# (TREC-style) Question Answering systems

Christopher Manning
CS224N/Ling 237 2004

(includes slides borrowed from Sanda Harabagiu, ISI, Nicholas Kushmerick)

## Question Answering from text

- An idea originating from the IR community
- With massive collections of full-text documents, simply finding *relevant documents* is of limited use: we want *answers* from textbases
- QA: give the user a (short) answer to their question, perhaps supported by evidence.
- The common person's view? [From a novel]
  - "I like the Internet. Really, I do. Any time I need a piece of shareware or I want to find out the weather in Bogota … I'm the first guy to get the modem humming. But as a source of information, it sucks. You got a billion pieces of data, struggling to be heard and seen and downloaded, and anything I want to know seems to get trampled underfoot in the crowd."
    - M. Marshall. *The Straw Men*. HarperCollins Publishers, 2002.

## People *want* to ask questions…

Examples from AltaVista query log
who invented surf music?
how to make stink bombs
where are the snowdens of yesteryear?
which english translation of the bible is used in official catholic liturgies?
how to do clayart
how to copy psx
how tall is the sears tower?
Examples from Excite query log (12/1999)
how can i find someone in texas
where can i find information on puritan religion?
what are the 7 wonders of the world
how can i eliminate stress
What vacuum cleaner does Consumers Guide recommend
Around 10–15% of query logs

## AskJeeves

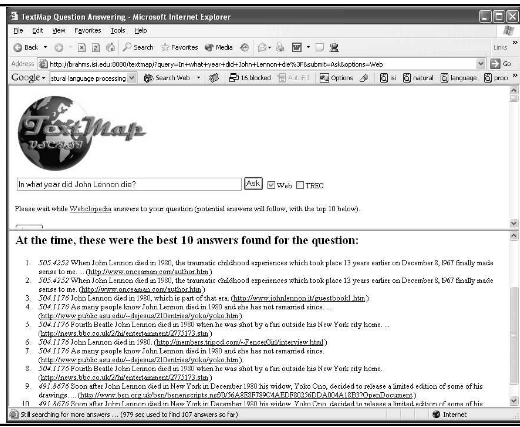
- **AskJeeves** is probably most hyped example of "Question answering"
- It largely does pattern matching to match your question to their own knowledge base of questions
- If that works, you get the human-curated answers to that known question
- If that fails, it falls back to regular web search
- A potentially interested middle ground, but a fairly weak shadow of real QA

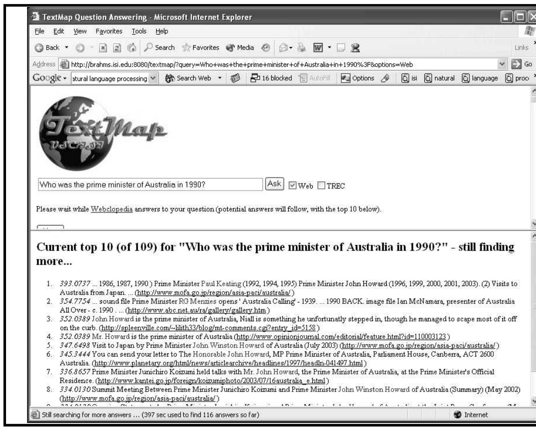
## A Brief (Academic) History

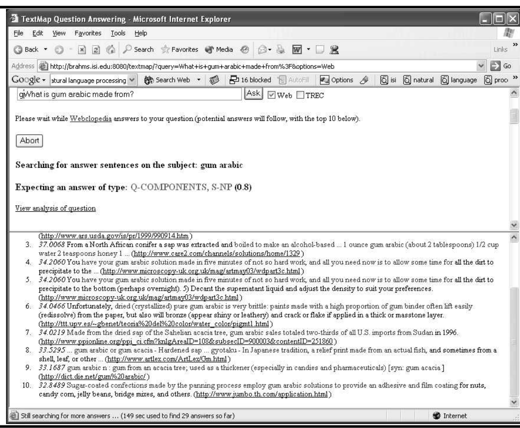
- In some sense question answering is not a new research area
- Question answering systems can be found in many areas of NLP research, including:
  - Natural language database systems
    - A lot of early NLP work on these
  - Spoken dialog systems
    - Currently very active and commercially relevant
- The focus on open-domain QA is new
  - MURAX (Kupiec 1993): Encyclopedia answers
  - Hirschman: Reading comprehension tests
  - TREC QA competition: 1999–

## Online QA System Examples

- Examples
  - **AnswerBus** is an open-domain question answering system: www.answerbus.com
  - **Ionaut:** http://www.ionaut.com:8400/
  - **LCC:** http://www.languagecomputer.com/
  - **EasyAsk, AnswerLogic, AnswerFriend, Start, Quasm, Mulder, Webclopedia, etc.**
  - **ISI TextMap** http://brahms.isi.edu:8080/textmap/




## Question Answering at TREC

- Question answering competition at TREC consists of answering a set of 500 fact-based questions, e.g., *"When was Mozart born?"*.
- For the first three years systems were allowed to return 5 ranked answer snippets (50/250 bytes) to each question.
  - IR think
  - Mean Reciprocal Rank (MRR) scoring:
    - 1, 0.5, 0.33, 0.25, 0.2, 0 for 1, 2, 3, 4, 5, 6+ doc
  - Mainly Named Entity answers (person, place, date, …)
- From 2002 the systems are only allowed to return a single *exact* answer and the notion of confidence has been introduced.

## The TREC Document Collection

- The current collection uses news articles from the following sources:
  - AP newswire, 1998-2000
  - New York Times newswire, 1998-2000
  - Xinhua News Agency newswire, 1996-2000
- In total there are 1,033,461 documents in the collection. 3GB of text
- Clearly this is too much text to process entirely using advanced NLP techniques so the systems usually consist of an initial information retrieval phase followed by more advanced processing.
- Many supplement this text with use of the web, and other knowledge bases

## Sample TREC questions

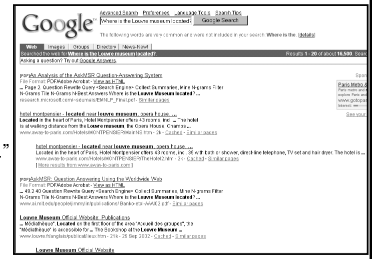
1. Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?
2. What was the monetary value of the Nobel Peace Prize in 1989?
3. What does the Peugeot company manufacture?
4. How much did Mercury spend on advertising in 1993?
5. What is the name of the managing director of Apricot Computer?
6. Why did David Koresh ask the FBI for a word processor?
7. What debts did Qintex group leave?
8. What is the name of the rare neurological disease with symptoms such as: involuntary movements (tics), swearing, and incoherent vocalizations (grunts, shouts, etc.)?

## Top Performing Systems

- Currently the best performing systems at TREC can answer approximately 70% of the questions
- Approaches and successes have varied a fair deal
  - Knowledge-rich approaches, using a vast array of NLP techniques stole the show in 2000, 2001
    - Notably Harabagiu, Moldovan et al. – SMU/UTD/LCC
  - AskMSR system stressed how much could be achieved by very simple methods with enough text (and now various copycats)
  - Middle ground is to use large collection of surface matching patterns (ISI)

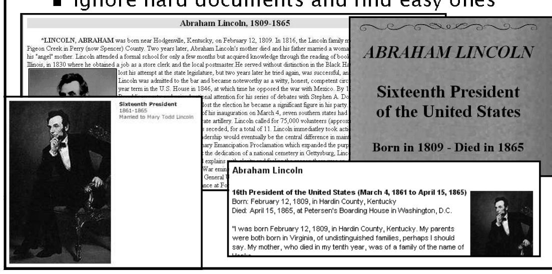
## AskMSR

- ***Web Question Answering: Is More Always Better?***
  - Dumais, Banko, Brill, Lin, Ng (Microsoft, MIT, Berkeley)
- **Q: "Where is the Louvre located?"**
- Want "Paris" or "France" or "75058 Paris Cedex 01" or a map
- <u>Don't</u> just want URLs

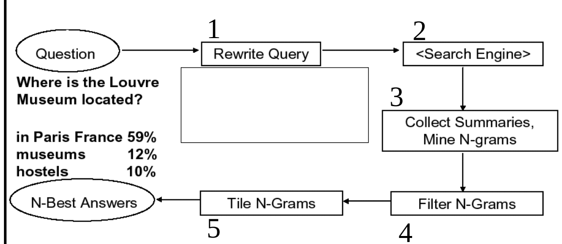



## AskMSR: Shallow approach

- *In what year did Abraham Lincoln die?*
- Ignore hard documents and find easy ones




## AskMSR: Details




## Step 1: Rewrite queries

- Intuition: The user's question is often syntactically quite close to sentences that contain the answer
  - Where <u>is</u> <u>the</u> <u>Louvre</u> <u>Museum</u> <u>located</u>?
  - <u>The</u> <u>Louvre</u> <u>Museum</u> <u>is</u> <u>located</u> in ***Paris***
  - Who <u>created</u> <u>the</u> <u>character</u> <u>of</u> <u>Scrooge</u>?
  - ***Charles Dickens*** <u>created</u> <u>the</u> <u>character</u> <u>of</u> <u>Scrooge</u>.

## Query rewriting

- Classify question into seven categories
  - <u>Who</u> is/was/are/were…?
  - <u>When</u> is/did/will/are/were …?
  - <u>Where</u> is/are/were …?

a. Category-specific transformation rules
   eg "For Where questions, move 'is' to all possible locations"
   "Where <u>is</u> the Louvre Museum located"
   → "<u>is</u> the Louvre Museum located"
   → "the <u>is</u> Louvre Museum located"
   → "the Louvre <u>is</u> Museum located"
   → "the Louvre Museum <u>is</u> located"
   → "the Louvre Museum located <u>is</u>"

Nonsense, but who cares? It's only a few more queries to Google.

b. Expected answer "Datatype" (eg, Date, Person, Location, …)
   <u>When</u> was the French Revolution? → DATE

- Hand-crafted classification/rewrite/datatype rules (Could they be automatically learned?)

## Query Rewriting - weights

- One wrinkle: Some query rewrites are more reliable than others

Where is the Louvre Museum located?

**Weight 1**
Lots of non-answers could come back too

**Weight 5**
if we get a match, it's probably right

+"the Louvre Museum is located"

+Louvre +Museum +located

## Step 2: Query search engine

- Send all rewrites to a Web search engine
- Retrieve top N answers (100?)
- For speed, rely just on search engine's "snippets", not the full text of the actual document

## Step 3: Mining N-Grams

- Unigram, bigram, trigram, … N-gram: list of N adjacent terms in a sequence
- Eg, "Web Question Answering: Is More Always Better"
  - Unigrams: Web, Question, Answering, Is, More, Always, Better
  - Bigrams: Web Question, Question Answering, Answering Is, Is More, More Always, Always Better
  - Trigrams: Web Question Answering, Question Answering Is, Answering Is More, Is More Always, More Always Betters

## Mining N-Grams

- Simple: Enumerate all N-grams (N=1,2,3 say) in all retrieved snippets
  - Use hash table and other fancy footwork to make this efficient
- Weight of an n-gram: occurrence count, each weighted by "reliability" (weight) of rewrite that fetched the document
- Example: "Who created the character of Scrooge?"
  - Dickens - 117
  - Christmas Carol - 78
  - Charles Dickens - 75
  - Disney - 72
  - Carl Banks - 54
  - A Christmas - 41
  - Christmas Carol - 45
  - Uncle - 31

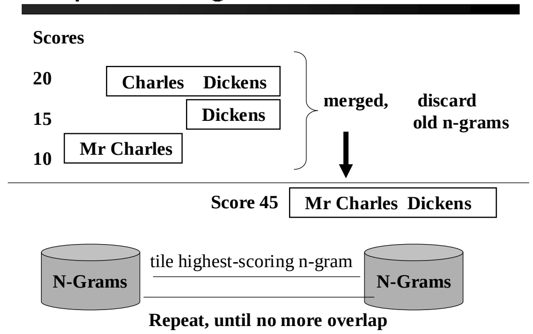
## Step 4: Filtering N-Grams

- Each question type is associated with one or more "**data-type filters**" = regular expression
- When…
- Where…
- What …
- Who …

**Date**
**Location**
**Person**

- Boost score of n-grams that do match regexp
- Lower score of n-grams that don't match regexp
- Details omitted from paper….

## Step 5: Tiling the Answers



## Results

- Standard TREC contest test-bed:
  ~1M documents; 900 questions
- Technique doesn't do too well (though would have placed in top 9 of ~30 participants!)
  - MRR = 0.262 (ie, right answered ranked about #4-#5 on average)
  - Why? Because it relies on the enormity of the Web!
- Using the Web as a whole, not just TREC's 1M documents... MRR = 0.42 (ie, on average, right answer is ranked about #2-#3)

## Issues

- In many scenarios (e.g., monitoring an individuals email...) we only have a small set of documents
- Works best/only for "Trivial Pursuit"-style fact-based questions
- Limited/brittle repertoire of
  - question categories
  - answer data types/filters
  - query rewriting rules

## ISI (Ravichandran and Hovy 2002): Surface patterns approach

- Use of Characteristic Phrases
- "When was <person> born"
  - Typical answers
    - "Mozart was born in 1756."
    - "Gandhi (1869-1948)..."
  - Suggests phrases like
    - "<NAME> was born in <BIRTHDATE>"
    - "<NAME> ( <BIRTHDATE>-"
  - as Regular Expressions can help locate correct answer

## Use Pattern Learning

- Example:
  - "The great composer Mozart (1756-1791) achieved fame at a young age"
  - "Mozart (1756-1791) was a genius"
  - "The whole world would always be indebted to the great music of Mozart (1756-1791)"
  - Longest matching substring for all 3 sentences is "Mozart (1756-1791)"
  - Suffix tree would extract "Mozart (1756-1791)" as an output, with score of 3
- Reminiscent of IE pattern learning

## Pattern Learning (cont.)

- Repeat with different examples of same question type
  - "Gandhi 1869", "Newton 1642", etc.
- Some patterns learned for BIRTHDATE
  - a. born in <ANSWER>, <NAME>
  - b. <NAME> was born on <ANSWER> ,
  - c. <NAME> ( <ANSWER> -
  - d. <NAME> ( <ANSWER> - )

## Experiments (Ravichandran and Hovy 2002)

- 6 different Question types
  - from Webclopedia QA Typology (Hovy et al., 2002a)
    - BIRTHDATE
    - LOCATION
    - INVENTOR
    - DISCOVERER
    - DEFINITION
    - WHY-FAMOUS

## Experiments: pattern precision

- BIRTHDATE table:
  - 1.0 <NAME> ( <ANSWER> - )
  - 0.85 <NAME> was born on <ANSWER>,
  - 0.6 <NAME> was born in <ANSWER>
  - 0.59 <NAME> was born <ANSWER>
  - 0.53 <ANSWER> <NAME> was born
  - 0.50 - <NAME> ( <ANSWER>
  - 0.36 <NAME> ( <ANSWER> -
- INVENTOR
  - 1.0 <ANSWER> invents <NAME>
  - 1.0 the <NAME> was invented by <ANSWER>
  - 1.0 <ANSWER> invented the <NAME> in

## Experiments (cont.)

- WHY-FAMOUS
  - 1.0 <ANSWER> <NAME> called
  - 1.0 laureate <ANSWER> <NAME>
  - 0.71 <NAME> is the <ANSWER> of
- LOCATION
  - 1.0 <ANSWER>'s <NAME>
  - 1.0 regional : <ANSWER> : <NAME>
  - 0.92 near <NAME> in <ANSWER>
- Depending on question type, get high MRR (0.6-0.9), with higher results from use of Web than TREC QA collection

## Shortcomings & Extensions

- Need for POS &/or semantic types
  - "Where are the Rocky Mountains?"
  - "Denver's new airport, topped with white fiberglass cones in imitation of the Rocky Mountains in <u>the background</u> , continues to lie empty"
  - <NAME> in <ANSWER>
- NE tagger &/or ontology could enable system to determine "background" is not a location

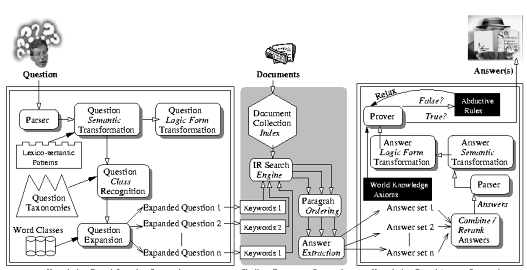
## Shortcomings... (cont.)

- Long distance dependencies
  - "Where is London?"
  - "London, which has one of the most busiest airports in the world, lies on the banks of the river Thames"
  - would require pattern like: <QUESTION>, (<any_word>)*, lies on <ANSWER>
  - Abundance & variety of Web data helps system to find an instance of patterns w/o losing answers to long distance dependencies

## Shortcomings... (cont.)

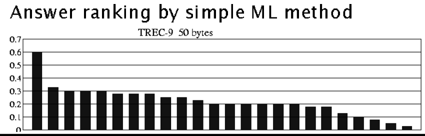
- System currently has only one anchor word
  - Doesn't work for Q types requiring multiple words from question to be in answer
    - "In which county does the city of Long Beach lie?"
    - "Long Beach is situated in Los Angeles County"
    - required pattern: <Q_TERM_1> is situated in <ANSWER> <Q_TERM_2>
- Does not use case
  - "What is a micron?"
  - "...a spokesman for Micron, <u>a maker of semiconductors</u>, said SIMMs are..."
- If Micron had been capitalized in question, would be a perfect answer

## LCC: Harabagiu, Moldovan et al.

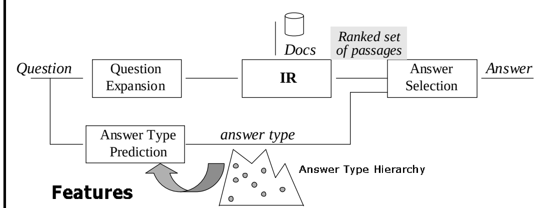


## Value from sophisticated NLP – Pasca and Harabagiu (2001)

- Good IR is needed: SMART paragraph retrieval
- Large taxonomy of question types and expected answer types is crucial
- Statistical parser used to parse questions and relevant text for answers, and to build KB
- Query expansion loops (morphological, lexical synonyms, and semantic relations) important
- Answer ranking by simple ML method




## *Answer types in State-of-the-art QA systems*




**Features**

- Answer type
  - Labels questions with answer type based on a taxonomy
  - Classifies questions (e.g. by using a maximum entropy model)

## QA Typology (from ISI USC)

- Typology of typical Q forms—94 nodes (47 leaf nodes)
- Analyzed 17,384 questions (from answers.com)

```
(THING
 ((AGENT
   (NAME (FEMALE-FIRST-NAME (EVE MARY ...))
         (MALE-FIRST-NAME (LAWRENCE SAM ...))))
         (COMPANY-NAME (BOEING AMERICAN-EXPRESS))
         JESUS ROMANOFF ...)
   (ANIMAL-HUMAN (ANIMAL (WOODCHUCK YAK ...))
                 PERSON)
   (ORGANIZATION (SQUADRON DICTATORSHIP ...))
   (GROUP-OF-PEOPLE (POSSE CHOIR ...))
   (STATE-DISTRICT (TIROL MISSISSIPPI ...))
   (CITY (ULAN-BATOR VIENNA ...))
   (COUNTRY (SULTANATE ZIMBABWE ...))))
 (PLACE
   (STATE-DISTRICT (CITY COUNTRY...))
   (GEOLOGICAL-FORMATION (STAR CANYON...))
   AIRPORT COLLEGE CAPITOL ...)
 (ABSTRACT
  (LANGUAGE (LETTER-CHARACTER (A B ...)))
  (QUANTITY
   (NUMERICAL-QUANTITY INFORMATION-QUANTITY
   MASS-QUANTITY MONETARY-QUANTITY
   TEMPORAL-QUANTITY ENERGY-QUANTITY
   TEMPERATURE-QUANTITY ILLUMINATION-QUANTITY
   (SPATIAL-QUANTITY
    (VOLUME-QUANTITY AREA-QUANTITY DISTANCE-QUANTITY)) ...
     PERCENTAGE)))
 (UNIT
  ((INFORMATION-UNIT (BIT BYTE ... EXABYTE))
   (MASS-UNIT (OUNCE ...)) (ENERGY-UNIT (BTU ...))
   (CURRENCY-UNIT (ZLOTY PESO ...))
   (TEMPORAL-UNIT (ATTOSECOND ... MILLENIUM))
   (TEMPERATURE-UNIT (FAHRENHEIT KELVIN CELCIUS))
   (ILLUMINATION-UNIT (LUX CANDELA))
   (SPATIAL-UNIT
    ((VOLUME-UNIT (DECILITER ...))
     (DISTANCE-UNIT (NANOMETER ...))))
     (AREA-UNIT (ACRE)) ... PERCENT))
 (TANGIBLE-OBJECT
  ((FOOD (HUMAN-FOOD (FISH CHEESE ...)))
   (SUBSTANCE
    ((LIQUID (LEMONADE GASOLINE BLOOD ...))
     (SOLID-SUBSTANCE (MARBLE PAPER ...))
     (GAS-FORM-SUBSTANCE (GAS AIR)) ...))
   (INSTRUMENT (DRUM DRILL (WEAPON (ARM GUN)) ...)
   (BODY-PART (ARM HEART ...))
   (MUSICAL-INSTRUMENT (PIANO)))
   ... *GARMENT *PLANT DISEASE)
```

## *Extracting Answers for Factoid Questions*

- In TREC 2003 the LCC QA system extracted 289 correct answers for factoid questions
- The Name Entity Recognizer was responsible for 234 of them

| | | | | | |
|---|---|---|---|---|---|
| QUANTITY | 55 | ORGANIZATION | 15 | PRICE | 3 |
| NUMBER | 45 | AUTHORED WORK | 11 | SCIENCE NAME | 2 |
| DATE | 35 | PRODUCT | 11 | ACRONYM | 1 |
| PERSON | 31 | CONTINENT | 5 | ADDRESS | 1 |
| COUNTRY | 21 | PROVINCE | 5 | ALPHABET | 1 |
| OTHER LOCATIONS | 19 | QUOTE | 5 | URI | 1 |
| CITY | 19 | UNIVERSITY | 3 | | |

## *Special Case of Names*

*Questions asking for names of authored works*

| |
|---|
| 1934: What is the play "West Side Story" based on?<br>Answer: Romeo and Juliet |
| 1976: What is the motto for the Boy Scouts?<br>Answer: Be Prepared |
| 1982: What movie won the Academy Award for best picture in 1989?<br>Answer: Driving Miss Daisy |
| 2080: What peace treaty ended WWI?<br>Answer: Versailles |
| 2102: What American landmark stands on Liberty Island?<br>Answer: Statue of Liberty |

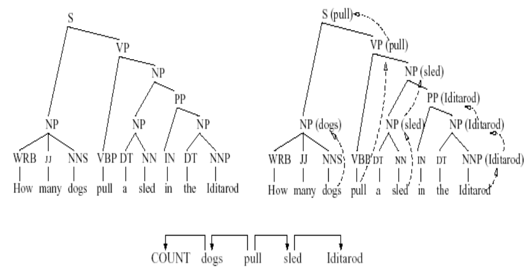
## *NE-driven QA*

- The results of the past 5 TREC evaluations of QA systems indicate that current state-of-the-art QA is determined by the recognition of Named Entities:
  - *Precision of recognition*
  - *Coverage of name classes*
  - *Mapping into concept hierarchies*
  - *Participation into semantic relations (e.g. predicate-argument structures or frame semantics)*

## *Concept Taxonomies*

- For 29% of questions the QA system relied on an off-line taxonomy with semantic classes such as:
  - Disease
  - Drugs
  - Colors
  - Insects
  - Games
- *The majority of these semantic classes are also associated with patterns that enable their identification*

**Syntax to Logical Forms**




- Syntactic analysis plus semantic -> logical form
- Mapping of question and potential answer LFs to find the best match

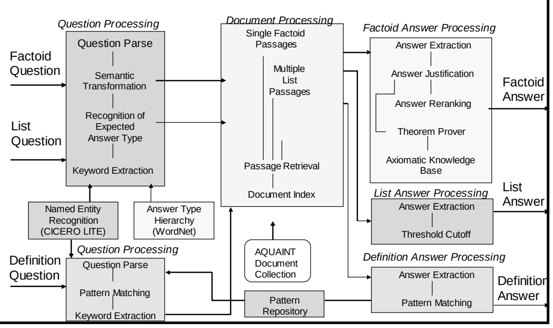
## Abductive inference

- System attempts inference to justify an answer (often following lexical chains)
- Their inference is a kind of funny middle ground between logic and pattern matching
- But quite effective: 30% improvement
- *Q: When was the internal combustion engine invented?*
- *A: The first internal-combustion engine was built in 1867.*
- invent -> create_mentally -> create -> build

## Question Answering Example

- How hot does the inside of an active volcano get?
- get(TEMPERATURE, inside(volcano(active)))
- "lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit"
- fragments(lava, TEMPERATURE(degrees(300)),
  belched(out, mountain))
  - volcano ISA mountain
  - lava ISPARTOF volcano
  - lava inside volcano
  - fragments of lava HAVEPROPERTIESOF lava
- The needed semantic information is in WordNet definitions, and was successfully translated into a form that was used for rough 'proofs'

## *The Architecture of LCC's QA System around 2003*




## *Definition Questions*

- They asked about:
  - PEOPLE (most of them starting with "Who")
  - other types of NAMES
  - general concepts
- People questions
  - Many use the PERSON name in the format [First name, Last name]
    - examples: Aaron Copland, Allen Iverson, Albert Ghiorso
  - Some names had the PERSON name in format [First name, Last name1, Last name2]
    - example: Antonia Coello Novello
  - Other names had the name as a single word → very well known person
    - examples: Nostradamus, Absalom, Abraham
  - Some questions referred to names of kings or princes:
    - examples: Vlad the Impaler, Akbar the Great

## *Answering definition questions*

- Most QA systems use between 30-60 patterns
- The most popular patterns:

| Id | Pattern | Freq. | Usage | Question |
|---|---|---|---|---|
| 25 | person-hyponym QP | 0.43% | The doctors also consult with former Italian Olympic skier Alberto Tomba, along with other Italian athletes | 1907: Who is Alberto Tomba? |
| 9 | QP, the AP | 0.28% | Bausch Lomb, the company that sells contact lenses, among hundreds of other optical products, has come up with a new twist on the computer screen magnifier | 1917: What is Bausch & Lomb? |
| 11 | QP, a AP | 0.11% | ETA, a Basque language acronym for Basque Homeland and Freedom _ has killed nearly 800 people since taking up arms in 1968 | 1987: What is ETA in Spain? |
| 13 | QA, an AP | 0.02% | The kidnappers claimed they are members of the Abu Sayaf, an extremist Muslim group, but a leader of the group denied that | 2042: Who is Abu Sayaf? |
| 21 | AP such as QP | 0.02% | For the hundreds of Albanian refugees undergoing medical tests and treatments at Fort Dix, the news is mostly good: Most are in reasonable good health, with little evidence of infectious diseases such as TB | 2095: What is TB? |

## *Complex questions*

- Characterized by the need of domain knowledge
- There is no single answer type that can be identified, but rather an answer structure needs to be recognized
- Answer selection becomes more complicated, since inference based on the semantics of the answer type needs to be activated
- Complex questions need to be decomposed into a set of simpler questions

## *Example of Complex Question*

*How have thefts impacted on the safety of Russia's nuclear navy, and has the theft problem been increased or reduced over time?*

**Need of domain knowledge**

*To what degree do different thefts put nuclear or radioactive materials at risk?*

**Question decomposition**

***Definition questions:***
- *What is meant by nuclear navy?*
- *What does 'impact' mean?*
- *How does one define the increase or decrease of a problem?*

***Factoid questions:***
- *What is the number of thefts that are likely to be reported?*
- *What sort of items have been stolen?*

***Alternative questions:***
- *What is meant by Russia? Only Russia, or also former Soviet facilities in non-Russian republics?*

## *Semantic inference for Q/A*

- The problem of classifying questions
  - E.g. "manner questions":
  - Example *"How did Hitler die?"*
- The problem of recognizing answer types/structures
  - Should *"manner of death"* by considered an answer type?
  - What other manner of event/action should be considered as answer types?
- The problem of extracting/justifying/ generating answers to complex questions
  - Should we learn to extract "manner" relations?
  - What other types of relations should we consider?
  - Is relation recognition sufficient for answering complex questions? Is it necessary?

## *Manner-of-death*

In previous TREC evaluations 31 questions asked about manner of death:
  - *"How did Adolf Hitler die?"*
- State-of-the-art solution (LCC):
  - *Considered "Manner-of-Death" as an answer type, pointing to a variety of verbs and nominalizations encoded in WordNet*
  - Developed text mining techniques for identifying such information based on lexico-semantic patterns from WordNet
  - Example:
    - [*kill #sense1 (verb)* – *CAUSE* → *die #sense1 (verb)*]
  - Source of the troponyms of the [*kill #sense1 (verb)*] concept are candidates for the MANNER-OF-DEATH hierarchy
    - e.g., drown, poison, strangle, assassinate, shoot

## *Practical Hurdle*

- Not all MANNER-OF-DEATH concepts are lexicalized as a verb

  → set out to determine additional patterns that capture such cases
- *Goal:* (1) set of patterns

  (2) dictionaries corresponding to such patterns

  → well known IE technique: (IJCAI'99, Riloff&Jones)

| | |
|---|---|
| X {DIE / be killed} in ACCIDENT | seed: (ACCIDENT) train, accident, car wreck |
| X {DIE / be killed} {from\|of} DISEASE | seed: (DISEASE) cancer AIDS |
| X {DIE} {after suffering / suffering of} MEDICAL CONDITION | seed: (ACCIDENT) stroke, complications caused by diabetes |

- *Results: 100 patterns were discovered*

## *Applying Frame Structures to QA*

- Parsing Questions

***Q**: What kind of materials were stolen from the Russian navy?*

*FS(Q): What [GOODS: kind of nuclear materials] were [Target-Predicate:stolen] [VICTIM: from the Russian Navy]?*

- Parsing Answers

***A(Q)**: Russia's Pacific Fleet has also fallen prey to nuclear theft; in 1/96, approximately 7 kg of HEU was reportedly stolen from a naval base in Sovetskaya Gavan.*

*FS(A(Q)): [VICTIM(P1): Russia's Pacific Fleet] has also fallen prey to [Goods(P1): nuclear ] [Target-Predicate(P1): theft]; in 1/96, [GOODS(P2): approximately 7 kg of HEU] was reportedly [Target-Predicate (P2): stolen] [VICTIM (P2): from a naval base] [SOURCE(P2): in Sovetskawa Gavan]*

## *Additional types of relations*

- *Temporal relations*
- *Causal relations*
- *Evidential relations*
- *Part-whole relations*

- ... plenty of research to be done to be able to answer these kinds of questions!

## References


- **AskMSR: Question Answering Using the Worldwide Web**
  - Michele Banko, Eric Brill, Susan Dumais, Jimmy Lin
  - http://www.ai.mit.edu/people/jimmylin/publications/Banko-etal-AAAI02.pdf
  - In Proceedings of 2002 AAAI SYMPOSIUM on Mining Answers from Text and Knowledge Bases, March 2002
- **Web Question Answering: Is More Always Better?**
  - Susan Dumais, Michele Banko, Eric Brill, Jimmy Lin, Andrew Ng
  - http://research.microsoft.com/~sdumais/SIGIR2002-QA-Submit-Conf.pdf
- D. Ravichandran and E.H. Hovy. 2002. **Learning Surface Patterns for a Question Answering System.** ACL conference, July 2002.


## References


S. Harabagiu, D. Moldovan, M. Paşca, R. Mihalcea, M. Surdeanu, R. Bunescu, R. Gîrju, V.Rus and P. Morărescu. *FALCON: Boosting Knowledge for Answer Engines*. The Ninth Text REtrieval Conference (TREC 9), 2000.

**Marius Pasca and Sanda Harabagiu, High Performance Question/Answering, in *Proceedings of the 24th Annual International ACL SIGIR Conference on Research and Development in Information Retrieval (SIGIR-2001)*, September 2001, New Orleans LA, pages 366-374.**

L. Hirschman, M. Light, E. Breck and J. Burger. *Deep Read: A Reading Comprehension System*. In Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics, 1999.

C. Kwok, O. Etzioni and D. Weld. *Scaling Question Answering to the Web*. ACM Transactions in Information Systems, Vol 19, No. 3, July 2001, pages 242-262.

M. Light, G. Mann, E. Riloff and E. Breck. *Analyses for Elucidating Current Question Answering Technology*. Journal of Natural Language Engineering, Vol. 7, No. 4 (2001).

M. M. Soubbotin. *Patterns of Potential Answer Expressions as Clues to the Right Answers*. Proceedings of the Tenth Text REtrieval Conference (TREC 2001).