force or at least prevails. Only words and relationships used by authors to express a distinct idea shall ideally be chosen as stuff to model the prototype – our perception of the real world - in the ICLO. They are encountered in valid, real world NL source texts.

- Sometimes, however, this is not possible, e. g. if the ontological construct needs explicit inference of general knowledge or of knowledge tacitly contained in sources. This calls for source and source text validation, and raises the question of responsible choice,

[21] For different modes of reality to be modelled in the ICLO, see RC 21.

inference, and documentation, in short: responsible editorship, and choices performed to be made transparent.

- But even the most valid source text may be found not free from tacit knowledge or knowledge about its respective environment that is necessary to be known by man ***and*** machine in order to get the full meaning of the message. So, some essential parts of this tacit knowledge also need to be entered in the ontology.
- Such knowledge environment may be contextual, linguistic, situational or other. It is for this reason why valid source texts must still be validated for the purpose of knowledge extraction.

  If, e. g. in a German text it reads: "Arbeitsamtumfrage" (labour administration investigation), the term to be extracted to serve as a proxy (DESC) must be changed into "Umfrage ***des*** Arbeitsamts" (investigation ***by*** the labour administration), whereas "Arbeitsamtumfrage" has to be kept as AAE, since this much more current expression, of course, is indispensable in text browsing.

- As seen from a reuser of the information in the ICLO, he/she must be given a facility, open at any time, of looking into the whys and hows of the intellectual choices that are behind the intellectual processes of source validating, knowledge extraction, stipulating and formalizing the knowledge, in the ontologic model.
- This calls for rigorous documentation of every source text and corresponding validated text from which a stipulated relationship is drawn. This resource must be kept open as a separate file for instant, uncomplicated access by constructors, curators and users on all stages of their intellectual work with and on the ontology, and this facility must be maintained and guaranteed at any given moment of the ICLO's life-time [6]. Instant, easy, reliable and sustained documentation of all intellectual processes and motives behind decisions in knowledge organization is a long-standing postulate.

RC 7

Making tacit knowledge about relationships explicit is admitted when seen necessary with a view to automatic processes of logic and calculus. Ideally, this process should also be based on equally reliable, well documented NL textual sources.

Comments:

Formalizing tacit knowledge that must be made explicit is a great theme in today's ontological engineering.

As a rule, the aim of such explicitation is to support logic reasoning and automatic calculus on the content of the entries.

- In the ICLO, the potential for application of logic and automatic calculus, however, is limited. The ICLO is no heavyweight among the known ontologies. "The semantic matter is continuous and moving, and the differentials are gradual", justly states LAZARD[22]. On soft, floppy ground like this, equality of meaning, the basis of reasoning, is difficult to state and define. Which parts of first order logic are applicable must be left for intense discussion in detail. Suffice it to say here that transitivity, e.g., should be restricted to taxonomic isness, as well as for the Arsenal's canonical group of partitive relations. This is for precaution, since the respective class criteria must remain undefined (impossible to generalize). It should be stressed that the ICLO in its unconventional approach, rather focuses on its many types of semantic relations which with their still softer, fuzzier inter-relationships open up new fields of automatic and MA exploitation, possibly beyond logic and reasoning.

---

[22] LAZARD, G. 1992: *Y a-t-il des catégories* **[17]** 431

- Explicitation of tacit knowledge, therefore, is not expected to play a pre-eminent role here. Experience tells us that one can well do with the assumption that where tacit knowledge occurs, this same knowledge can be supposed to be present and active also on the side of the user. This way, much effort in explicitation, and ensuing spill of means, can be avoided.
- Rather on the contrary, experience gained on practical applications has shown that unlimited explicitation can spell an enormous intellectual temptation. In real-world applications, it tends to absorb valuable human and machine resources, even opens the door to idiosyncratic use of terminology. Why? Corresponding textual sources for knowledge that is tacit are difficult to trace and spot.
- Stipulating and thus formalizing a relationship in a responsible way to make it work safely in a complex linguistic systems environment like the ICLO is a rather expensive act.
- However, doing this once and for many more cases of reuse and sharing surely is worth the greater effort, all the more as a potentially world-wide audience on the W3 alone is sure.

Rationale:

- The advantage over traditional methods of knowledge organization is obvious. Just think of indexing and classification, where a substantial effort also is invested - in single (possibly short-lived) documents.

RC 8

An argument either stands for a category or for a case (of a category). In the ontological model, the descriptor functions as the sole, authoritative representative (proxy) of the argument.

Comments:

The distinction between concepts and instances is a well-known feature in ontologies, although by no means in all of them **[12], [14].**

- Concepts also come under the denomination of *category, universal, subject*, *class,* sometimes also *topic*, or *theme*. Boundaries are fluent. Categories and classes are often seen as emanations of a general human perception of the world, explained by equal human experience in equal or similar cases (as with, say, "typhoons").
- Individual cases (instances) are modelled as instantiations of their respective categories or classes. They are defined as singular events in space and time, and normally addressed by means of an individual name. Thus, the name of an individual case in the category, or class of typhoons may be "Cathrina".

Frontiers:

It must be admitted, however, that what has to be regarded as an instance is all but crystal clear. Singularity in space and time sounds good but is fuzzy in dimensions of Star Wars and Big Bang. Are Rock'n Roll, Paris, the River Nile, singular in space and time, and thus instances?

One sees that the link to a respective category and Upper Ontology Qualifier (UOQ) is essential, but we have to make up choices that are down-to-earth. In the ICLO, one such choice was to stipulate states as universals (not instances), and to mark states as a special category inferred from an Upper Ontology.

Comments:

- Instances are sometimes seen as instances of a letter or a word in a text, mostly counted for the purpose of probability research. Let us here put this special understanding aside.
- In the ICLO, knowledge comes in first line from semantics. Here, an instance is understood as standing in a defined relation to a specific concept or category of which it is seen a singular case in space and time. This makes three crucial differences:
    - The instance in the ICLO is (a) independent of any particular document in which it "occurs". (b) If it is found modelled in the ontology, it "exists" only in relation to its respective concept (category), and (c) its existence and state in the model depends on ever changing states or parameters – whether the "case" is still up and alive, has changed or grown older, whether a record is worth to be held in the ontology and for how long, whether actions of the instance, and reactions once modelled are obsolete, and, if yes, whether they are for deletion or must be kept in memory, etc., etc.

Considering instances in such a way puts a heavy charge, and much responsibility, on editors and managers of the ontology.

There is a tremendous amount of movement in instances while concepts (categories) use to enjoy a relative diachronic stability **[6].**

RC 9

A descriptor standing for a category can take any length as in characters or words that may be required for any of the languages implemented in the ICLO, on condition that:

> ***[1] the cognitive content (Chomsky) of the argument is expressed the shortest possible but univocal way, whereby***
> ***[2] the expressiveness offered by the ICLO's canon of relations (which is independent of a particular language) must be exhaustively exploited, and***
> ***[3] no subordinate sentences can be tolerated.***

Note: NL phrases chosen as descriptors not necessarily need be nominal phrases in a strict grammatical sense.

Deliberation:

- That descriptors don't get too long should remain under reasonable control by condition [3] in first line, but also by [2]. The "shortest possible" recommendation of condition [1] works in the same direction.
- Condition [2] should work as a device to support the ICLO's constructivist potential, especially so since in multilingual environments it could be proved that in some way it "channels" , thus avoids, double implementation in the different languages.

Conditions:

This new liberty does not dispense with internal terminology control. It is not that type of terminology control known from classic thesaurus construction, which means forcing indexers and users to use a preferential term instead of a free one. It is terminology control in a much more advanced sense: Knowing at any given moment of the ontology's life cycle whether - and how - a given, free term is "known" by, and used in, the ontology. On the actual use of terms, the ICLO offers utmost, instant transparency. And such transparency must be offered to anybody dealing with the ontology at any time.

How can this be achieved?

By means of what in library practice is known as Keyword-in-Context register –KWIC. Thanks to advanced technology, it now comes in a much more usable, practical, revamped form: the KWIC (Key-Word-In-Context) Scroll Lens. The entries are shown scrolled and magnified as if under a magnifying glass. The lens was proposed for the visitor system of the EXPO 2000 World Exhibition in Hanover **[3].** Two samples of how it works on different term strings with equal meaning is given in ***Fig. 10 and 11***.

The KWIC should be available in four versions: (1) OntoDesc only, (b) AAE, (c) MULTI, (d) consolidated, with corresponding status marks. Such powerful, ergonomic alphabetical term control comprising every word in a descriptor line would have to be established for each language implemented in the ontology.

Comments:
For long, length restrictions have been forcing practice into a jungle of uniterm type polysems along with still other heavy constraints and provisions. On technical level, meanwhile, fetters of the like are tending to become obsolete.

1. To-day, concepts can be used in a more NL like way, which is the way we find them in the source texts. For stipulation in the ICLO, e.g., a concept like "*wear of woman's Islamic headdress at school",* need not be slashed into *wear of woman's Islamic headdress* and *schools*. Anyhow, "cognitive content" is neither *wear of woman's Islamic headdress* in general, nor *schools* alone. The full concept then opens ways of formalizing some much more specific enunciations, such as: [*wear of woman's Islamic headdress at school* – IS_IN_PLACE_OF – *Lycée Louis Pasteur]* – IS_IN_GEOPLACE_OF - *Paris*[23]. ***Figs. 5a and b.***
2. As a result, the construction (syntax) of the utterance in the ontology draws nearer to the syntax of the sources. This alone facilitates the question to be transposed into linguistic authenticity.
3. In short: Instead of improved term control for indexation of single texts the focus now is on better knowledge of the language itself, with a view to come forth with the closest possible linguistic approach to a text string that may appear in any text from the most heterogeneous sources.
4. As this comes with improved search string differentiation (for the "Terminological Tentacle", see CR 26 further below), the quality of results in searches can be markedly enhanced.
5. In Boolean searches, all this works as both a recall and a precision device.
6. The ICLO approach also spells a unique lift in potential explicitness. The ontology is now capable of better reflecting complex real-world problems, and change in word use can be monitored and adjusted on very short terms.

Rationale:
From improved term control in indexation of single texts, the focus now switches to better knowledge of the language itself. The aim is to get the closest possible linguistic approach to a text string standing for an argument or instance of argument that may appear in any text.

- This also means that the ICLO can do without format requirements such as the ones recommended for stipulation of descriptors in standard thesauri (e.g. the *nom. sing*. rule*)* **[28]**.
- For use in the wording of the DESCs, and to highlight the property of the descriptors as a class, the use of plural forms is rather recommended.

[23] SCHMITZ-ESSER, W. 2003: *Thésaurus* **[28]**

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Fig. 10. The KWIC-Scroll, a useful tool for the constructor, since not all in ICLO relates to a lexicon. Best shown on example in German

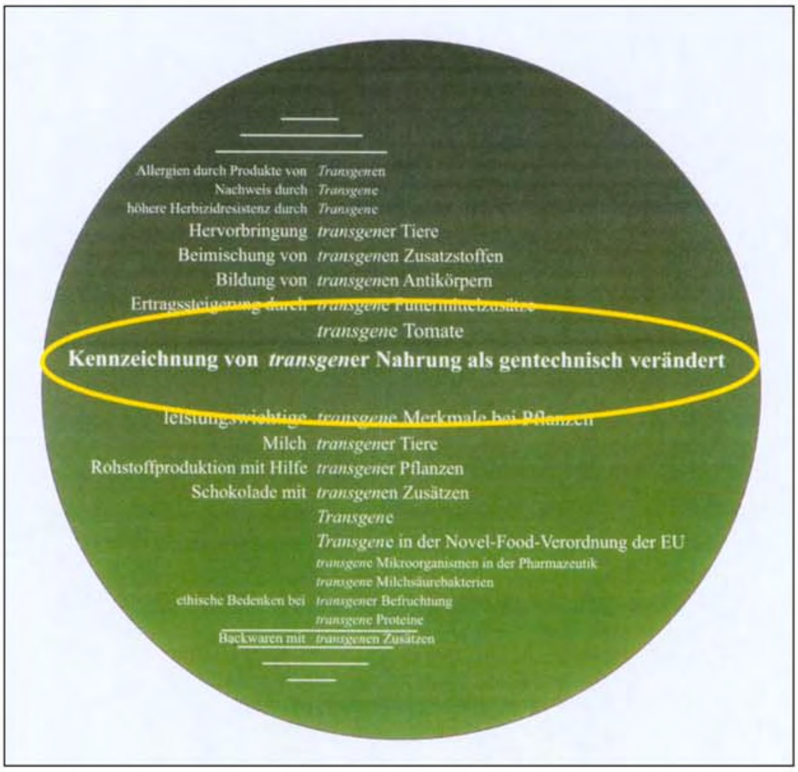


Fig. 11. The KWIC-Scroll gives a full overview on all words „known" by the ontology at any given moment of construction.

RC 10

In the ontological model, the descriptor (Onto-DESC) functions as the sole proxy among all expressions with equal meaning that are detected. They are stipulated and grouped on what in classic thesaurus construction is known as "synonym chain". In the ICLO, and for a given language implemented, this is the Equivalence Chain of Expressions – ECE.

Of course, it is not possible to relate the meaning of every term with that of every other term. The story is known from classic thesaurus construction: Among expressions with equivalent meaning, one preferential expression must be stipulated which is appointed the role of unique and authoritative representative of the meaning.

Comments:

- The descriptor is stipulated the sole and exclusive representative of the argument in the ICLO. The descriptor is selected by high-level, skilled manpower (peers/curators) so as to express the integral meaning of the argument in NL terms, in a clear and univocal way, independently from any context, be it term-wise, situational or other. Its sole function is that of an authoritative proxy of the semantics of the argument in the respective language **(*Fig. 2a*).**
- The trope of "descriptor" is drawn from classic thesaurus engineering, although its functionality and formation rules in the ICLO are slightly different.
- Even in classic thesauri, the descriptor does not really "describe" the semantic content. However, designed and stipulated to function as a preferential expression over expressions with equal meaning (ideally these should be synonyms), the descriptor is given the authoritative power of a proxy in a company of synonyms (lined up on the synonym chain). It is because of this known role that the expression is maintained for the ICLO. But to mark the little difference, the denomination of Onto-DESC is used for this type of descriptor, or just DESC in straight cases.

RC 11

The Equivalence Chain of Expressions (ECE) consists of a maximum of three types of members or groups of members:

> ***[1] The descriptor (Onto-DESC), which is obligatory;***
> ***[2] Equivalent expressions with no other meaning than that of the DESC (univocal equivalences). They are Additional Access Equivalents - AAE;***
> ***[3] Equivalent expressions with still other meanings (polysemic equivalences). These are the Equivalents with Multiple Meaning – MULTI [9].***

Comments:

1. Among expressions with same or equivalent meaning as that of the DESC it must be distinguished between two different types: The univocal and the polysemic type, AAE and MULTI, respectively.
2. Synonym chains of additional Access Equivalents – AAE – including the OntoDESC can be directly taken to function as search chains for programmed browsing in Boolean searches **(*Fig.12*)**. The single members of the chain then are linked by a (non exclusive) OR operator. This function is a most powerful recall device, welcome by both human researchers and agents.
3. Expression equivalents with multiple meaning – MULTI – then are the root of two or more synonym chains, each with a group of their own, their AAE and OntoDESC.

Corresponding chains, when defined, can be used for a BUT NOT operator in Boolean searches as a precision device ***(Fig. 2b, "unwanted worlds").***

In Boolean searches, precision is generally understood to be enhanced in unavoidable trade-off with recall, and vice-versa. If we accept the underlying definition of this standard iron paradigm (which is weak), our ontology enables it to be dismissed. The ICLO would enable to enhance both recall ***and*** precision.

- Among expressions with same or equivalent meaning, the ICLO distinguishes between two different types: univocal or polysemic. Let us call the univocal ones *Additional Access Expressions - AAE*, and the ones with multiple meaning *MULTI*. Then,
- an equivalence chain of the univocal type, including the OntoDESC itself, can be directly taken to function as search chain for programmed "OR" browsing in Boolean searches. This renders a powerful recall device.
- Expressions with more than one meaning lined up on that same equivalence chain then can be used as a precision device: By withdrawal of the MULTI from the search while excluding all terms on its other ECEs (that of the unwanted worlds) mediating the BUT NOT operator.
- In early Boolean approaches there was much ado about synonyms which under a more critical approach do not always stand the test. In all languages hardly any two words exist that have exactly the same meaning. Then the idea came up about quasi-synonymy. But this opened a bottle that was difficult to plug up. After quite some dispute, and with a view to practical application, practitioners now prefer to talk of "semantic equivalence".

But how far does equivalence reach when it comes to semantics? "When it turns into the false", says Margotti **[14],** thus paying tribute to Lazard's earlier observation that "the semantic matter is continuous and flexible, and the differences we encounter are gradual" **[15].** Here we find it: A robust, practical working position well suited for the ICLO.

RC 12

A descriptor standing for an instance consists of the full individual name of the instance plus a formalized addendum in plain words that unmistakably warrants the singularity of the instance as in space and time.

Ever since mankind emerged from history, the question was about naming individuals, acts and events. In a ubiquitiously available model of the world, names of what in this model is called "instances" are of prime importance.

- Rights of the individual and rights of the society are concerned and sometimes at stake. A mix-up of names could have fatal consequences in a real world knowledge machine.
- Other than with universals, the name of a person or a business is seldom spelled out in full in source texts. As a rule, this forces the ontology modeller to edit a special addendum to the DESC of a person, or company, etc., for completion which, as a rule, means that some supplementary research for this purpose is required. This may turn out as burdensome, even arduous, or impossible at all.

  Just to give an example: Find out the seat and constitution of firm X that offers illegal software on the internet. Or: Find out the name and domicile of a young, unknown pseudonymous writer!.
- A supplemented descriptor of this type and genesis cannot, of course, be expected to occur in regular source texts and therefore may be of little value in browsing. But clear-cut autonomous denomination of what is meant by the DESC has absolute priority. As to browsing, the currently much shorter names of the instance which normally are

encountered in valid source texts would be member(s) anyhow of the respective synonym chain (ECE), and no valid name in the string needs to be lost for the sake of browsing.

- But what with a name string extracted from a source text that would be mixed up with the name of somebody else because, speaking in terms of symbol sequence, the string is a MULTI? Both ECEs, with their respective DESCs and AAEs, would end up in the same MULTI. Then, in searches, there would be four options: [1] The MULTI is held back from a Boolean matching procedure, and search results rely on DESC and AAEs only. [2] The MULTI is held back from the search as before, and in addition, unwanted worlds of meaning as indicated by "alien" ECEs are excluded on purpose by activation of the Boolean AND NOT operator **(*Fig. 2b*).** Or, it may be felt that [3] the respective Main Type of Instances (MTI) could help in disambiguation. Finally [4], if the MTI does not hold, the Upper Ontology Qualifier (UOQ) which is derived from the instance's category may serve as a last resort.

  This way, e.g., a trade mark and a business having the same name and being kept apart in the ontology, could be distinguished and identified..
- If all four options fail to yield a fairly precise search result, there is no other way than falling back to normal Boolean search practice, throwing the MULTI in, and accepting the negative side of the recall.

Motion:
All over the world, there is rich experience in the art of identifying instances: in keeping media morgues, in handbook and dictionary writing, in the encyclopaedic business, in the archival, library and documentation business. Many ways and kinds of authentification exist in the different countries and in the different languages and could be used.[24] It would be most wishful if this wealth could be activated to bring forth some commonly accepted formula for addenda to the trope of an Onto-DESC for instances, along with recommendations of good practice in searches.

As a suggestion for short, labour-conscious but practicable addenda, an example each for the standard case of a person and the standard case of a business company in a real-world application, are given in ***Fig. 13 and 14.*** Both reflect experience gained in earlier projects of the author **[3] [14] [17].**

RC 13

Special rules must apply for the stipulation of the different categories of instances, the Main Types of Instances – MTI. This is an internal, ICLO standard proposal permitting still more adapted, special solutions.

As to this seemingly simple point, no international standards were seen on which the ICLO could lean or which it could apply. The background is a long standing, controversial matter, with the most divergent interests involved, calling for solutions on broad, international level, which obviously are difficult to attain.

Differentiation of different types of instances is all but theoretical. In fact, workable and useful differentiation is of greatest practical and economic importance. The discussion on standardization of just some different types, like businesses, shows that this is a real big issue.

[24] For persons, e. g. the PND (Personen-Norm-Datei) of The German Library (Die Deutsche Bibliothek). http://www.ddb.de/de/standardisierung/pdf/pnd_4.PDF

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Search string generated by ICLO for Boolean searches in English from the EN language ECE

Search for:
truck tankers OR tanker trucks OR tanker lorries OR tank trucks OR bulk liquid tanker trucks (sing. forms to be automatically added to the search chain by appropriate linguistic programme)

Omit:
"tankers", since this has multiple meaning, is a MULTI. Don't use Boolean "NOT" (would be counterproductive).

But INCLUDE :
"tankers", if search is for maximum recall. This then must be triggered by the searcher on purpose.

Fig. 12. The Equivalence Chain of Expressions - ECE in Boolean searches

Addenda in Argument set for DESC of a person: Gerhard R. Baum
(redundant presentation)

MNIN, VEx, ESI, UOQ, Pers, DESC, EN, Gerhard Rudolf Baum, German politician, FDP, born Dresden 1932-10-28
MNIN, VEx, ESI, UOQ, Pers, AAE, EN, Gerhard R. Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, EN, G. R. Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, EN, liberal Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, MULTI, EN, Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, MULTI, EN, Baum
MNIN, VEx, ESI, UOQ, Pers, DESC, FR, Gerhard Rudolf Baum, homme politique allemand du parti libéral allemand (FDP), né le 28 octobre1932 à Dresde
MNIN, VEx, ESI, UOQ, Pers, AAE, FR, ancien ministre allemand Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, FR, libéral allemand Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, FR, ex-ministre allemand Gérard Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, FR, Gerhard Baum, libéral allemand
MNIN, VEx, ESI, UOQ, Pers, MULTI, FR, Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, MULTI, FR, Baum
MNIN, VEx, ESI, UOQ, Pers, DESC, GE, Gerhard Rudolf Baum, deutscher Politiker (FDP), geb. 28. 10. 1932 in Dresden
MNIN, VEx, ESI, UOQ, Pers, AAE, GE, Innenminister Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, GE, Innenminister Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, GE, Bundesminister des Innern, Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, AAE, GE, FDP Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, MULTI, GE, Gerhard Baum
MNIN, VEx, ESI, UOQ, Pers, MULTI, GE, Baum

Legend: MNIN = Meta Name ID Number (indicating that it is an instance)
VEx = Entry on validity and expiration
ESI = External source Interface
UOQ = Upper Ontology Qualifier
Pers = Entry on a Person as Main Type of Instances – MTI

Fig. 13. Additional research is needed to avoid confusion with same names of other persons, or with concepts (like Baum = tree)

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Addenda in Argument set for a corporate entity: Telecom Italia
(redundant presentation)

MNIN, VEx, ESI, UOQ, Comp, DESC, EN, Telecom Italia, Milan, Piazza degli affari 2
MNIN, VEx, ESI, UOQ, Comp, AAE, EN, Telecom Italia
MNIN, VEx, ESI, UOQ, Comp, AAE, EN, Telecom Italy
MNIN, VEx, ESI, UOQ, Comp, AAE, EN, Italian Telecom
MNIN, VEx, ESI, UOQ, Comp, AAE, EN, Italy's Telecom
MNIN, VEx, ESI, UOQ, Comp, AAE, EN, Italy's National Telecom
MNIN, VEx, ESI, UOQ, Comp, AAE, EN, Italy's National Telecom
MNIN, VEx, ESI, UOQ, Comp, MULTI, EN, Telecom
MNIN, VEx, ESI, UOQ, Comp, DESC, FR, Telecom Italia, Milan, Piazza degli affari 2
MNIN, VEx, ESI, UOQ, Comp, AAE, FR, Télécom Italia
MNIN, VEx, ESI, UOQ, Comp, AAE, FR, Télécom nationale italienne
MNIN, VEx, ESI, UOQ, Comp, AAE, FR, Telecom Italia
MNIN, VEx, ESI, UOQ, Comp, MULTI, FR, Télécom
MNIN, VEx, ESI, UOQ, Comp, DESC, GE, Telecom Italia, Mailand, Piazza degli affari 2
MNIN, VEx, ESI, UOQ, Comp, AAE, GE, Telecom Italia
MNIN, VEx, ESI, UOQ, Comp, AAE, GE, Telecom Italien
MNIN, VEx, ESI, UOQ, Comp, AAE, GE, Italienische Telecom
MNIN, VEx, ESI, UOQ, Comp, AAE, GE, Italiens Telecom
MNIN, VEx, ESI, UOQ, Comp, AAE, GE, Italiens staatliche Telecomgesellschaft
MNIN, VEx, ESI, UOQ, Comp, MULTI, GE, Telecom

Legend: MNIN = Meta Name ID Number (indicating that it is an instance)
VE = Entry on validity and expiration
ES = External source Interface
UO = Upper Ontology Qualifier
Comp = Entry on a company as Main Type of Instances – MTI

Fig. 14. Additional research as needed for Instantiation „Telecom Italia" to be safely represented in the ontology.

[Main Types of arguments]

2. Main Type of Instances – MTI
(as used in an ontology project on general knowledge for deciders)

Pers Human beings
Begs Non-human beings
Comp Companies and other private groups, also when wholly or partly nationalized
Govt Governments and governmental organizations
NGO Non-government organizations
Prse Products and services
Proj Projects
Evts Events
Trmk Trademarks
Styles Styles

This is a simple example, but it is robust and almost intersection-free. Choosing the right types for a planned or given application is a task that should not be underestimated. Compatibility should be sought with existing, sound typefications that come nearest to the purpose of the application. The aspect of sustainability must not be lost of sight.

Thought might be given to the idea of introducing (sub-) types attributed to such robust main types which might be mandatory in more complex, sophisticated applications.

Fig. 15. Main Types of Instances - MTI

- Persons, bodies by law, informal groups, public programs and stratagems, events, titles, trade marks, just to name what may be the most important, all require a special way to be stipulated as descriptors in a correct, exhaustive, widely predictable and confusion-free manner.

On the level of the ICLO this calls for a committed scheme of Main Types of Instances – MTI which come in two versions, one for the instances and one for the categories. To have it as a standard, either some convention must be brought about if it does not emerge from de-facto acceptance. An example for main types required in a proprietary project on business deciders knowledge designed by the author is presented in ***Fig. 15.*** Formal conformance of implementation of the different types can be program enforced independently from general acceptance. Anyhow, reliability and consistency of the ontology would benefit from this.

RC 14

External resources can be linked to each canonical ICLO argument. These may be terms or term strings from third ontologies or other KO works, like descriptors from thesauri, DDC numbers, DUNs codes, or just URLs. The device for this is the ESI, the External Source Interface.

Remarks:

1. The ICLO's link to the outside world. To be exact: the link of a model widely controlled in terms of general semantics and terminology to a world of terms and meanings of heterogeneous descent, rules and parameters.

   That effective linking from ontology to external sources is viable on the basis of ESI constructions was studied and exemplified in various seminars held by the author in the past three years at the FB BUI of the University of Applied Sciences in Hamburg: on linking of DUN's codes, Dewey classification, Thesaurus Wirtschaft (HWWA Hamburg, and the Institut für Weltwirtschaft, Kiel).

- In Topic Maps, this could be modelled as occurrences or as (different relationship types.
- The challenge of defining a general, robust and flexible ESI still lies ahead, and a corresponding initiative needs to be taken. A prime point will be to take provision that the door is kept wide open for the extension of the resource base during operation as the ontology grows.

Rationale:

1. Integration of existing KO resources is the driving force behind the ICLO with its initial I in the banner that stands for integration.
2. Experience over almost 30 years has shown **[14] [3] [5]** that definition of the undefined even in large worlds of general knowledge can be achieved by relatively small segments of knowledge provided these are well defined and predictable, and that work on this can well be performed by different teams, and at different places, provided the rules for definition are tight and "rigor of encoding" **[12]** can be effectively enforced.
3. With its External Source Interface (ESI), the ICLO can be used as a work bench and semantic interpreter for other ontologies and related resources, like thesauri, authority files, classifications, search word lists. This is a most useful and promising facility derived from the ICLO, with immediate value.

   In the SERUBA project, e.g. a device similar to the ESI was used as an identifier of search words semantically related to "sex and crime" which in the offer of a TV and Internet service customer were required to be blended out during afternoon hours in order to protect kids. It was an easy exercise.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 15

An Upper Ontology Qualifier – UOQ is attributed to each argument. Instances carry the same UOQ as their category counterparts. (See Roberto Poli's subsequent paper which will also deal with the draft Upper Level Ontology discussed on the recent uos-convene[25]).

The UOQ asserts, underlines, and helps to discriminate the meaning of the formalized argument.

The UOQ substantiates the categorical character of an instance stipulated in the ontology. The instances being defined as singular cases of their respective categories, and no instantiation occurring without respective formal definition (which will be program enforced), the same UOQ applies for categories and instances alike.

The UOQ characterizes the argument in a highly abstract way. Applied in conjunction with the External Source Interface, the UOQ gives the ICLO the power of semantically defining knowledge organized in other ontologies and related knowledge order works and systems such as classifications, taxonomies, and thesauri, and therefore is capable of "judging" degrees of commitment, expressiveness, and compatibility, and of indicating ways for knowledge merging and integration.

The UOQ provides still other unique, powerful means.

1. As to the interpretation of unknown source texts and ensuing knowledge generation: Since the onomasiologic principle prevails in ontology construction, the human ontology constructor is in steady search for valid, authors' NL representatives of the meaning that are adequate, equivalent, new, and fitting.. Then, e. g., in capturing the intended meaning in an unknown source text, it would be important for him/her to know whether OUT OF JOB, as it stands in the ontology, is to be perceived as "process" or "determination", since this, in a search for NL equivalences, makes quite a difference. On the other hand, when producing a machine-generated answer, it is important for the machine to know whether a PROCESS is finished or still going on, or whether some past event A has IMPACT ON process B today, etc.
2. As to searches: It is foreseeable that in Boolean searches not all cases of multiple meaning (MULTI) can be cleared just by exclusion of un-wanted term worlds, all the less since these worlds cannot be expected to have been stipulated in full at any stage of development of the ICLO. The UOQ then may help in disambiguation, and even more sophisticated disambiguation procedures could be developed on the basis of the UOQ.
3. As to knowledge generation, still: The position of the UOQ in the system enables the ICLO to render the most appropriate words from an Equivalence Chain of Expressions (ECE) when it comes to the generation of knowledge by virtue of the ontology proper **[16].**
4. In implementing the ontology on practical level, the rigor required by the upper ontology resource helps to codify. The UOQ raises consensuality among divergent user groups, and since, as we all know, the ICLO is designed to be constructed, updated, used and reused on interoperable pluri-lingual level, consensuality spells a top priority issue.

[25] SMITH, B. 2006: *Beginnings* **[31]**

RC 16

Each language implemented in the ICLO requires an Equivalence Chain of Expressions (ECE) to be made up for each DESC, irrespective of whether the DESC represents a category or an instance. DESCs stipulated in the different languages for the same meaning form the Equivalence Chain of Descriptors – ECD. They share the same MLIN, or MNIN, respectively.

It can be assumed that considerable demand exists for cross-language knowledge organization – CLKO, although this may not come up to the surface in full extent for supposed or de-facto reasons of limited feasibility. Most of the very few ontology-based CL approaches known are derivatives from solutions on English, the predominant language, which hampers CL access, reuse, and acceptance in other language communities, and on top poses additional barriers, e. g. on grounds of political and other correctnesses.

But building an ontology apt to cope with one and the same meaning in different languages without classic translation is demanding in many respects, since it heavily relies not only on the principle, but also on practical feasibility of the onomasiologic approach.

1. This rule is a command of logic, but not easy to implement on onomasiologic conditions. It is hard to come to know which expressions for a given idea out for implementation at a given moment in time are in use in the target language, and it may be even harder to find the respective valid source texts in this language exactly at the time when this is needed[26].
2. New knowledge that comes up in language L1, along with its sources and new expressions (neologisms), often has no counterpart yet in language L2 and L3, neither as in pure lexical units nor in terms of the much more articulated and varied expressions encountered in valid textual sources.
3. Typically, there are phase shifts in the discussion of a specific problem in different language communities and/or countries. What to-day is discussed in U.S. papers may be taken up in German media only weeks later. Since the onomasiologic way of construction requires true and documented NL evidence also in German, the default of sources causes problems. Just a dictionary is not enough.
4. Another aspect is that work of this sort cannot be performed in a responsible way without sufficient knowledge and understanding of what the world segment to be modelled is about. Polyglot peers being specialists in their respective domains of knowledge are required.
5. The rise of world-wide communication stimulates the use of borrowed terms, especially so expressions in English. With its equivalence chain construct for every language, the ICLO nicely matches a requirement of this sort. Thus, an English term can well appear among (or instead of) a couple of German language expressions on an Equivalence Chain for the German language.
6. This construction not only enables to stipulate "Espíritu Santo" as in use in German papers for a Portuguese bank, but also to fix the German "Handy" as the German language equivalent of what in English is "mobile".

[26] ***Fig. 1b*** renders a nice example for this. The topic of "comptes rémunérés" recently came up in French papers during discussions in Parliament on new bank legislation permitting savings banks to offer their clients special interest-bearing deposit accounts as part of their regular banking services. The issue passed unnoticed in papers of England and Germany where such accounts since long exist.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 17

Whether also for categories there is a need to distinguish different types is an open question. This then would be the Main Types of Categories feature – MTC, the respective counterpart of the Main Types of Instances - MTI.

RC 18

An argument in the ICLO then has the following eight elements:

Element 1: ***MLIN*** or ***MNIN*** as ID number (or any other ID), depending on whether the argument is a category or an instance. If it is a category, the ID be called the Meta-Language Identification Number – MLIN, if instance Meta-Name Identification Number – MNIN.

Element 2: (void in the default case): ***VEx*** = Validity of argument expired on [*timeline*] but argument is kept as a historical feature

Element 3: ***ESI***, the External Source Interface, enabling the link to different sources not formalized under the ICLO.

Element 4: ***UOQ*** – Upper Ontology Qualifier. If the argument represents an instance, the UOQ would be attributed (by program) from its respective category.

Element 5: ***MTC*** (of a category) or ***MTI*** (of an instance), respectively. Main Types of Category are still be defined and could even be recommended for committed use. Main Types of Instances in a business application, e.g., under the ICLO may read as shown in **Fig 15.** They appear as (individual) names:

Pers	Human beings	*Prse*	Products and services
Begs	Non-human beings[27]	*Proj*	Projects
Comp	Companies and other private groups, also when wholly or partly nationalized	*Evts*	Events
Govt	Governments and governmental organizations	*Trmk*	Trademarks
NGO	Non-government organizations	*Styles*	Styles

Element 6: ***TECE*** – Type on ECE, the Equivalence Chain of Expressions. This may be: DESC if it is a Descriptor, AAE = Additional Access Equivalents of the descriptor (sole, univocal synonym or acronym), or MULTI = Expressions equivalent to the descriptor, but with multiple meaning (polysems or polysemic acronyms).

Element 7: ***LAc*** = Language Acronym for the language in which the NL term string is expressed. Used in this paper: EN for English, FR for French, GE for German.

Element 8: ***TStr*** = Term String, a single NL string of terms among an unlimited number of others in an argument set.

[27] See **RC13**

RC 19

Relations R in this ontology are semantic relations of a language-independent type, the „Essential Relations“, or “Class-2 relations”[28], whereas language-dependent semantic relations, the “Class-1 relations” are considered and represented in their respective ECEs. See *Fig. 4*: The Multilingual Semantic Net.

Distinction must be made between semantic relations that root in a given language (like synonyms, polysems, etc.) or whether they exist in all languages, independently of a particular language. Since almost all of to-day’s ontologies are made up in one language only, it doesn’t come as a surprise that a corresponding demand for distinction up to now is not really being felt. But the distinction is crucial in a multi-lingual set-up.

Note: **Both classes are semantic relations. In a multilingual environment they must be kept separate and dealt with in their respective, appropriate way.**

RC 20

In the ontological model, semantic equivalents of Class-2 relations require their respective Equivalence Chains of Relational Expressions – ECRE, one each only for each language.

Like an argument, a semantic relation can be found expressed in many ways. Nevertheless, onomasiological experience shows that the way our most vital semantic relations are found expressed in real-world texts is not totally unpredictable and that the number of these relations is small.

While widening and deepening the ontology, application of the onomasiologic principle on new, unknown texts would of course give the peers or curators an opportunity to also detect hitherto unknown expressions authors use for what in the ontology is a canonized semantic relation. The finding then would be included in the ECRE. This renders both best coverage of meaning for future searches and an ever more realistic, solid grounding for constant completion and update of the ontology.

So, it is in pursuit of the onomasiologic principle on source texts that on a given relation an ever more complete and authoritative Equivalence Chain of Relational Expressions – ECRE can be assembled. ***Fig 16*** shows this on argument “cell motility” harvested from the relationship of “Regulation”.

As many Equivalent Chains of Relational Expressions - ECRE are needed as are implemented in the ontology. In the example given this is English, French and German. Still another advantage comes with the ICLO’s multi-lingual set-up. No middle language (inter-lingua) is needed. There are just as many language pairs as languages are implemented. This has considerable advantages over the middle language approach.

RC 21

In the ICLO, each of the canonical relations comes in four different modes:
[1] assertive, [2] modal, [3] intention, [4] negation.

Comment: More modes exist, but these are the most important ones. The yield is much improved expressiveness. Also, the meaning is clear, and no further comment is needed.

[28] proposed by SCHMITZ-ESSER, W. 1999: *Thesaurus and Beyond* **[9]** 11f.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Argument set for a category (concept, universal, class)
Redundant presentation

MLIN, VEx, ESI, UOQ, MTC, DESC, EN, cell motility
MLIN, Vex, ESI, UOQ, MTC, AAE, EN, motility of cells
MLIN, VEx, ESI, UOQ, MTC, AAE, EN, cell mobility
MLIN, VEx, ESI, UOQ, MTC, AAE, EN, mobility of cells
MLIN, Vex, ESI, UOQ, MTC, AAE, EN, motility of the cells
MLIN, Vex, ESI, UOQ, MTC, AAE, EN, cells wandering
MLIN, Vex, ESI, UOQ, MTC, AAE, EN, cell migration
MLIN, VEx, ESI, UOQ, MTC, MULTI, EN, motility
MLIN, VEx, ESI, UOQ, MTC, MULTI, EN, mobility
MLIN, VEx, ESI, UOQ, MTC, DESC, FR, motilité cellulaire
MLIN, Vex, ESI, UOQ, MTC, AAE, FR, motilité des cellules
MLIN, Vex, ESI, UOQ, MTC, AAE, FR, mobilité cellulaire
MLIN, VEx, ESI, UOQ, MTC, AAE, FR, mobilité des cellules
MLIN, VEx, ESI, UOQ, MTC, AAE, FR, migration cellulaire
MLIN, VEx, ESI, UOQ, MTC, AAE, FR, migration des cellules
MLIN, VEx, ESI, UOQ, MTC, MULTI. FR, motilité
MLIN, Vex, ESI, UOQ, MTC, MULTI, FR, mobilité
MLIN, VEx, ESI, UOQ, MTC, DESC, GE, Zellmotilität
MLIN, Vex, ESI, UOQ, MTC, AAE, GE, Motilität der Zellen
MLIN, Vex, ESI, UOQ, MTC, AAE, GE, Motilität der Zellen
MLIN, Vex, ESI, UOQ, MTC, AAE, GE, Motilität bei den Zellen
MLIN, Vex, ESI, UOQ, MTC, AAE, GE, Mobilität der Zellen
MLIN, Vex, ESI, UOQ, MTC, AAE, GE, Zellmigration
MLIN, VEx, ESI, UOQ, MTC, AAE, GE, Zellwanderung
MLIN, VEx, ESI, UOQ, MTC, MULTI, GE, Motilität
MLIN, VEx, ESI, UOQ, MTC, MULTI, GE, Mobilität

Legend:
MLIN = Meta Language ID Number (indicating a category)
VEx = Entry on validity and expiration
ESI = External source Interface
UOQ = Upper Ontology Qualifier
MTC = Main Type of Category (optional)
DESC = Descriptor, proxy of respective AAE and MULTI
AAE = Additional Access Equivalents to the DESC
MULTI = Equivalent Expression but with still other meanings

Fig. 16. Argument „cell motility", the case of a category.

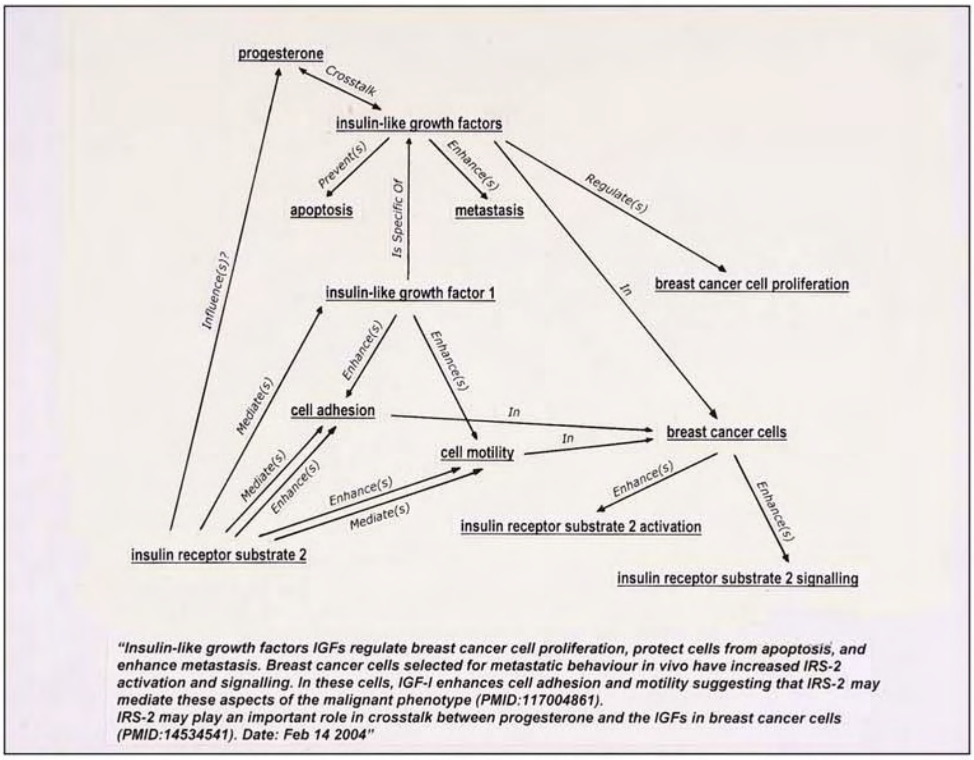


Fig. 17. Argument „cell motility" as modelled from knowledge contained in note on insulin receptor substrate 2 published in Cancer Gene Card.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 22

Two classes of essential relations are admitted for use in the ICLO: The defined, authoritative, or canonical ones (from the "Arsenal") and an open number of particular ones which may be defined along the same lines by interested users as need for them arises. Backed by research reaching down till the beginning of the nineties [9] and further work conducted with Roberto Poli, the following 23 defined Class-2 relations are proposed to be contained in the "Arsenal":

1. Instantiation	*A IS_INSTANCE_OF B*
2. "Is a" type of isness (Taxonomy)	*A IS B*[29]
3. Partitive	*A IS_PART_OF B*
3.1 Location (part/whole)	*A IS_IN_PLACE_OF B*
3.2 Geographic part/whole	*A IS_IN_GEOPLACE_OF B*
3.3 Pertainance (institutional p/wh)	*A IS_INSTPART_OF B*
3.4 Composition (matter)	*A IS_MATTER_OF B*
4. Causality	*A IS_CAUSE_OF B*
4.1 Regulation	*A IS_REGULATOR_OF B*
4.11 Enhancement	*A IS_ENHANCER_OF B*
4.111 Beneficial	*A IS_BENEFICIAL_FOR B*
4.112 Detrimental	*A IS_DETRIMENTAL_TO B*
4.12 Diminution	*A IS_DIMINISHER_OF B*
4.13 Prevention	*A IS_PREVENTER_OF B*
4.2 Instrumental	*A IS_INSTRUMENTAL_FOR B*
4.3 Process	*A IS_PROCESS_APPLIED_IN B*
4.31 Mediation	*A IS_MEDIATOR_IN B*
4.311 Communication	*A IS_MESSENGER_TO B*
4.312 Cooperation	*A IS_COOPERATING_WITH B*
4.313 Unification	*A IS_UNITING_WITH B*
4.4 Descendancy	*A IS_DESCENDANT_OF B*
5. Presentation (appearance)	*A IS_PRESENTED_AS B*
5.1 State/Form	*A IS_STATE/FORM_OF_B*

Examples, one for each relationship, mode options included:

1. Instantiation	Salman Rushdie, India-born writer – IS_INSTANCE_OF [1] – writers in exile
2. "Is a" type of isness	hibiscus – IS [1] – malvaceae
3. Partitive	missile boosters – IS_PART_OF [2] – missiles
3.1 Location	receptor activation – IS_IN_PLACE_OF [2] – breast cancer cells
3.2 Geographic p/whole	Urals – IS_IN_GEOPLACE_OF [1] – Russia
3.3 Pertainance p/whole	City of Westminster – IS_INSTPART_OF [1] – Greater London
3.4 Composition	Earth Core – IS_MATTER_OF [1] – iron
4. Causality	polygamy – IS_CAUSE_OF [2] – social unrest (Le Monde: riots by African immigrants in suburbs of French cities, 11/2005)
4.1 Regulation	thermostats – IS_REGULATOR_OF [1] – room heating
4.11 Enhancement	additional search words – IS_ENHANCER_OF [2] – searchability of crude texts
4.111 Beneficial	fish staircases – IS_BENEFICIAL_FOR [1] – protection of living marine resources

[29] The case "A is a B" is implied in the very definition of instance: The instance is a case of the category (class).

4.112 Detrimental	polygamy – IS_DETRIMENTAL_TO [1] – womans rights
4.12 Diminution	high protein food – IS_DIMINISHER_OF [2] – cardiovascular risk
4.13 Prevention	insulin-like growth factors IGF – IS_PREVENTER_OF [2] – apoptosis
4.2 Instrumental	raising the VAT – IS_INSTRUMENTAL_FOR [2] – cut in state deficit
4.3 Process	carved laquer technique – IS_PROCESS_APPLIED_IN [1] – antique Japanese jar production
4.31 Mediation	even platinum surface – IS_MEDIATOR_IN – exhaust gas purification
4.311 Communication	RNA – IS_MESSENGER_TO [1] – DNA
4.312 Cooperation	Sharp – IS_COOPERATING_WITH [1] – Fujitsu
4.313 Unification	Frankfurt Stock Exchange – IS_UNITING_WITH [4] – London Stock Exchange
4.4 Descendancy	negative film – IS_GENERATOR_OF [1] – film paper copies
5. Presentation	portals – IS_PRESENTED_AS [2] – Roman arcs
5.1 State/Form	steam – IS_STATE/FORM_OF [1] - water

Note:

- No transitivity is admitted except on
  (a) Relation 2, the *"Is a"* type of isness which is that of the classic taxonomy: *oaks IS trees IS plants.*
  (b) The family of the partitive relations 3 through 3.4
- Relations of the *4.31 Mediation family* may have a bi-directional character. Relations 4.312 Cooperation and 4.313 Unification (those with the "WITH") are always bi-directional. The system must adjust to both logic and linguistics as to conform with sententional logic which then could be taken over by appropriate extra lingware. So, in the case of Relation 4.311 Communication, an internal ECRE can be built on the basis of *"InCrosstalkWith"*.
- More specific relations may be defined by appliers as need for them arises.
- Formally they should in line with those of the Arsenal type, defined in such a way that they can be smartly ranged under one of the five main groups shown in **RC 22**.
- Such free-lance type, more specialized relations be called here ROSA – Relations made Operable for Special Applications. Typically, such ROSA would be PRODUCER – produces – PRODUCT, or OWNER – owns – OWNED. Examples of application of some ROSA, even on triadic relationships, like PROPOSER – *ProposesTheObjec*t – OBJECT_PROPOSED – *OnOfferTo* - BUYER may be found in **[8]**[30].

RC 23

Each R carries two time-lines: that of the validity and that of the entry (date of stipulation or update respectively).

This RC copes with a widely deplored deficiency, especially so encountered on URLs. Nobody will sincerely doubt that this is a most wishful feature. Modelled in Topic Maps, this type of entry could be feasible as "occurrence" or validity relation.

[30] SCHMITZ-ESSER, W. 2003: *Meaning, Understanding* **[8]** 161f.

RC 24

Each canonical relation is described and defined in terms of NL. Here comes the complete set of relations as listed in RC 22. They are seen expressed in the form required for implementation as Topic Maps - TM:

Lack of definition what a relation means is a major deficiency in most modern ontologies, spells built-in fuzzyness and results in a source of constant semantic confusion. It does not suffice to say that category A is lower in hierarchy than category B. It is necessary to say what an attributed position in a hierarchical relationship *means*. Thorough definition is crucial for all semantic relations applied.

RelClass2, Type 1 Instantiation

In ICLO statements, the Instantiation relationship expresses that argument B. the Category, concept, class, or topic [1] is, or [2] may be, or [3] is intended to be, or [4] is not that of the Instance A, which is the Instance.

Role of arguments:	A is the Instance B is the Category

Meaning of Instantiation as modelled in this ontology: The concept of the relationship between a state of affairs B, the Category, which, within the realm of application of this ontology, the value of being Instance [1] is, or [2] may be, or [3] is intended to be, or [4] is not, attributed to A, the Instance.

Example:	The Beatles (A)	IS_INSTANCE_OF[1]	Pop Music groups (B)
	Belgium (A)	IS_INSTANCE_OF[4]	States

RelClass2, Type 2 "Is a" type of isness (Taxonomy)

In ICLO statements, the "Is a" relationship expresses that argument B, the being [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, argument A, which is the specific being of argument A.

Role of arguments:	A is the specific being of B B is the being

Meaning of "Is a" type of isness as modelled in this ontology: The concept of the relationship between a state of affairs B, the being, which, within the hierarchical system of established human knowledge, [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, attributed a more specific state of affairs A, the specific being A, which from in B differs in at least one classic "*differentia specifica*" however, undefined in the ontology.

Example;	hibiscus (A)	IS{1]	malvaceae (B)
	killing of cats for experiments (A)	IS[3]	murder(B)
	whales (A)	IS[4]	fish[B]

RelClass2, Type 3 Partitive

In ICLO statements, the Partitive relationship expresses that argument B, which is the whole, [1] includes, or [2] may include, or [3] is intended (claimed) to include, or [4] does not include, argument A, the part (the included).

Role of arguments;	A is the part B is the whole

Meaning of Partitive as modelled in this ontology: The concept of the relationship between a state of affairs B, the whole (the including) B, which, within the hierarchical system of established human knowledge, [1] includes, or [2] may include, or [3] is intended (claimed) to include, or [4] does not include, the part (the included), which is A.

Examples:	derricks(A)	IS_PART_OF[1]	offshore drilling platforms
	peace offer	IS_PART_OF[2]	deliberations
	real estate	IS_PART_OF[3]	ancient heritage
	wheels	IS_PART_OF[4]	maglev vehicles

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RelClass2-Type 3.1 Location (part/whole)

In ICLO statements, the Location (part/whole) relationship expresses that argument B, which is the whole, [1] includes, or [2] may include, or [3] is intended (claimed) to include, or [4] does not include, argument A, the part (the included), as by location in the widest sense, except geographic location.

Role of arguments:	A is the part as by location
	B is the whole as by location

Meaning of Location (part/whole) in this ontology: The concept of the relationship between a state of affairs B, the whole as by location B, which, within the hierarchical system of established human knowledge, but with the exception of geographic location, [1] includes, or [2] may include, or [3] is intended (claimed) to include, or [4] does not include, the part (the included), which is A as by location.

Examples:	nanotech particles	IS_PLACE_OF[1]	intelligent laquer surfaces
	Cancer Cells	IS_IN_PLACE_OF[2]	lungs
	CYP1A2 ferment	IS_IN_PLACE_OF[4]	patient's liver

RelClass2 –Type 3.2 Geographic (part/whole)

In ICLO statements, the Geographic (part/whole) relationship expresses that argument B, which is the whole, [1] includes, or [2] may include, or [3] is intended (claimed) to include, or [4] does not include, argument A, the part (the included), as by geographic location on Earth and in space.

Role or arguments:	A is the geographic part
	B is the geographic whole

Meaning of Location (part/whole) in this ontology: The concept of the relationship between a state of affairs B, the whole as by geographic place, which, within the hierarchical system of established human knowledge, [1] includes, or [2] may include, or [3] is intended (claimed) to include, or [4] does not include, the part (the included), which is A as by location on Earth or in space.

Examples:	Rio Branco	IS_IN_PLACE_OF[1]	Amazon Basin
	Old Atlantis	IS_IN_PLACE_OF[3]	North Frisian coast

RelClass2-Type 3.3 Pertainance

In ICLO statements, the Pertainance relationship expresses that argument B, the pertaining, [1] pertains, or [2] may pertain, or [3] is intended to pertain, or [4] does not pertain, to argument A, which is the pertainee, as by law, jurisdiction, convention, established opinion or expectation, religion and belief.

Role of arguments:	A is the pertainee
	B is the pertaining

Meaning of Pertainance as modelled in this ontology: The concept of the relationship between a state of affairs B, the pertaining, which, within an established system of social values and postulates, [1] pertains to, or [2] may pertain to, or [3] is intended to pertain to, or [4] does not pertain to, a hierarchically higher state of affairs, which is A, the pertainee, as stipulated by law, jurisdiction, convention, determination or any other aspect that is generally accepted as factual or mandatory in the world of values the ICLO is designed to work in/for.
:

Examples:	France	IS_INSTPART_OF [1]	European Union
	poverty rule	IS_INSTPART_OF [2]	Christian monasteries
	Sachalin	IS_INSTPART_OF [3]	Japan

(capital punishment IS_INSTPART_OF[4] criminal law) - IS_IN_PLACE_OF[1] - Germany

RelClass2-Type 3.4 Composition (matter)

In ICLO statements, the Composition (matter) relationship expresses that argument B, the component, [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, a constituent of A, which is the composed, the whole, as by matter or stuff.

Role of arguments:	A is the component (part)
	B is the composed (whole)

Meaning of Composition (matter) as modelled in this ontology: The concept of the relationship between a state of affairs B, the composed whole, which, within the hierarchical system of established human knowledge on matter or stuff, [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, a component of A. There may be other A that [1] are, or [3] are intended (claimed) to be, or [4] are not, components of B.

Examples:	epicatechine (A)	IS_MATTER_OF[1]	cacao beans (B)
	diamonds (A)	IS_MATTER_OF[1]	carbon

RelClass2-Type 4 Causality

In ICLO statements, the Causality relationship expresses that argument B, the caused, [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, the cause of A, which is the cause.

Role of arguments: A is the cause
B is the caused

Meaning of Cause as modelled in this ontology: The concept of the relationship between a state of affairs B, the caused, which, within the structures of established human knowledge, [1] brings about, or [2] may bring about, or [3] is intended (claimed) to bring about, or [4] does not bring about, A to happen as a result.

Examples:	deep sea earthquakes (A)	IS_CAUSE_OF[1]	tsunamis (B)
	deep sea earthquakes (A)	IS_CAUSE_OF[2]	tsunami catastrophes (B)

RelClass2-Type 4.1 Regulation

In ICLO statements, the Regulation relationship expresses that argument B, which is the regulated, [1] is regulated, or [21] may be regulated, or [3] is intended (claims) to be regulated, or [4] is not regulated, by the regulator, which is A.

Role of arguments: A is the regulator
B is the regulated

Meaning of Regulation as modelled in this ontology: The concept of the relationship between a state of affairs B, the regulated, which [1] is regulated, or [2] may be regulated, or [3] is intended (claimed) to be regulated, or [4] is not regulated, by regulator A within proper parameters, whereby B is supposed to react in conformance with the command of A. The relationship implies a process of response or retroaction by B to stimuli or messages emitted by A, and insofar is an expression of interaction between A and B, whereby it remains open what the reaction is or in which way the reaction is performed.

Example: insulin-like growth factors IGF (A) IS_REGULATOR_OF[2] breast cancer cell proliferation (B)

RelClass2-Type 4.11 Enhancement

In ICLO statements, the Enhancement relationship expresses that argument B, which is the enhanced [1] is enhanced, or [2] may be enhanced, or [3] is intended (claims) to be enhanced, or [4] is not enhanced, by the enhancer,, which is A.

Role of arguments: A is the enhancer
B is the enhanced

Meaning of Enhancement as modelled in this ontology: The concept of the relationship between a state of affairs B, the enhanced, which [1] is enhanced, or [2] may be enhanced, or [3] is intended (claimed) to be enhanced, or [4] is not enhanced, by the enhancer, making B greater as in value, activity, performance, desirability, but also as in negative values like fear, deficiency, morbidity.

Example: addition of search words to free text (A) IS_ENHANCER_OF[2] recall in full text browsing (B)

RelClass2-Type 4.111 Beneficial

In ICLO statements, the Beneficial relationship expresses that argument B, which is the beneficiary [1] benefits, or [2] may benefit, or [3] is intended (claimed) to benefit, or [4] does not benefit, from beneficial action or attitude by the beneficent, which is A.

Role of arguments: A is the beneficent
B is the beneficiary

Meaning of *beneficial* as modelled in this ontology : The concept of the relationship between a state of affairs B, the beneficiary, who/which [1] benefits, or [2] may benefit, or [3] is intended (claimed) to benefit, or [4] does not benefit, from A, the beneficient, as to usefulness and greater value, which has to be judged on the desirability of sustainability, not short term profit **[9],** 17.

Example: alphabetisation (A) IS_BENEFICIAL_FOR[1] higher quality of life (B)

RelClass2-Type 4.112 Detrimental

In ICLO statements, the Detrimental relationship expresses that argument B, which is the harmed [1] is harmed, or [2] may be harmed, or [3] is intended (claimed) to be harmed, or [4] is not harmed, by the harmful., which is A.

Role of arguments: A is the harmful
B is the harmed

Meaning of *detrimental* as modelled in this ontology : The concept of the relationship between a state of affairs B, the harmed, who/which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not harmed by A, the harmful, as to detrimental effects suffered , which has to be judged on values of desirability of sustainability, not short term damage **[9],** 17.

Example: overfishing (A) IS_DETRIMENTAL_TO marine living resources

RelClass2-Type 4.12 Diminution

In ICLO statements, the Diminution relationship expresses that argument B, which is the diminished [1] is diminished, or [2] may be diminished, or [3] is intended (claimed) to be diminished, or [4] is not diminished, by the diminishing, which is A.

Role of arguments: A is the diminishing
B is the diminished

Meaning of *diminution* as modelled in this ontology : The concept of the relationship between a state of affairs B, the diminished, which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not diminished by A, the diminishing, with the result of lessened effort, value, authority, renown, but also fear, danger, and damage.

Example: Aspirin (A) IS_DIMINISHER_OF[1] headache (B)
Aspirin (A) IS_DIMINISHER_OF[2] risk of heart attack (B)

RelClass2-Type 4.13 Prevention

In ICLO statements, the Prevention relationship expresses that argument B, which is the prevented [1] is prevented, or [2] may be prevented, or [3] is intended (claimed) to be prevented, or [4] is not prevented, by A, the preventer.

Role of arguments: A is the preventer
B is the prevented

Meaning of prevention as modelled in this ontology : The concept of the relationship between a state of affairs B, the prevented, which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not prevented by A, the preventer, as to keep an unwanted result from happening. Ideally, preventing implies that an un-surmountable obstacle is placed by A to stop the course of events which if not prevented by A, is imminent.

Example: Insulin-like growth factors (A) IS_PREVENTER_OF[1] apoptosis (B)

RelClass2-Type 4.2 Instrumental

In ICLO statements, the Instrumental relationship expresses that argument B, which is the instrument [1] is used, or [2] may be used, or [3] is intended (claimed) to be used, or [4] is not used, to achieve the aim of use to be attained, which is A, the purpose.

Role of arguments: A is the purpose
B is the instrument

Meaning of prevention as modelled in this ontology : The concept of the relationship between a state of affairs B, the prevented, which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not prevented by A, the preventer, as to keep an unwanted result from happening. Ideally, preventing implies that an un-surmountable obstacle is placed by A to stop the course of events which if not prevented by A, is imminent.

Example:	gas pipeline (A)	IS_INSTRUMENTAL_FOR[1]	transport of gas (B)
	risk diversification (A)	IS_INSTRUMENTAL_FOR[2]	minimizing risk exposure (B)
	venture capital investment (A)	IS_INSTRUMENTAL_FOR[3]	bringing innovation about (B)
	Agent Orange (A)	IS_INSTRUMENTAL_FOR[4]	tropical warfare (B)

RelClass2-Type 4.3 Process

In ICLO statements, the Process relationship expresses that argument B, the processed, which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, processed by the processor, as to any process, or kind of processing that might be intended or involved.

Role of arguments: A is the processor
B is the processed

Meaning of *processing* as modelled in this ontology : The concept of the relationship between a state of affairs B, the processed, which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not processed by A, the processor which is A. Any kind of process, treatment, processing or proceedings is addressed.

Example:	fully automated assembly line (A)	IS_PROCESS_APPLIED_IN[1]	car production (B)
	silver process (A)	IS_PROCESS_APPLIED_IN[2]	imaging (B)

RelClass2-Type 4.31 Mediation

In ICLO statements, the Mediation relationship expresses that argument B, the mediated, [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, mediated, by A, the mediator, by acting as go-between or as mediating agent, in actions, processes, chemical reactions, disputes, and aspirations.

Role of arguments: A is the mediator
B is the mediated

Meaning of *mediating* as modelled in this ontology : The concept of the relationship between a state of affairs B, the mediated, which [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, mediated by A, the mediator. Any type of intervention between A and B in order to achieve a specific end is addressed.

Example: Insulin receptor substrate 2 (A) IS_MEDIATOR_IN[2] Insulin-like growth factor IGF1 (B)

RelClass2-Type 4.311 Communication

In ICLO statements, the Communication relationship expresses that argument B, the receiver, [1] receives,, or [2] may receive, or [3] is intended (claimed) to receive, or [4] does not receive, a message from A, the messenger.

Role of arguments: A is the messenger
B is the receiver

Meaning of *communicating* as modelled in this ontology : The concept of the relationship between a state of affairs B, the receiver who/which [1] receives, or [2] may receive, or [3] is intended (claimed) to receive, or [4] does not receive, a message carried from A, the messenger. Any type of messaging , signalling, or transfer is addressed by this relationship. To model "IN_CROSSTALK_WITH", a separate statement has to be added, where B is the messenger, and A is the receiver.

Example: RNA (A) IS_MESSENGER_TO[1] DNA (B)

RelClass2-Type 4.312 Cooperation

To model "Cooperation", the same principle as in RelClass2-Type 431 Communication can apply, namely, that the mutual character of the relationship is expressed by a separate, inverse statement. If there are more than two partners that cooperate, however, the relationships may become confusing . To avoid wildcat cross-linking, the following procedure is proposed:

The mutual relationship is covered from as many sides as there are partners, each partner demanding a separate ICLO statement with respect to himself and only one of the partners who, in mutual agreement, functions as "body" of the cooperation which then can be stipulated as "partnership" . This allows for expression of different attitudes (modes).the partners may have vis-à-vis their common partnership. Relationships that exist between all other partners.(those who not function as the "partnership") must be modelled the ordinary way.

Having said this, the Cooperation relationship expresses that argument B1, who is the partner No. 1 in the partnership A, [1] cooperates, or [2] may cooperate, or [3] intends to cooperate, or [4] does not cooperate, with A, the partnership that unites B1 and all other partners of the partnership with the partner stipulated as "partnership", respectively.

Role of arguments: A is partnership
B is partner B1 in partnership

Meaning of *cooperation* as modelled in this ontology : The concept of the relationship between a state of affairs B, the partner B1 of A, the partnership, who/which [1] cooperates, or [2] may cooperate, or [3] intends to cooperate, or [4] does not cooperate, with A with the aim to attain mutual benefit.

Example: Sharp (A) IS_COOPERATING_WITH[1] Toyota (B)
Toyota (A) IS_COOPERATING_WITH[1] Sharp (B)

RelClass2-Type 4.313 Unification

Also a mutual relationship that must be considered from as many sides as there are candidates that unite, each candidate demanding a separate, double ICLO statement with respect to himself and the unified body, so as to allow for expression of different attitudes (modes).the candidates may have vis-à-vis their unification. To avoid wildcat cross-linking, one out of the two first partners, would be stipulated as "unified body A".(In a classical merger of two corporations, the "unified body A" need not always be the actually established firm.)
The Unification relationship expresses that argument B1, the candidate No. 1 in the unified body A, [1] unites, or [2] may unite, or [3] intends to unite, or [4] does not unite, with A, the unified body that unites B1 and all other candidates of the unification with the candidate stipulated as "unified body A", respectively.

Role of arguments: A is the unified body
B is candidate B1 in unification

Meaning of *cooperation* as modelled in this ontology : The concept of the relationship between a state of affairs B, which is the candidate B1 of A for the unified body which [1] cooperates, or [2] may cooperate, or [3] intends to cooperate, or [4] does not cooperate, with A, the unified body.

Example: Daimler-Benz AG (A) IS_UNITING_WITH[1] Chrysler Corporation (B)

RelClass2-Type 4.4 Descendancy

In ICLO statements, the Descendancy relationship expresses that argument B, the descendant, [1] deriving , or [2] possibly deriving, or [3] intended (claimed) to derive, or [4] not deriving, from A, the ancestor, by birth, school of thought, experience, invention or other source.
Role of arguments: A is the predecessor
B is the descendant

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Meaning of *descending* as modelled in this ontology : The concept of the relationship between a state of affairs B, the descendant, who/which [1] derives, or [2] may derive, or [3] is intended (claimed) to derive, or [4] does not derive, from A, the predecessor, as by birth, school of thought, experience, invention or other source.

Example:

ICE train 3 (A)	IS_DESCENDANT_OF[1]	ICE train 2 (B)
Anna Anderson (A)	IS_DESCENDANT_OF[3]	Nicholas II of Russia (B)
Anna Anderson (A)	IS_DESCENDANT_OF[4]	Nicholas II of Russia(B)

RelClass2-Type 5 Presentation (Appearance)

In ICLO statements, the Presentation (Appearance) relationship expresses that argument B, the appearing, [1] is created, or [2] may be created, or [3] is intended (claimed) to be created, or [4] is not created, by A, the creator,. as to the outward aspect or impression of what appears.

Role of arguments: A is the creator
B is the appearing

Meaning of *presentaion and appearance* as modelled in this ontology : The concept of the relationship between a state of affairs B, the appearing that [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, created by A, the creator, as to the outward aspect or impression of the appearing.

Example:

Nebra Disc (A) IS_PRESENTED_AS[1] golden sun (B)

RelClass2-Type 5.1 State/Form

In ICLO statements, the State/Form relationship expresses that argument B, the matter, [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, the state or form A of the matter, all this in the broadest sense.

Role of arguments: A is the state/form of matter
B is the matter

Meaning of *state and form* as modelled in this ontology : The concept of the relationship between a state of affairs B, the matter (or stuff, idea, concept) that [1] is, or [2] may be, or [3] is intended (claimed) to be, or [4] is not, in the state or form A of the matter (or stuff, idea, concept), in a very broad sense.

Examples:

ice (A)	IS_STATE/FORM_OF[1]	water (B)
amber (A)	IS_STATE/FORM_OF[1]	pine resin (B)
peace offer (A)	IS_STATE/FORM_OF[3]	threat (B)

RC 25

The relations from the "Arsenal" are inter-related, similar to the arguments with their respective R. In the ICLO, one single relationship is believed to suffice, which be the RR, - just a loose hierarchy according to points 2 to 5 in the table of relations proposed in RC22 above. So,

3.4 Composition – IS_NARROWER_RELATION_THAN – 3. Partitive.
4.13 Prevention – IS_NARROWER_RELATION_THAN – 4.1 Regulation, whereby
4.1 Regulation - IS_NARROWER_RELATION_THAN – 4. Causality.

The inverse side of the RR is BROADER RELATION (not shown here).

The RR has no distinction after modes.

Rationale:
The semantic matter is continuous and fluent, and the differences are gradual, observes G. Lazard[31], The ICLO has to face this dilemma when set at understanding unknown, heterogeneous texts, e. g. in semantic browsing, and high-end text mining. Rough, pragmatic inter-relating (broader/narrower) of the commonly understandable, well-defined, relationships offered by the Arsenal is expected to respond to this requirement in some degree.

Main advantages could be:

- In stipulating: Higher relations (hypernyms) can be offered as default positions.
- In searches: Hypernyms can augment recall and probably precision as well.
- Easy to implement on TM.
- Inter-relating of typed, well-defined relations makes the ontology more flexible in meeting demands of fuzzy, fluent expressiveness while protecting against arbitrary proliferation of new relationships whose semantics may be not really clear[32].

RC 26

The face side of the formalized Relational Operator – Rface internally can serve as a proxy for the different NL equivalences on the Equivalence Chain of Relational Expressions – ECRE. One ECRE each is needed for the languages implemented in the ontology.

Ideally, the ontology should "know" all the different expressions in common use for relations from the Arsenal, in all languages implemented, in order to be "aware of" of how they can be expected to occur in fresh textual messages of all kinds and origin.

- Such goal, of course, will never be attained in full, but a mechanism can be installed, and provision can be taken in such a way that the ontology systematically is enriched with new, related knowledge from the source texts it is put to browse and read, – both by intellectual and MA inference.

To this end, some device must be constructed which, when applied on the terms of unknown texts, functions like the tentacles of a medusa: They stretch wide out in order to yield the most of what can be found within reach. Here is an example for this:

Following is an example of the "terminological tentacle" that spreads out from an Rface of the Relation 4.11 Prevention, in English, French and German.

The Rface of Relation 4.13 Prevention is:

As seen from the active argument A:

A IS_PREVENTER_OF B

English language expressions
Phrases* A *prevent(s)* Phrases B
Phrases A *protect against* Phrases B
Phrases A *keep(s) back* Phrases B
Phrases A *stop(s)* Phrases B
Phrases A *counteract(s)* Phrases B

[31] LAZARD, G. 1992: *Y a-t-il des catégories* **[17].**
[32] SIGEL, A. 2003: *Topic Maps* **[18]**.437

Phrases A *hinder(s)* Phrases B
Phrases B *block(s)* Phrases B

And all inverse, seen from the passive argument B ("prevented by", and other equivalents), and machine-enriched by past tense and pp lingware for English is required. Open for inclusion of more candidates as may be encountered in future NL source texts.

French language expressions
Phrases A prévient/préviennent Phrases B
Phrases A empèche(nt) Phrases B
Phrases A stoppe(nt) Phrases B
Phrases A arrète(nt) Phrases B
Phrases A bloque(nt) Phrases B
Phrase A mis/mise/mises à l'abri de Phrases B

And all inverse and machine-enriched by other most used tenses. Lingware for French required. Open for new candidates as encountered in future NL source texts. Open for inclusion of more candidates as encountered in future NL source texts.

German language expressions
Phrases A beugt/beugen Phrases B [2]** vor
Phrases A blockt/blocken Phrases B [2]**ab
Phrases A blockiert/blockieren Phrases B
Phrases A schützt/schützen vor Phrases B
Phrases A stoppt/stoppen Phrases B
Phrases A verhindert/verhindern Phrases B

And all inverse, and machine-enriched by other most-used tenses. Lingware for German required. Open for new candidates, like "*Vorbeugung vor* Phrases B*** *durch* Phrases A"

**) Phrases are either (1) given, defined, "known" NL text strings from the ECEs standing for the argument (universals or instances inferred from the ontology), or (2) searched, by nature undefined, "unknown" NL character strings of variable length expected to occur in NL text corpora browsed.*
***) till 2 words after*
****) till 3 words after*

Comments:

The tentacle stretches wide out also for near equivalences, set to achieve enhanced IR results. But this then, inevitably will have to be traded off against problems of adaptation when it comes to generate NL answers from the core of the ontology, the BSRS, which cannot find the way back to the original co-ordination that stems from the source text.

- So, expressed by the preferential term of the relation, the generated output in NL may produce some linguistically strange, or surprising, results. Anyhow, being univocal machine-generated productions, it is safe they will all remain human-understandable utterances.
- Whether also on the semantic conformation of generated NL expressions for relations some sort of UOQ Relational Operator can be constructed on the basis of its Upper Ontology Qualifier? – it just comes to the mind.

The idea, at least, appears charming: Programmed "re-construction" of NL enouncements whereby what on formalized model level is (a) Argument with its ECE and (b) R with its Rface and ECRE would both appear co-ordinated in a safe(r) way.

Rationale:
With the "terminologic tentacle", and with the whole of the ICLO construction, a point is reached where gathering of new knowledge from heterogeneous NL texts seems possible, if not on programmed level then at least on the level of MA. This, then, could be used in order to widen and deepen the knowledge in the ICLO. Also, it would offer a platform for constant MA learning while the ontology is in use.

So, a question can be posed to an unknown corpus of NL texts of the sort:

Uphrases? – R1 – A2

Or

Uphrases ? – R1 – A2
|
R2

Or other combinations that all would be defined and predictable as in number and meaning.

Thereby "Uphrases?" would be unknown phrase fragments to be searched for in the NL texts as well as in the ontology itself.

Provided the ontology is complete and all entries are checked and correct, a question as complex as shown in the following example could be posed. It could be prompted after programmed decomposition (for logic exploitation and knowledge interchange):

> ***Public aid in France for damage caused by catastrophes since 1996: Which cases?; Who paid (central state, regions, departments, communities? EU?, public agencies?); What sort of catastrophes (flooding, oil spills, storms, etc)?; Which storms?; Which tankers? Who paid? How much?***
>
> The ontology then would "know" what "public" means, what "public aid" means, and from which institutions public aid can be expected to come, it would know which of such instances exist and which of them are related to France, and it would know their names, etc. Correct decomposition of the query can be supposed a matter well under control since pertinent networked knowledge in the ontology is at hand.

The answer would then come (a) as NL texts found in matches as well as (b) prompts from the ontology's kernel (the BSRS) in the form of generated, human-understandable units of enouncements, all along, on demand, with sound documentation of decisions and sources.

As offspring from the new NL texts discovered, fresh, hitherto un-noticed expressions and facts would be spotted and could be yielded on top of it. Corresponding organization and mechanisms of processing would have to be developed. This would be ideally suited to amplify, assert and solidify the knowledge in the ICLO which the ICLO then checks and adds and relates to its formalized "known" and NL contents.

Methods then can be developed how a distributed, programmed and MA knowledge acquisition can be performed. Sustainable, sound editorial, commercial and financial platforms would have to be secured, and corresponding mechanisms and software for interoperable reuse were to be created.

This would be the prerequisite that from a very humble beginning, and after attentive "pump priming" in a special breeding environment, the ontology can be constantly widened, deepened and updated in a decentralized, interoperable way, on matters of which it is safe that (a) they are part of the current world-wide social discourse and (b) that the respective organized knowledge from the ICLO not only can be shared but is actually required by the respective user and language communities.

Concluding remark

That an ontology can feed and support itself (at least in part) from external NL textual knowledge sources, that it can understand them (within still narrow limits) and insofar is sustainable and able to grow, is a long-desired, ultimate goal in linguistic and ontological engineering. A device such as the ICLO, with the necessary complexity indicated in this outline (and probably a little bit more needed for this), one day may make it come true.

Not discussed here:
Questions of mark-up in engineering. Extensions foreseeably required on the TM model. The External Source Interface - ESI as part of the arguments. The KWIC SCROLL magnifying "glass". Questions of construction, update and maintenance, reuse and sharing (segmentation, merging) under decentralized, semantically and technically interoperable conditions. The easy, really ergonomic, multi-lingual ICLO interface for (a) constructors, and (b) for users, and its programming. The editorial set-up, copyright and privacy issues, and other juridical questions. The human resource, its qualification, training and standing. Investment and financing, and possible business models. The political frame and questions of political correctness in the multi-cultural environment. Life-cycle of the ontology, just to name these points.

References

[1] Gruber, T. R. 1993: *Toward principles for the design of ontologies used for knowledge sharing*. In: Guarino, N., Poli, R. (eds). Intl. Workshop on Conceptual Analysis and Knowledge Representation, Padova, Italy.. Proc. Kluwer Academic Publishers, Deventer, The Netherlands. http://citeseer.nj.nec.com/gruber93toward.html .

[2] Schmitz-Esser, W. 2005: *Ontology-based understanding of scientific natural language texts*. In: Hitzler, P., Lutz, C., Stumme, G. (eds) Foundational Aspects of Ontologies. Fachberichte Informatik, FB4 Universität Koblenz-Landau, 2005. 34-45. http://www.uni-koblenz.de/fb4/publikationen/gelbeReihe/RR-9-2005.pdf

[3] Schmitz-Esser, W. 2000: *Expo 2000 – INFO 2000. Visuelles Besucherinformationssystem für die Weltausstellung*. Springer Heidelberg, 2000. p. 10, 18.

[4] Schmitz-Esser, W. 1999: *Gedankenraumreisen – Neue Thesaurusstrukturen, multimedial präsentiert, machen Anregung, Spielen, Lernen, Finden möglich für jedermann*. In: Ohly, H., Rahmstorf, G., Sigel, A. (Hrsg). Fortschritte in der Wissensorganisation. Proc. 6. Tagung der Deutschen Sektion der Internationalen Gesellschaft für Wissensorganisation, Hamburg. 1999, Bd. 6 (FW – 6), Ergon, Würzburg. P. 197-210.

[5] Schmitz-Esser, W. 2000: *SERUBA. A new search and learning technology for the Internet and Intranets*. Proc. 11th ASIS&T SIG/CR Classification Research Workshop, Chicago, IL, Nov. 12, 2000. pp.91-102. Accessed in Oct. 2001 at http://uma.info-science.uiowa.edu/sigir/papers/sigir00schmitz.pdf http://www.gca.org/papers/xmleurope2001/papers/pdf/s04-3.pdf

[6] Schmitz-Esser, W. 2000: *How to cope with Dynamism in Ontologies*. In: Beghtol, C., Howarth, L.C., Williamson, N.J. (eds). Dynamism and Stability in Knowledge Organization. Proc. 6th International ISKO Conference, July 2000, Toronto, Canada. Ergon, Würzburg 2000, pp. 83-89.

[7] Poli, R., Schmitz-Esser, W., Sigel, A. (2006): *Introducing Ontologies.* Workshop. Proc. 9th International ISKO Conference, Vienna, 2006……..

[8] Schmitz-Esser, W. (2003): *Meaning, understanding, and the organization of knowledge in a multilingual world – New tools for new tasks: Ontologies.* In: Vielberth, J., Drexel, G. (eds) Linguistic cultural identity and international communication.. AQ-Verlag, Saarbrücken, 2003, 149-171.

[9] Schmitz-Esser, W. (1999): *Thesaurus and Beyond. An advanced formula for linguistic engineering and information retrieval*. Knowl. Org. 26(1999), No. 1, 10-22

[10] Ceusters, W., Smith, B., Goldberg, L., (2005): *A terminological and Ontological Analysis of the NCI Thesaurus*. http://ontology.buffalo.edu/medo/NCIT.pdf

[11] Poli, R. (1996: *Ontology for Knowledge Organization***.** In: Green, R. (ed). Knowledge Organization and Change. Proc.4th Int. ISKO Conference, Washington DC 1996. Indeks Verlag, Frankfurt/Main. 1996. 313-319.

[12] Lassila, O., McGuinness. D. (2001): *The role of frame-based representation on the Semantic Web*. Technical report KSL-0102. Knowledge Systems Laboratory, Stanford, University, Stanford, CAL.

[13] Gómez-Pérez, A., Fernandez-López, M., Corcho, O.: *Ontological engineering*. Springer-Verlag, London.

[14] Margotti, F. W. (2003): *Sinonímia e paráfrase: Algumas considerações a partir de dados do Atlas Lingüistico-etnográfico da Região Sul-ALERS.* In: Linguagem em discurso, 3/2, Tubarão, Brasil.

[15] Lazard, G. (1992): *Y a-t-il des catégories interlangagières?* In: Anschütz, S.R. (ed) Texte, Sätze, Wörter und Moneme. Festschrift für Klaus Heger zum 65. Geburtstag. Heidelberg: Heidelberger Orientverlag, 427-434.

[16] Poli, R. (2006): *Role of upper ontology in a non-committed ontology ??* (forthcoming)

[17] see 15

[18] Sigel, A. (2003): *Topic Maps in Knowledge Organization*. In: Park, J., Hunting, S. (eds.). XML Topic Maps. Creating and Using Topic Maps for the Web. Addison-Wesley, Boston, 2003. 338-476.

[19] Poli, R. (2003): *Descriptive, Formal and Formalized Ontologies.* In: Fisette D. (Ed.): Husserl's Logical Investigations Reconsidered.: Dordrecht: Kluwer, 2003, 183-210

[20] Jaenecke, P. (1996): *Elementary Principles for Representing Knowledge.* KO 23(2)88-102.

[21] Soergel, D. et al.(2003) : *Reengineering thesauri for new applications.* 5th NKOS Workshop, Building a more meaningful Web. As part of JCDL, 2003. Houston, TX, http://jodi.tamu.edu/Articles/v04/i04/Soergel/

[22] Körner, H.G. (2001): *Cross-Language ontologies and topic maps.* Proc. Infonortics, Boston MASS. 2002. http://www.infonortics.com/searchengines/sh01/slides-01/seruba.pdf

[23] Schmitz-Esser, W., Algermissen, J. (2001): *SERUBA. A new search and learning technology for the Internet and Intranets.* SML Europe, Berlin, 2001

[24] Drexel, G., (2001): *Knowledge engineering for intelligent information retrieval.* Proc. CICLing 2001, Mexico City, Mexico, 495-504

[25] W3C Editor's Draft, 10 February 2006: *Guidelines for RDF/Topic Maps Interoperability.* http://www.ontopia.net/work/guidelines.html

[26] Fuchs, C. (1997): *Diversité des représentations linguistiques* : Quels enjeux pour la cognition ? In : Fuchs., C., Robert, S. (eds).Diversité des langues et représentations cognitives. Ophrys, Paris, 5-24.

[27] Robert, S. (1997): *Variations des représentations linguistiques : Des unités à l'énoncé-* In : Fuchs., C., Robert, S. (eds). Diversité des langues et représentations cognitives. Ophrys, Paris, 5-24.

[28] Schmitz-Esser, W. (2003): *Thésaurus, texte intégral et web sémantique : rupture ou continuité ?* Proc. Journée d'étude organisée par l'ADBS et l'IUT de Paris 5. Le Web sémantique : de nouveaux enjeux documentaires ? Paris, 2003.

[29] Fensel, D. (2002) : *Language standardisation for the Semantic Web : The long way from OIL to OWL.* In: Proc. 4th Int. Conference on Distributed Communities on the Web. Sydney, Australia, 2002. www.sigmod.org/dblp//db/indices/a-tree/f/Fensel.Dieter.html

[30] Bateman, J. A. (2006): *The Bremen Ontology Research Group.* http://www.fb10.uni-bremen.de/anglistik/ontology/ Last page update: 8 March 2006.

[31] Smith, B. (2006): *Beginnings of a draft Upper Level Ontology.* http://ontolog.cim3.net/forum/uos-convene/2006-02/msg00059.html 26 Feb. 2006

[32] CyCorp, (2006): *CYC* http://www.cyc.com/

[33] Bateman, J. A. (2006): *The Generalized Upper Model.* http://www.fb10.uni-bremen.de/anglistik/langpro/webspace/jb/gum/index.htm Last Update: 17 March 2006.

[34] Hovy, E. (2002): *Comparing Sets of Semantic Relations in Ontologies.* In: Green, R., Bean, C.A., Myaeng, S. H. (eds): Semantics of Relationships: An Interdisciplinary Perspective.. Kluwer, 2002, 91-110. http://www.isi.edu/natural-language/people/hovy/papers/02ontolbook-methodology.pdf

[35] Rahmstorf, G. (2006, forthcoming): *Sprachliche Aspekte der Wissensorganisation.* 10th Workshop of the German ISKO Chapter, Vienna, 2006.

[36] Rahmstorf, G. (1983): *Die semantischen Relationen in nominalen Ausdrücken des Deutschen.* Diss. Mainz, 1983.

[37] Newell, A., Simon, H. A. (1976): *Computer Science as Empirical Inquiry: Symbols and Search.* In: Comm. ACM, Vol. 9, 3 (1976) 113-126.

[38] Schmitz-Esser, W. (2005): *More information on some innovative projects.* (German) http://www.schmitz-esser.de

[39] Bechhofer, S., Goble, C., Horrocks, I. (2003): *Requirements of ontology languages. Deliverable 4.1 of the OntoWeb project.* http://ontoweb.aifb.uni-karlsruhe.de/About/Deliverables.

Citation: **Winfried Schmitz-Esser (2006): Introduction to Integrative Cross-Language Ontology (ICLO): Formalizing and interrelating textual knowledge to enable intelligent action and knowledge sharing**, pp. 54-113, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Acronyms, and use of terms in this paper

AAE	Additional Access Expressions
Addendum	a specific formalized addition to the Onto-DESC standing for an instance, in order to name the instance in an unmistakable way as an entity singular in space and time.
AGROVOC	Agricultural Agency of the UN
AI	Artificial Intelligence Research and Development (R&D)
Arg	Argument
Argument	semantically related, formalized object of thought on model level.
Category	unit of thought sharing the same characteristics. Universals, concepts, classes**,** topics and themes all are considered categories in this paper.
CL	Cross-Language
Class	category in this paper
CLIR	Cross-Language Information Retrieval
CLKO	Cross-Language Knowledge Organization
Concept	category in this paper
CSt	Chain of Statements formed by a Joint Identification Number JtID
ECE	Equivalence Chain of Expressions
ECD	Equivalence Chain of Descriptors
ECRE	Equivalence Chain of Relational Expressions
ESI	External Source Interface of the canonical ICLO argument
FOL	First Order Logic. ICLO regulations may restrain full application of FOL
Formalize	to make formal in order to enable sentential logic and restricted FOL and some calculus in the ICLO. Also, renders output in natural language expressions more predictable.
ICLO	Integrative Cross-Language Ontology
IT	Information Technology
Joint	a common ID number – JtID that binds together two or more binary or ternary statements (by means of their StID) to form a formalized statement of higher complexity.
JtID	Joint ID
KO	Knowledge Organization
KOS	Knowledge Organization Systems
KR	Knowledge Representation
KRS	Knowledge Representation Systems
KWIC	Keyword-In-Context Index
Lac	Language Acronym for language of argument
MA	Machine-aided, machine-assisted
MLIN	Meta Language Identification Number, indicating that the ECD stands for a universal (category)
MNIN	Meta Name Identification Number, indicating that the ECD stands for an instance
MTC	Main Type of Categories
MTI	Main Types of Instances.
MULTI	Term or term string with more than one meaning (polysem)
NCI	National Cancer Institute
NL	natural language as in general use in a language community, including the languages of specific domains, but excluding symbolic languages.
Onto-DESC	Descriptor as defined in the ICLO, or just DESC.
Osource	NL text source revealing a sought-for variety of expression for the meaning in the process of onomasiological ontology construction
Phrases	may be NL text strings formalized on ECEs or Uphrases
PSSH	Physical Symbol Systems Hypothesis after NEWELL, A., SIMON, H. A. 1976: *Computer Science*
R	symbol for the relational operator standing for the semantic relation on model level. In the ICLO, it functions as the representative of the different NL expressions encountered in source texts which are stipulated for this relation.
RC	Rules and principles of construction of the ICLO

Restricted FOL	First Order Logic applied under ICLO restrictions.
Rface	Acronym for the formalized relation in the ICLO as seen from the active side A (face side). To give an example: "IS_INSTPART_OF" is the Rface of the Pertainance relation.
ROSA	Relational Operator denominating semantic relations not contained in the "Arsenal" but needed for Specific Applications.
RR	Relation between relations, a loose hierarchy (narrower/broader) of relations listed in points 2 to 5 of table of canonic relations proposed ("Arsenal"). Just this one type of RR is admitted in the ICLO.
Statement	formalized minimal terminological enunciation representing a segment of world knowledge on model level.
StID	Statement ID of binary or ternary statement.
TECE	Type of Term on an Equivalence Chain of Expressions – ECE
TM	Topic Maps
TStr	Term String in natural language – NL
Uphrases	Unknown phrases or phrase fragments searched for in NL texts (as referents of their relation)
UOQ	Upper Ontology Qualifier
VEx	Validity of argument expired on [timeline]. Then the respective argument is kept as a historical feature.

S.-E./s.
21. June 2006

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Alexander Sigel (sigel@wim.uni-koeln.de)
University of Cologne, Department of IS and IM, Cologne, Germany

First Idea Sketch on Modelling ICLO with Topic Maps

Abstract: Within the workshop "Introducing Terminology-based Ontologies" by Poli, Schmitz-Esser, and Sigel, this paper accompanies a 20 minute presentation of a first idea sketch on how the Integrative Cross-Language Ontology (ICLO) by Schmitz-Esser (2006) (in this paper, pp. 54-113) may be modelled with Topic Maps. The main claims are:
(1) Essential parts of ICLO, up to now only presented as text with graphics, can be adequately modelled in the Topic Maps paradigm.
(2) A working implementation of ICLO will help making ICLO both more elegant and precise. During the process of modelling ICLO constructs, several inconsistencies were found, and simplifications proposed.

Status: This paper is a preliminary technical report, intended as a communication of early work in progress to foster exchange and inspiration, not meant as a stable or normative solution. Feedback by readers is explicitly solicited.
Recommendation: In order to understand this paper, readers should get the LTM text files discussed from the author and load them with the topic map engine of choice (see e.g. OKS, www.ontopia.net, or TM4J, tm4j.sf.net).

1. Introduction

The question of this paper is:

How can one represent (model) ICLO concepts (arguments) and semantic relations (statements) with topic maps in a natural way?

The following are the most important abbreviations used (for more abbreviations used see the acronym resolution section in the ICLO chapter, p. 112f.):

- **PRI**: Published Resource Identifier (according to the Published Subject (Identifier) concept as advocated by Steve Pepper (Pepper, 2003, 2006a, 2006b; Pepper & Schwab, 2003)
- **RC:** Rules and principles of construction of the ICLO (see the respective RC sections as proposed in the ICLO chapter by Winfried Schmitz-Esser, pp. 54-113.
- **LTM:** The Linear Topic Map Notation (Garshol, 2006), a special serialization of topic maps easily understandable by humans is used in the topic map examples.
- **UOQ:** Upper Ontology Qualifier (see the respective elements of the upper ontology as proposed by Roberto Poli)

Subsequently, all RCs (#1-#26) are briefly discussed. However, their presentation does not follow the order of the ICLO RCs, but rather moves – for didactic purposes – from simple to complex.

For each RC, a table with three areas is used:

1. ICLO: A restatement/reinterpretation of the ICLO RC
2. topic map: An interpretation of the ICLO RC in topic map terms, discussing modelling alternatives
3. Example: A topic map example provided in LTM (if provided)

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

1. Expressing concepts (arguments)

RC 18: An argument

ICLO	The most atomar ICLO element is an argument, which is defined as the 8-tupel (RC 18) consisting of the following sub-elements: 1. MLIN/MNIN 2. VEx, 3. ESI 4. UOQ 5. MTC/MTI 6. TECE, 7. Lac 8. TStr
topic map	An argument is a topic representing a concept. To decide how to best model the argument's record structure (8 attribute – value pairs), we need to know if the attributes need to be topics (first class topic map constructs). They should be topics if more assertions shall be made about them, or if they shall be used in semantic retrieval. To be on the safe side, it is recommended to make all attributes topics. Therefore, it is proposed to introduce for each sub-element a new semantic association type (where such a sub-element is needed), with the attribute being the relation type, the value being the new topic, and the other end being the argument topic. The alternative (not introducing topics for the attributes) would establish the argument's record structure by using an inline occurrence of the occurrence type of the attribute, and the value to fill it. 1: IDs (MLIN/MNIN): IDs can be automatically generated. Topics in topic maps already have auto-generated IDs. In addition, reification (reinterpreting a topic map construct as a topic) needs IDs. There is no need to discern between a MLIN (Meta Language Identification Number, category/type) number range and a MNIN (Meta Name Identification Number, instance) number range, since an instance (MNIN) should be connected with with the instance-of association to a category (for each MNIN, the corresponding MLIN should be defined). This means that MNINs can be recognized by playing the instance role. 2: Validity in time (VEx): (see RC 23): Probably not needed, since an elegant alternative to modelling VEx as a sub-element is scoping each argument with VEx. For dates and date/time terms, the Semagia PRI sets could be used (http://psi.semagia.com/iso8601/, http://psi.semagia.com/datetime/). 3. External Source Interface (ESI): This is not needed since the topic map constructs subject identifier/locator and typed occurrence with URI can link to external sources 4. UOQ: A statement is related to one or more UOQs using an association of type statement_UOQ with names ("has UOQ"/"is UOQ for"). An argument can have multiple UOQs. 5. MTI/MTC: analogous to UOQ 6. TECE: It is not clear if this is needed, given that DESC is a display name, AAE a base name, and MULTI a basename of the same character string, but within a different scope. Further thinking is needed here.

	7. LAc: Not needed, since it is modelled as language scope 8. TStr: Not needed, since this is a basename In sum, associations to UOQ and MTI/MTC are needed. TECE seems to be handled by the topic map constructs. Note that only the elements 7. LAc and and 8. TStr must be filled out, the rest is optional, or automatically generated.
Example	---

RC 8: Argument sub-element MLIN/MNIN

ICLO	An argument is either a category (type), or an instance (case). The descriptor is the sole proxy of the argument.
topic map	Arguments not only assume various roles, but also have fixed positions, either connected via taxonomy (superclass-subclass), or via instantiation (class-instance). The subject identifier identifies the PRI which is the sole proxy of the argument (which is a topic). A descriptor is a basename, typically within a language scope.
Example	---

RC 23: Argument sub-element VEx

ICLO	Each statement is assigned the timeline of its validity (time frame of the statement's validity) and the entry date (stipulation or update, i.e. administrative metadata).
topic map	(see RC 18): As an elegant alternative to modelling VEx as an element is scoping each argument with VEx. For dates and date/time terms, the PRI sets http://psi.semagia.com/iso8601/ and http://psi.semagia.com/datetime/ could be used. Alternatively, a validity relation or a validity inline occurrence to represent a property (attribute-value) are possible.
Example	---

RC 14: Argument sub-element ESI

ICLO	Each argument can be linked to external resources via the ESI (External Source Interface).
topic map	Either entries within other KOS (Knowledge Organization Systems) are referenced as equivalent via subject indicators, the subject of the argument can be specified by the subject locator, or external resources are relevant to the argument and can be linked as occurrences using appropriate occurrence types.

Example	`/* ICLO RC 14 LTM */`
	
	`/* occurrence types */`
	`[government_source = "Government source"]`
	
	`/* topics */`
	`[t7 = "sectorial aid"`
	`@"http://www.websters-online-dictionary.org/definition/SECTORIAL+AID"]`
	
	`{t7,government_source,`
	`"http://ec.europa.eu/comm/competition/state_aid/decisions/n253_2005/en.pdf"}`

RC 15: Argument sub-element UOQ

ICLO	Each argument has an arbitrary number of UOQs (set)
topic map	A set of PRIs for UOQs will be published, aiding humans in classifying the argument with the appropriate UOQ. The corresponding element position will contain a set of subject identifiers. Topic maps-based semantic web services could provide inferences on restrictions implied by UOQs.
Example	---

A selection of UOQs from the upper ontology as defined by Roberto Poli:

Item
1. Substance
    - i. "Nature"
        1. Particular
        2. Non-particular
    - ii. Structure
        1. Holon
            - a. Aggregate
            - b. Whole
            - c. System
        2. Part
        3. Boundary
            - a. Position
            - b. Nature
            - c. Form
            - d. Maintenance
    - iii. Level
        1. Material
        2. Psychological
        3. Social
        4. Ideal
2. Determination
    - i. Quality
    - ii. Power
    - iii. Connection
    - iv. Configuration
3. Modifier
    - i. Window
    - ii. View

RC 17: Argument sub-element MTC (MLIN)

ICLO	MTCs
topic map	A set of PRIs for such MTCs will be published. Each MTC must be bound to at least one UOQ.
Example	---

RC 13: Argument sub-element MTI (MNIN)

ICLO	MTIs
topic map	A set of PRIs for such MTIs will be published. Each MTI must be bound to at least one UOQ.
Example	---

RC 18: Argument sub-element LAc

ICLO	The configuration of languages for which ICLO is built
topic map	Some language scoping topics

Example	`/* language scopes */`
	`[English = "English"]`
	`[French ="French"]`
	`[German = "German"]`

RC 9: Argument sub-elements TECE, TStr

ICLO	A descriptor can have arbitrary length.
topic map	Topic names can have arbitrary length, and the number of names a topic can have is also unlimited.
Example	---

RC 12: Argument sub-elements TECE, TStr

ICLO	From the descriptor standing for an instance it must become clear that it stands for a singular instance.
topic map	Checks will ensure that the descriptor does not match concepts with the same name.
Example	---

RC 10: Argument sub-elements TECE, TStr

ICLO	ECE: Equivalence Chain of Expressions ("synonym chain") for each language, connected to ECD: Equivalence Chain of Descriptors. The (Onto-)DESC is the designated descriptor standing for all other synonymous expressions.
topic map	All synonyms are basenames for a topic proxying a concept. They do not have to be linked since they are direct basenames of the same topic, or share the same subject identifier and become merged. The designated descriptor could be just the display name within the language scope (preferred). Alternatively, the scope "designated DESCriptor" could be used, or a property could be asserted as inline occurrence.
Example	(see code below)

```
/* ICLO RC 10 LTM */

 [display_name = "Display name"
  @"http://www.topicmaps.org/xtm/1.0/core.xtm#display"]

/* topics */
[t8 = "t8"
    =  "interest-bearing  deposit  accounts  offered  by  savings  banks";  ;
"interest-bearing deposit accounts offered by savings banks" / English
    = "interest bearing deposit-account at the savings bank" / English
    = "savings banks' interest-bearing deposit accounts" / English
    = "savings banks provide interest bearing deposit accounts" / English

    = "comptes rémunérés des caisses d'espargne"; ; "comptes rémunérés des
caisses d'espargne" / French
    = "comptes rémunérés dans le cadre d'un package de service bancaire des
caisses d'espargne" / French
    = "rémunération des dépôts de caisses d'espargne" / French

    = "Extra-Zinskonten der Sparkassen"; ; "Extra-Zinskonten der Sparkassen"
/ German
    = "Kundenzinskonten der Sparkassen" / German
    = "Sparkassen geben Zins ab 1. Euro" / German
    = "Zinskonto bei der Sparkasse" / German
    = "Zinsen auf Sparkonto bei den Sparkassen" / German
```

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 11: Argument sub-elements TECE, TStr

ICLO	ECE consists of at least DESC, and can additionally have AAEs (additional access equivalents, univocal) and MULTIs (polysemic equivalences).
topic map	A consistency check ensures that all expressions (arguments) have a display name. AAEs are additional basenames. MULTIs: Topics can have the same basenames without having to be automatically merged (TNC: topic naming constraint), if TNC is not followed and the topics have different subject identifiers. In any case, scope could be used to discern between MULTIs. Retrieval mechanisms (e.g. tolog value-like) can find identical basenames for different arguments.
Example	---

RC 16: Argument sub-elements TECE, TStr

ICLO	For each language in the ICLO, there must be a DESC and a corresponding ECE.
topic map	The language set for a specific ICLO application is configured as a set of scoping topics of type language. A consistency ensures that every argument will have a display name. AAEs are considered optional. Class-1-relations hold within ECEs (RC 19).
Example	---

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

2. Expressing relations (statements)

RC 1: A statement

ICLO	A statement S is a binary, typed semantic relation R_i which links its two arguments A_1 and A_2. The relation can only be a type II relation (RC 19, RC 22, RC 24). Their role players are the argument roles (RC 24). Rface is one natural language name for the relation, namely the name as seen from the side of the role player playing the active role.
topic map	See RC 18 on how an argument can be represented. A statement S can be represented as an instance of a binary association type (RC 22) with two corresponding association role types (RC 24). A set of PRIs for such association types of type I and II, and their association role types will be published. In a statement S, each argument becomes an association role which plays the role as specified by the association role type. It is documented which association role type plays the active and which the passive side (e.g. by active and passive role types connecting a active-passive association type). Rface is a basename for the association type, as seen from the active side.
Example	see code below

```
/* ICLO RC 1 LTM */

/* role types */
[rt_cause : ICLO_role_type = "cause"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_cause"]
[rt_caused : ICLO_role_type = "caused"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_caused"]

/* association types */
[r2_1 : ICLO_relation = "Cause-Caused"
                                        = "Causality"
                                        = "RelClass2-Type 4"
                                        = "is cause of" / rt_cause
                                        = "is caused by" / rt_caused
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_1"]

/* topics and associations */
[t1 = "deep sea earthquakes"]
[t2 = "tsunamis"]

r2_1(t1 : rt_cause, t2 : rt_caused)
```

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 2: Inverse relations

ICLO	Reciprocal relations and directionality
topic map	By definition, the association types hold between the two association role types, hence there is no need to define reverse relations. Association types do not imply a direction, they just connect back and forth. Within the scope of an association role type, the name of the association type representing S is an additional basename as seen from this association role type, valid only in this scope. In the ontology, additional characteristics of association types can be modelled, e.g: transitivity, or the implication that if an argument is both role player of type receiver and sender, another statement with relation type "crosstalk", with both association role types "crosstalker" and with basename "is in crosstalk with" could be added dynamically by a tolog/TMQL rule. Although first work on coding tolog inference rules *within* topic maps exists (see e.g.: Siebers: Implementing Inference Rules in the Topic Maps Model, submitted to TMRA06), normally such rules are externally applied to topic maps. Therefore, the tolog rule is shown here as a *commentary* only. The reification of the dynamically computed statement "is in crosstalk with" is also not shown here, since in LTM only static statements can be reified because the IDs of the reified statements are not known in advance. Instead, only the reification of statement 9299 is shown.
Example	(see code below)

```
/* role types */
[rt_messenger : ICLO_role_type = "messenger"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_messenger"]
[rt_receiver : ICLO_role_type = "receiver"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_receiver"]

[rt_part_location : ICLO_role_type = "part as by location"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_part_location"]
[rt_whole_location : ICLO_role_type = "whole as by location"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_whole_location"]

[rt_crosstalker : ICLO_role_type = "crosstalker"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_crosstalker"]

/* association types */
[r2_4 : ICLO_relation = "Part_Location-Whole_Location"
                                      = "Location (part/whole)"
                                      = "RelClass2-Type 3.1"
                                      = "is in place of" / rt_part_location
                                      = "is place of" / rt_whole_location
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_4"]

[r2_18 : ICLO_relation = "Messenger-Receiver"
                                        = "Communication"
                                        = "RelClass2-Type 4.311"
                                        = "is messenger to" / rt_messenger
                                        = "is receiver from" / rt_receiver
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_18"]

[r2_18_crosstalk : ICLO_relation = "Crosstalker-Crosstalker"
                              = "Communication (Crosstalk)"
                              = "RelClass2-Type 4.311a"
                              = "is in crosstalk with" / rt_crosstalker
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_18_crosstalk"]

/* topics and associations */
[t3 = "progesterone"]
[t4 = "insuline-like growth factors"]
[t5 = "breast cancer cells"]

r2_18(t3 : rt_messenger, t4 : rt_receiver)
r2_18(t4 : rt_messenger, t3 : rt_receiver)

/* the inference rule above dynamically infers: "is in crosstalk with" */
/* Type and name have to be added */
```

RC 3: No orphan arguments

ICLO	Orphan arguments are not allowed: All arguments are connected by at least one semantic relation with another argument.
topic map	An application-specific consistency check procedure will ensure this, using an efficient graph theoretic algorithm. A list of orphan arguments will be output.
Example	---

RC 4: Nesting statements

ICLO	More complex statements can be created by nesting existing statements. To do so, a statement is reinterpreted (reified) as an argument. In an application this can either mean that an atomic argument becomes more refined (specifying its internal structure to consist of two arguments linked by a semantic relation), or that someone wants to make a statement about a statement in some discourse.
topic map	Within Topic Maps, reification means “making a topic represent the subject of another topic map construct in the same topic map” (Garshol & Moore, 2005). Here, a new topic (representing an atomic argument) represents the subject of the topic map construct representing the statement. Both use the same PRI. Note that both r2_18 relations have been changed in contrast to above, since they now need a statement ID. Note that reinterpreting a statement as argument opens a new 8-tuple statement to be filled out.

Example	`R2_18(t3 : rt_messenger, t4 : rt_receiver) ~ StID9299`
	`r2_18(t4 : rt_messenger, t3 : rt_receiver) ~ StID9300`
	
	`/* StID9299 is a statement which in turn could be reinterpreted as an argument. We give it a type and a name */`
	`[StID9299 : ICLO_statement = "progesterone IS MESSENGER TO insuline-like growth factors"]`
	`[StID9300 : ICLO_statement = "insuline-like growth factors IS MESSENGER TO progesterone"]`
	
	`/* the inference rule above dynamically infers: "is in crosstalk with" */`
	`/* Type and name have to be added */`
	
	`r2_4(StID9299 : rt_part_location, t5 : rt_whole_location)`

RC 5: Aggregating statements

ICLO	More complex statements can be created by aggregating existing statements, resulting in a semantic network which is an index adhering to a sophisticated index language with a grammar (Fugmann, 1993, 1999). Aggregation is defined as at least two statements sharing one argument.
topic map	The result of the aggregation is a small topic map (fragment). It might be possible to model the aggregation as being in the same scope of the JtID. Another possibility might be to use the existing association type "composition", or to add an association role type "aggregation". However, a topic map is already an aggregated semantic network in which all connections an argument plays roles in can be easily displayed from one virtual location. Hence, there seems no need for an aggregation relation. Instead, it is proposed to reify the whole semantic network, i.e. to use the construct for topic map reification (See (Garshol, 2006), 2.7.1 Topic Map directive), probably in combination with the mergemap directive. The whole topic map is reinterpreted as a statement. In this example, the topic map below shaded in grey is reinterpreted as topic called JtID2247. Reinterpreted as an argument, it could play further roles. Note that for simplicity, the further statements "is opposed to", and "is supporter of" have not been modelled here. Presumably the appropriate relations would be "beneficial" and "detrimental", respectively. Note that reinterpreting a topic map as a statement, and later reinterpreting the statement as argument opens a new 8-tuple statement to be filled out.
Example	(see code below)

```
/*further definitions */

/* role types */

/* 2_2 */
[rt_subclass : ICLO_role_type = "specific being"
                              = "subclass"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_specific_being"
  @"http://www.topicmaps.org/xtm/1.0/core.xtm#subclass"]
[rt_superclass : ICLO_role_type = "specific being"
                                = "superclass"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_being"
  @"http://www.topicmaps.org/xtm/1.0/core.xtm#superclass"]

/* 2_15 */
[rt_purpose : ICLO_role_type = "purpose"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_purpose"]
[rt_instrument : ICLO_role_type = "instrument"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_instrument"]

/* association types */
[r2_2 : ICLO_relation = "Taxonomy"
                      = "'Is a' type of isness"
                      = "Superclass/Subclass"
                      = "is subclass of" / rt_subclass
                      = "is superclass of" / rt_superclass
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_2"
  @"http://www.topicmaps.org/xtm/1.0/core.xtm#superclass-subclass"]

[r2_15 : ICLO_relation = "Purpose-Instrument"
                       = "Instrumental"
                       = "RelClass2-Type 4.2"
                       = "is instrumental for" / rt_purpose
                       = "is instrument in" / rt_instrument
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_15"]
```

	<pre>/* ICLO RC 5 LTM */
	
	/* topics and associations */
	[t6 = "government bail-outs for corporate entities"]
	[t7 = "sectorial aid"]
	[StID2222 = "Ministére des Finances, Paris IS BENEFICIAL TO Alstom S.A."]
	
	r2_2(t6 : rt_subclass, t7 : rt_superclass)
	r2_15(t6 : rt_instrument, StID2222 : rt_purpose)
	
	#TOPICMAP ~JtID2247
	[JtID2247 = "topic map for the aggregated statement 2247"]
	
	/* StID2247 is a statement which in turn could be reinterpreted as an
	argument. We give it a type and a name */
	[JtID9299 : ICLO_statement = "(government bail-outs for corporate entities
	IS SUBCLASS OF sectorial aid) IS INSTRUMENT IN (Ministére des Finances,
	Paris IS BENEFICIAL TO Alstom S.A.)"]</pre>

RC 22: Class-2 relations: Overview

ICLO	Definition of the class-2 relations
topic map	A set of PRIs for such association types of type II, and their association role types will be published. For mode, see RC 21. In the ontology, additional characteristics of association types can be modelled, e.g: transitivity, or the implication that if an argument is both role player of type receiver and sender, another statement with relation type "crosstalk", with both association role types "crosstalker" and with basename "is in crosstalk with" could be added dynamically by a tolog/TMQL rule.
Example	---

RC 24: Class-2 relations: Detail ("Arsenal")

ICLO	Definition of the arsenal
topic map	The PRI set for the arsenal will cover this
Example	---

RC 19: Class-1 vs. Class-2 relations

ICLO	Class-1-relations hold within ECEs, Class-2-relations between arguments.
topic map	A set of PRIs is defined for all Class-2 association types. Consistency checks will ensure that these association types can only take arguments to be role players. Class-1 association types could be explicitly modelled (e.g. use DESC as preferential expression for a non-descriptor AAE), but currently no need is seen for this.
Example	---

RC 25: Organizing relations by rendering them hierarchically

ICLO	RR, the hierarchical ordering of association types
topic map	The sub/superordinate_role_type pattern is used. Corresponding role players stand in a taxonomic (is-a) association. See: Published Subject Indicators For Modelling Hierarchical Relationships: http://www.techquila.com/psi/hierarchy/ Note that the active role in an association and the superordinate role need not coincide.

Example	`http://www.techquila.com/psi/hierarchy/#hierarchical-relation-type`
	`http://www.techquila.com/psi/hierarchy/#subordinate-role-type`
	`http://www.techquila.com/psi/hierarchy/#superordinate-role-type`

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop “Introducing Terminology-based Ontologies” (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

RC 21: Statement modes

ICLO	Each class-2 relation has exactly one of the following four modes: 1 – assertive (default) 2 – modal (possible world) 3 – intention 4 – negation
topic map	A mode is just a simple means of further characterizing the nature of a statement, and no intent is made to implement e.g. modal logics! PRIs will be published for the four modes. One possible implementation is to have four association subtypes according to the four modes, each one subclassing the class-2 association type. Alternatively, mode could be understood as a ternary relation (but this is not preferred here). As another alternative, the mode could be modelled as an additional relation, modifying the statement with its mode (this seems more attractive). This needs reification of the statement as argument playing the role of statement in the statement_mode relation. Modes are only types for statements. The modes should be used with care, since e.g. modal logics is not implied here.

Example

```
/* ICLO RC 21 LTM */

/* topic types */

[ICLO_statement = "ICLO statement"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/statement"]
/* also used as role type below */

[tt_mode = "mode"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/tt_mode"]
/* also used as role type below */

/* role types */

/* 2_15 */
[rt_purpose : ICLO_role_type = "purpose"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_purpose"]
[rt_instrument : ICLO_role_type = "instrument"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_instrument"]

/* mode*/
[rt_mode : ICLO_role_type = "mode"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_mode"]
[rt_statement : ICLO_role_type = "statement"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/rt_statement"]

/* predefined modes */

[mode_assertive : tt_mode = "assertive mode"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/mode_assertive"]
[mode_modal : tt_mode = "modal mode"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/mode_modal"]
[mode_intentional : tt_mode = "intentional mode"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/mode_intentional"]
[mode_negative : tt_mode = "negative mode"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/mode_negative"]

/* association types */

[r2_2 : ICLO_relation = "Taxonomy"
                  = "'Is a' type of isness"
                      = "Superclass/Subclass"
                  = "is subclass of" / rt_subclass
                  = "is superclass of" / rt_superclass
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r2_2"
  @"http://www.topicmaps.org/xtm/1.0/core.xtm#superclass-subclass"]

[r_statement_mode : ICLO relation = "Statement-Mode"
  = "has mode" / rt_statement
```

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

```
  = "is mode of" / rt_mode
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/r_statement_mode"]

/* topics and associations */

[t6 = "government bail-outs for corporate entities"]
[t7 = "sectorial aid"]

r2_2(t6 : rt_subclass, t7 : rt_superclass) ~ StID1442
[StID1442 : statement = "government bail-outs for corporate entities IS
SUBCLASS OF sectiorial aid"]

r_statement_mode(StID1442 : rt_statement, mode_modal : rt_mode)
```

3. Advanced Issues

RC 20: Natural language names for Class-2 relations

ICLO	ECRE (Equivalence Chain of Relational Expressions) is needed for equivalents of Class-2-relations.
topic map	To be developed.
Example	---

RC 26: Natural language names for ECREs

ICLO	Rface, the face side of the relational operator can internally serve as a proxy for the different NL equivalents on the ECRE.
topic map	Can/should one make the expressions first-class arguments?
Example	---

4. Considered out of the scope of topic maps modelling

The following two RCs are considered not to fall within the problem of adequate modelling of ICLO with topic maps. A methodology would have to be specified first, which what is implicit must be explicated, and what the onomasiological principle would imply here. This methodology would then have to be reformulated such that an application could support knowledge workers using topic map constructs. For now, RC 7 and RC 6 are considered adhortative to the modeller.

RC 7:

ICLO	Making tacit knowledge explicit during knowledge engineering.
topic map	Only possible once the theory will have been developed further
Example	---

RC 6:

ICLO	Adherence to the onomasiological principle (Riggs, 1996/1997), i.e. concept-to-term.
topic map	Only possible once the theory will have been developed further. Needs a theoretical model and will result in application-specific support based on it.
Example	---

5. Discussion of known limitations and further work planned

- This paper is only a draft, in particular in this section.
- Since this is only a first idea sketch which has been developed within short time, a full discussion of limitations is out of scope of this paper.
- Not all ICLO RCs could be dealt with in the same detail. Some RCs are considered out of scope of topic maps modelling.
- The argument structure has not been defined in detail. In particular, it is not yet fully understood if it is necessary to model the descriptor chains exactly as specified by the ICLO. It may well be that the Topic Maps paradigm already provides capabilities with which the same can be achieved, only simpler and more native to Topic Maps.
- No completeness criterion has been developed such that the claim that ICLO can be modelled with topic maps has been shown to be testable. The ideas presented do not prove that ICLO can be fully modelled with topic maps, or that topic maps modelling ICLO is the best approach. However, compared to the situation before this paper, a strong line of argument has been made that this approach is both feasible and useful, and that it even detects errors and omissions in ICLO, thus moving to a more formal model.
- The examples are very much simplified. E.g. the PRIs are not documented according to the minimal requirements.
- No complete example has been developed so far.
- The appendix is only a reminder to publish the example topic maps.
- No PRI set has been published yet (neither for the core ICLO meta concepts, nor for the arsenal itself). The same holds for e.g. UOQ, MCI and MTI sets.
- The examples should discern between defining the ICLO ontology and modelling instances (arguments and statements). In addition, there should be several topic maps, factoring out standard definitions and loading them in using mergemap.
- No example is shown for applying ICLO modelled with topic maps for retrieval purposes. The names (natural language expressions) to feed an information retrieval system could be extracted from the topic maps, and inter alia the value-like tolog predicate could be used.
- It is unknown if reifying that much is a good idea.
- Further work on dynamically adding inferences is needed.
- Further work in cooperation with Roberto Poli and Winfried Schmitz-Esser is under way.

6. Conclusion

A first sketch of ideas on how ICLO may be modelled using topic map constructs. Based on this work, it is claimed:
(1) Essential parts of ICLO, up to now only presented as text with graphics, can be adequately modelled in the Topic Maps paradigm.
(2) A working implementation of ICLO will help making ICLO both more elegant and precise.

7. References

Fugmann, R. (1993). *Subject Analysis and Indexing. Theoretical Foundation and Practical Advice*. Frankfurt/Main: Indeks.

Fugmann, R. (1999). *Inhaltserschließung durch Indexieren: Prinzipien und Praxis*. Frankfurt/Main: DGD.

Garshol, L. M. (2006, 2006-06-17). The Linear Topic Map Notation. Definition and introduction, version 1.3. Retrieved 2006-07-05, from http://www.ontopia.net/download/ltm.html

Garshol, L. M., & Moore, G. (2005, 2005-12-16). Topic Maps — Data Model. Final Draft International Standard. Retrieved 2006-04-24, from http://www.isotopicmaps.org/sam/sam-model/

Pepper, S. (2003, 2003-06-24). Published Subjects: Introduction and Basic Requirements. OASIS Published Subjects Technical Committee Recommendation. Retrieved 2006-04-26, from http://www.oasis-open.org/committees/download.php/3050/pubsubj-pt1-1.02-cs.pdf http://www.oasis-open.org/committees/tm-pubsubj/ http://www.oasis-open.org/committees/download.php/3052/pubsubj-pt1-1.02-cs.doc

Pepper, S. (2006a). *The Case for Published Subjects*. Paper presented at the WWW2006 Workshop: Identity, Reference, and the Web (IRW2006). Retrieved 2006-06-06, from http://www.ontopia.net/topicmaps/materials/The_Case_for_Published_Subjects.pdf (Final version) http://www.ibiblio.org/hhalpin/irw2006/spepper.pdf http://www.ibiblio.org/hhalpin/irw2006/spepper.html (Previous version, named: Towards the Semantic Superhighway. A Manifesto for Published Subjects) http://www.ontopia.net/topicmaps/materials/Public_Resource_Identifiers.ppt (Slides, named: The Case for Public Resource Identifiers: A Call for Action) http://www.ibiblio.org/hhalpin/irw2006/presentations/Public_Resource_Identifiers.ppt (Slides).

Pepper, S. (2006b). *A Proposal for Public Resource Identifiers*. Paper presented at the TMRA 2006. Retrieved 2006-06-12.

Pepper, S., & Schwab, S. (2003). *Curing the Web's Identity Crisis: Subject Indicators for RDF*. Paper presented at the XML Conference 2003. Retrieved 2006-04-24, from http://www.idealliance.org/papers/dx_xml03/html/abstract/05-01-05.html http://www.ontopia.net/topicmaps/materials/identitycrisis.html.

Riggs, F. W. (1996/1997). Onomantics and Terminology. Parts 1-4. *Knowledge Organization*, 23(21), 25-34, 23(23), 156-168, 123(154) 216-224, 124(151), 158-117.

Schmitz-Esser, W. (2006). *An Introduction to Integrative Cross-Language Ontology (ICLO)*. Paper presented at the Ninth International ISKO Conference 2006, Vienna, Austria.

Citation: **Sigel, Alexander (2006): First Idea Sketch on Modelling ICLO with Topic Maps**, pp. 114-130, in:
Winfried Schmitz-Esser & Alexander Sigel (2006): Introducing Terminology-based Ontologies.
Papers and Materials presented by the authors at the workshop "Introducing Terminology-based Ontologies" (Poli/Schmitz-Esser/Sigel) at the 9th International Conference of the International Society for Knowledge Organization (ISKO), Vienna, Austria, July 6th, 2006.
Published electronically on E-LIS (E-prints in Library and Information Science, http://eprints.rclis.org), 2006-07-14

Appendices

This appendix is only a reminder that the example topic maps should also be published here.
There should be a download address.
For the moment being, please ask the author for a current set of examples.

Appendix A: Core topic map

To be extended and adapted
More appendices with full Topic Maps to be shown.

```
/* ICLO types */
[ICLO_argument = "ICLO argument"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/argument"]
[ICLO_role_type = "ICLO role type"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/role_type"]
[ICLO_relation = "ICLO relation"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/relation"]
[ICLO_statement = "ICLO statement"
  @"http://www.schmitz-esser.de/PRI/ICLO/2006/statement"]
```

The ontology can contain inference rules, e.g. for RC 2:

```
/* ICLO inference rules */
/*
r2_18_crosstalk($CROSSTALKER, $CROSSTALKER) :-
        r2_18($MESSENGER, $RECEIVER),
        r2_18($RECEIVER, $MESSENGER).
*/
```