

Text-based Question Answering systems

Christopher Manning
CS224N/Ling 284

(includes slides borrowed from Sanda Harabagiu,
Roxana Girju)

Information overload

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, 2002.

Modern QA 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
- **Question Answering:** give the user a (short) answer to their question, perhaps supported by evidence.

Question Answering: IBM's Watson

- Won Jeopardy on February 16, 2011!

WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVIA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL



Bram Stoker



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 Quintex 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.)?

People *want* to ask questions...

Examples from AltaVista query log (late 1990s)

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% of early query logs

A Brief (Academic) History

- 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: e.g., LUNAR system
 - There's still Microsoft English Query
 - Spoken dialog systems
 - Currently very active and commercially relevant

A Brief (Academic) History

- Focusing on open-domain QA is newer
 - MURAX (Kupiec 1993): Encyclopedia answers
 - Hirschman: Reading comprehension tests
 - TREC QA competition: 1999–
- But not really new either: Simmons et al. 1970 *CACM*
 - http://brenocon.com/simmons70overview_qa.pdf
 - Take an encyclopedia and load it onto a computer
 - Take a question and parse it into a logical form
 - Perform simple information retrieval to get relevant texts, parse those into a logical form, match and rank
 - What do worms eat? **Worms eat ???**
 - Candidates
 - Worms eat grass
 - Grass is eaten by worms
 - Birds eat worms

Question Answering at TREC

- Question answering competition at TREC consisted 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.
 - 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 was introduced.

The TREC Document Collection

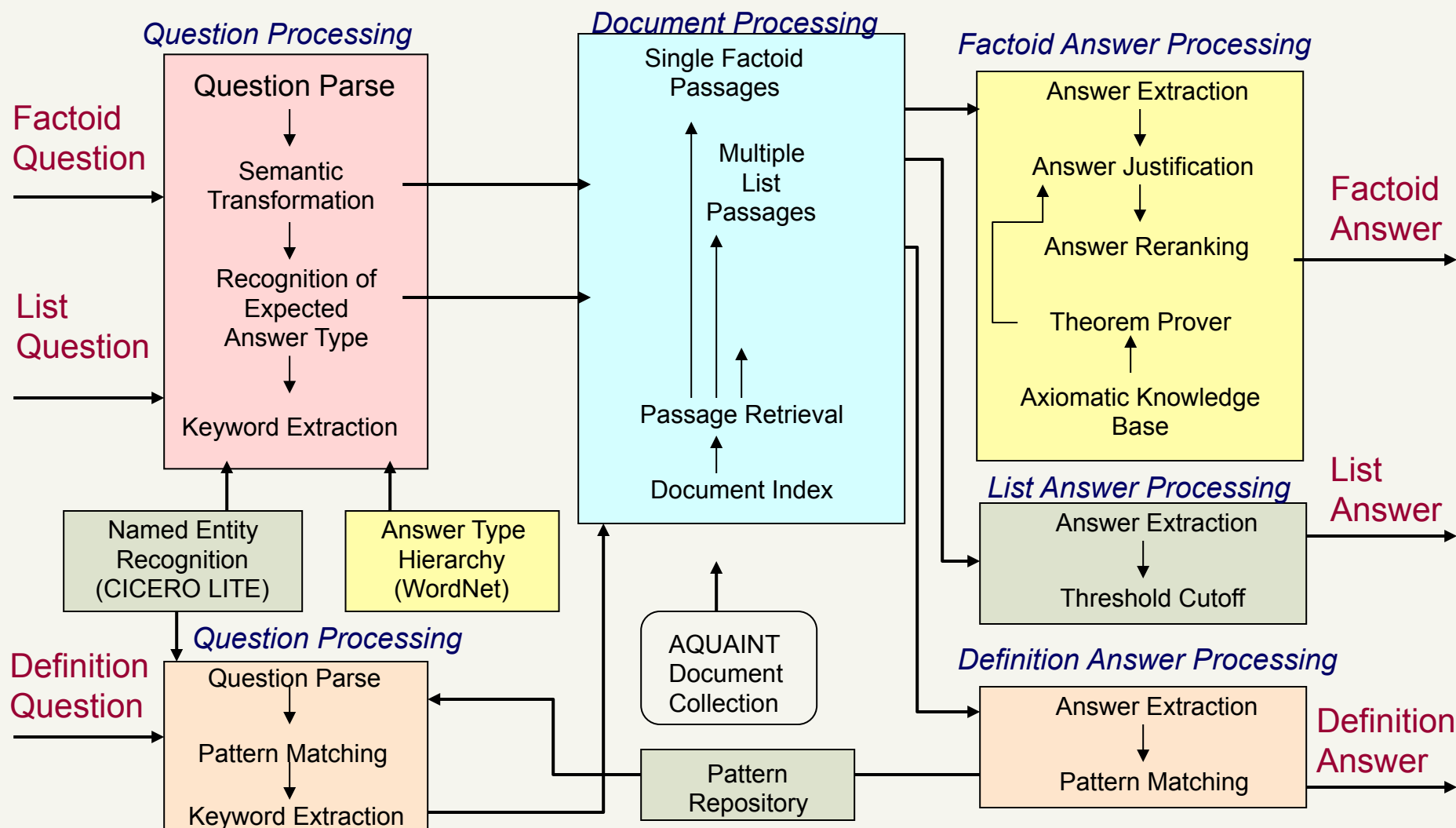
- The collection used news articles from the following sources:
 - AP newswire, 1998-2000
 - New York Times newswire, 1998-2000
 - Xinhua News Agency newswire, 1996-2000
- In total there are 1,033,461 documents in the collection. 3GB of text.
- This was a lot of text to process entirely using advanced NLP techniques so the systems usually consisted of an initial information retrieval phase followed by more advanced processing.
 - You'd probably do this differently in 2012
- Many supplemented this text with use of the web, and other knowledge bases, especially as time passed

Top Performing Systems

- The best performing TREC systems could 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)

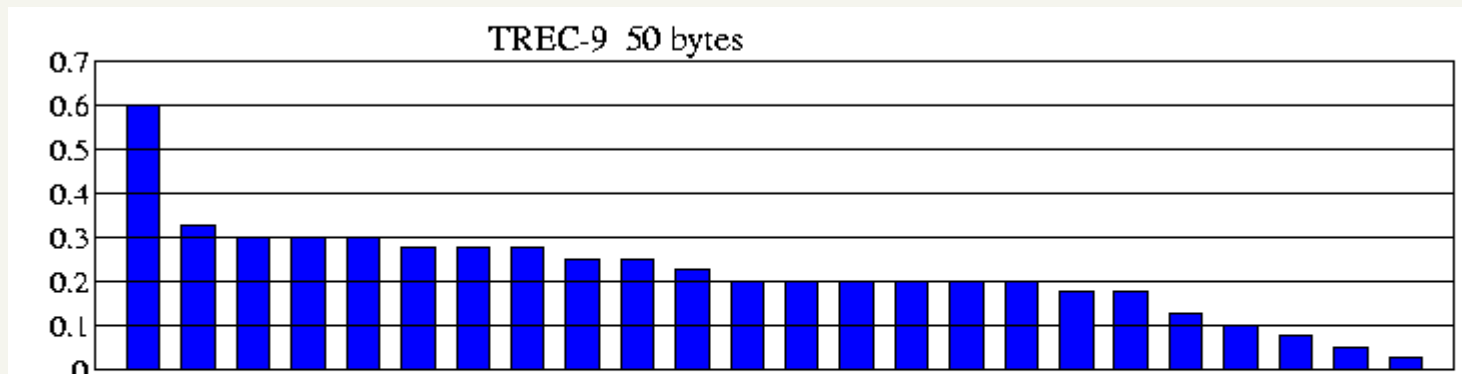
Full NLP QA: LCC (Harabagiu/Moldovan)

[below is the Architecture of LCC's QA system circa 2003]

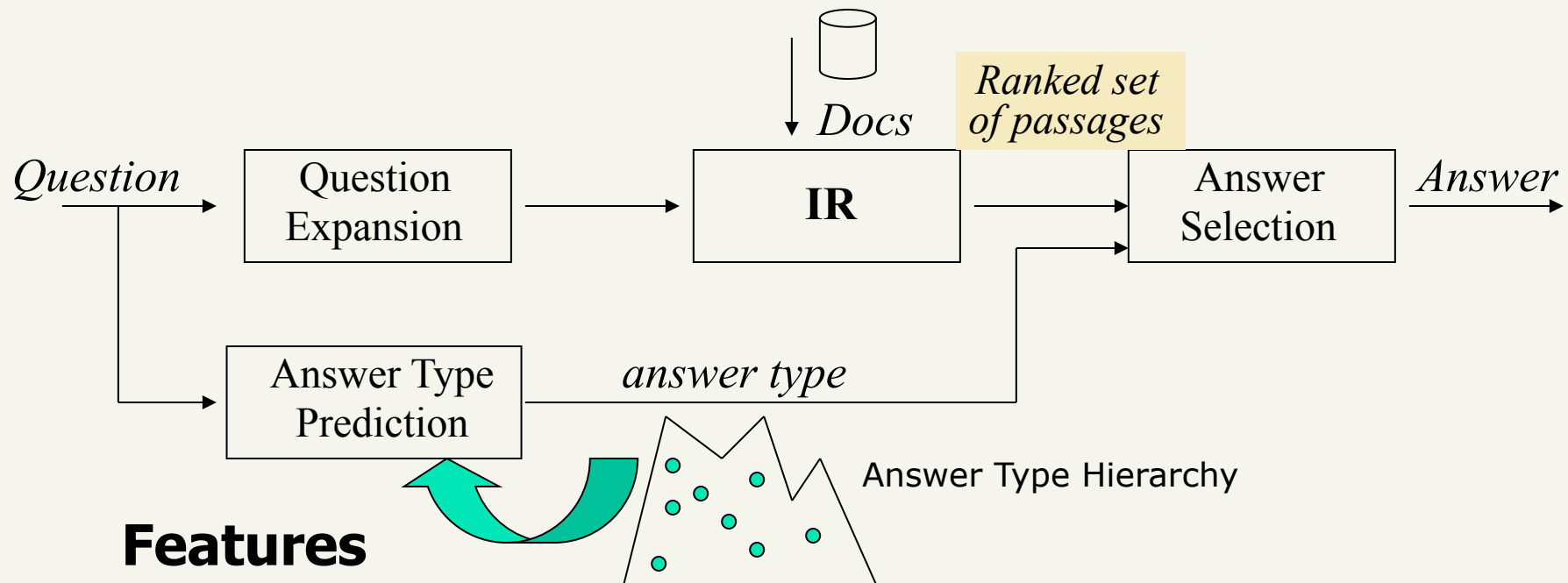


Value from sophisticated NLP – Pasca and Harabagiu (2001)

- 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
- Further value comes from deeper NLP and inferencing



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



Features

◆ Answer type

- Labels questions with answer type based on a taxonomy
 - Person, location, weight, temperature, year, vehicle
- Classifies questions (e.g. by using a maximum entropy model)

Answer Types

- Of course, determining the answer type isn't that easy...
 - **Who** questions can have organizations as answers
 - Who sells the most hybrid cars?
 - **Which** questions can have people as answers
 - Which president went to war with Mexico?

Lexical Terms Extraction as input to Information Retrieval

- Questions approximated by sets of words (lexical terms)
- Similar to bag-of-word IR models: but choose nominal non-stop words and verbs

Question (from TREC QA track)	Lexical terms
Q002: What was the monetary value of the Nobel Peace Prize in 1989?	monetary, value, Nobel, Peace, Prize, 1989
Q003: What does the Peugeot company manufacture?	Peugeot, company, manufacture
Q004: How much did Mercury spend on advertising in 1993?	Mercury, spend, advertising, 1993

Keyword Selection Algorithm

1. Select all non-stopwords in quotations
2. Select all NNP words in recognized named entities
3. Select all complex nominals with their adjectival modifiers
4. Select all other complex nominals
5. Select all nouns with adjectival modifiers
6. Select all other nouns
7. Select all verbs
8. Select the answer type word

Passage Extraction Loop

- Passage Extraction Component
 - Extracts passages that contain all selected keywords
 - Passage size dynamic
 - Start position dynamic
- Passage quality and keyword adjustment
 - In the first iteration use the first 6 keyword selection heuristics
 - If the number of passages is lower than a threshold \Rightarrow query is too strict \Rightarrow drop a keyword
 - If the number of passages is higher than a threshold \Rightarrow query is too relaxed \Rightarrow add a keyword

Passage Scoring

- Passage ordering is performed using a sort that involves three scores:
 - The number of words from the question that are recognized in the same sequence in the window
 - The number of words that separate the most distant keywords in the window
 - The number of unmatched keywords in the window

Rank candidate answers in retrieved passages

Q066: Name the first private citizen to fly in space.

- Answer type: **Person**

- Text passage:

“Among them was Christa McAuliffe, the first private citizen to fly in space. Karen Allen, best known for her starring role in “Raiders of the Lost Ark”, plays McAuliffe. Brian Kerwin is featured as shuttle pilot Mike Smith...”

- Best candidate answer: **Christa McAuliffe**

Extracting Answers for Factoid Questions: NER!

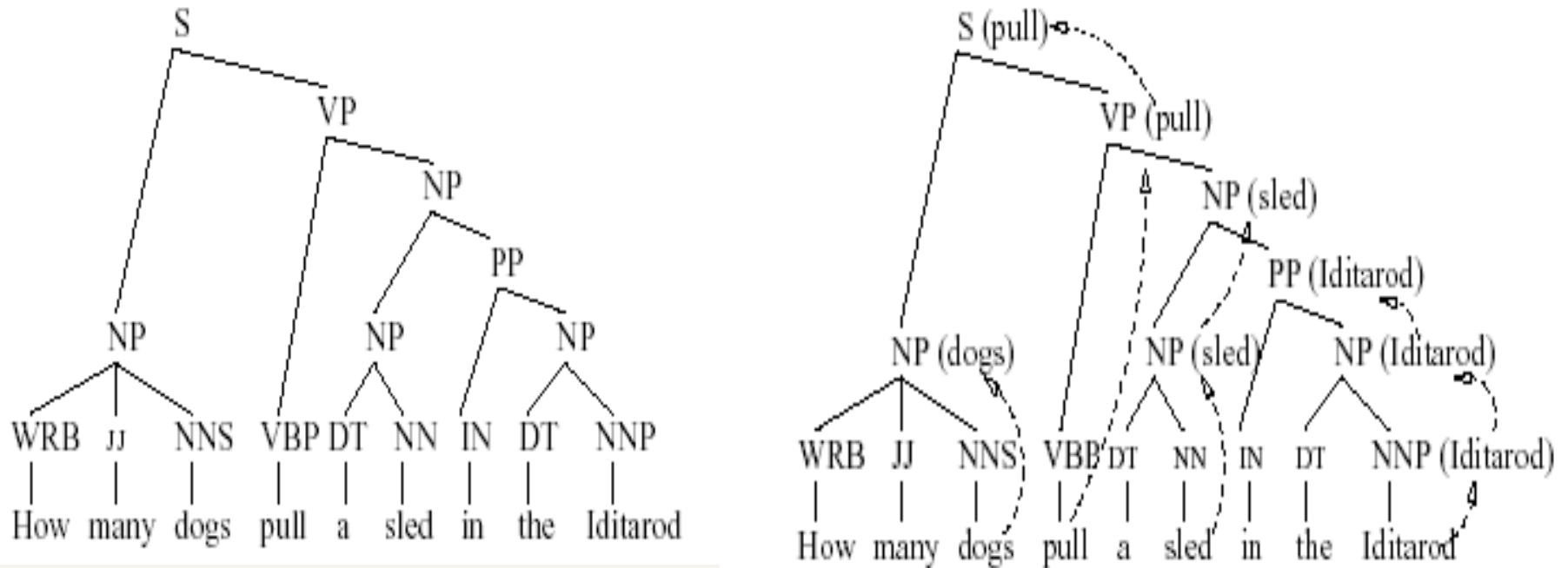
- In TREC 2003 the LCC QA system extracted 289 correct answers for factoid questions (out of 413)
- The Name Entity Recognizer was responsible for 234 of them
 - Current QA is largely based on the high accuracy recognition of Named Entities

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		

Semantics and Reasoning for QA: Predicate-argument structure

- Q336: *When was Microsoft established?*
- This question is **difficult** because Microsoft tends to establish lots of things...
Microsoft plans to establish manufacturing partnerships in Brazil and Mexico in May.
- Need to be able to detect sentences in which ‘Microsoft’ is **object** of ‘establish’ or close synonym.
- Matching sentence:
Microsoft Corp was founded in the US in 1975, incorporated in 1981, and established in the UK in 1982.
- Requires analysis of sentence syntax/semantics!

Semantics and Reasoning for QA: Syntax to Logical Forms



COUNT dogs pull sled Iditarod

- Syntactic analysis plus semantic → logical form (dependencies)
- 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 very 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

Q: How hot does the inside of an active **volcano** get?

- `get(TEMPERATURE, inside(volcano(active)))`

A: “**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’

QA beyond TREC

- Answers to complex questions that require a longer answer
 - *What is a PIC Freeze?*
 - *Can I travel with Ameripass in Mexico?*
 - *How do you change the spawn point in Minecraft?*

Not all problems are solved yet!

- Where do lobsters like to live?
 - on a Canadian airline
- Where are zebras most likely found?
 - near dumps
 - in the dictionary
- Why can't ostriches fly?
 - Because of American economic sanctions
- What's the population of Mexico?
 - Three
- What can trigger an allergic reaction?
 - ..something that can *trigger* an allergic reaction

References

- R. F. Simmons, Natural language question-answering systems: 1969. Communications of the ACM. Volume 13, 1970. http://brenocon.com/simmons70overview_qa.pdf
- **AskMSR: Question Answering Using the Worldwide Web**
 - Michele Banko, Eric Brill, Susan Dumais, Jimmy Lin
 - In Proceedings of 2002 AAAI SYMPOSIUM on Mining Answers from Text and Knowledge Bases, March 2002
- **Web Question Answering: Is More Always Better?**
 - S. Dumais, M. Banko, E. Brill, J. Lin, A. Ng
- Marius Pasca and Sanda Harabagiu, **High Performance Question/Answering**, in *ACM SIGIR-2001*, September 2001, New Orleans LA, pages 366-374.