Natural Language Processing

Question Answering

Dan Klein – UC Berkeley

The following slides are largely from Chris Manning, including many slides originally from Sanda Harabagiu, ISI, and Nicholas Kushmerick.
Large-Scale NLP: Watson

"a camel is a horse designed by"

Wiktionary

Entry Discussion Read Edit History Search

a camel is a horse designed by a committee

A camel is a horse designed by committee

Posted by Ruben P. Mendez on April 16, 2004

Does anyone know the origin of this maxim? I heard it way back at the United Nations, which is chockfull of committees. It may have originated there, but I'd like an authoritative explanation. Thanks

- Re: A camel is a horse designed by committee SR 16/April/04
  - Re: A camel is a horse designed by committee Henry 18/April/04
QA vs Search
People want to ask questions?

Examples of search queries

- 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?
- 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
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
  - Spoken dialog systems
    - Currently very active and commercially relevant
- The focus on open-domain QA is (relatively) new
  - MURAX (Kupiec 1993): Encyclopedia answers
  - Hirschman: Reading comprehension tests
  - TREC QA competition: 1999–
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 a notion of confidence has been introduced.
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, still do well
    - 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)
  - Emerging standard: analysis, soft-matching, abduction
Pattern Induction: ISI
Webclopedia Architecture

IR
- Steps: create query from question (WordNet-expand)
  retrieve top 1000 documents
- Engines: MG (Sydney)—(Lin)
  AT&T (TREC)—(Lin)

Segmentation
- Steps: segment each document into topical segments
- Engines: fixed-length (not used)
  TextTiling (Hearst 94)—(Lin)
  C99 (Choi 00)—(Lin)
  MAXNET (Lin 00; not used)

Ranking
- Steps: score each sentence in each segment, using WordNet expansion
  rank segments
- Engines: FastFinder (Junk)

Matching
- Steps: match general constraint patterns against parse trees
  match desired semantic type against parse tree elements
  match desired words against words in sentences
- Engines: matcher (Junk)

Ranking and answer extraction
- Steps: rank candidate answers
  extract and format them
- Engines: part of matcher (Junk)

Question parsing
- Steps: parse question
  find desired semantic type
- Engines: Identifinder (BBN)
  CONTEXT (Hemjakob)

Segment Parsing
- Steps: parse segment sentences
- Engines: CONTEXT (Hemjakob)

Match segments against question
Rank and prepare answers
Output answers

QA typology
- Categorize QA types in taxonomy (Gerber)

Constraint patterns
- Identify likely answers in relation to other parts of the sentence (Gerber)
Please wait while Wikipedia answers your question (potential answers will follow, with the top 10 below).

At the time, these were the best 10 answers found for the question:

1. 505 4252 When John Lennon died in 1980, the traumatic childhood experiences which took place 13 years earlier on December 3, 1967 finally made sense to me. (http://www.onceupon.com/author.htm)
2. 505 4252 When John Lennon died in 1980, the traumatic childhood experiences which took place 13 years earlier on December 3, 1967 finally made sense to me. (http://www.onceupon.com/author.htm)
3. 504 1176 John Lennon died in 1980, which is part of that era. (http://www.johnlennon.it/guestbook1.htm)
4. 504 1176 As many people know John Lennon died in 1980 and she has not remarried since. (http://www.public.ox.ac.uk/~depeus/212erkie/yoko/yoko.htm)
5. 504 1176 Fourth: Beatles John Lennon died in 1980 when he was shot by a fan outside his New York city home. (http://news.bbc.co.uk/2/hi/entertainment/2275173.stm)
6. 504 1176 John Lennon died in 1980. (http://members.tripod.com/~FenecGirl/interview.htm)
7. 504 1176 As many people know John Lennon died in 1980 and she has not remarried since. (http://www.public.ox.ac.uk/~depeus/212erkie/yoko/yoko.htm)
8. 504 1176 Fourth: Beatles John Lennon died in 1980 when he was shot by a fan outside his New York city home. (http://news.bbc.co.uk/2/hi/entertainment/2275173.stm)
Who was the prime minister of Australia in 1990?

Please wait while Wikipedia answers your question (potential answers will follow, with the top 10 below).

Current top 10 (of 109) for "Who was the prime minister of Australia in 1990?" - still finding more...

2. 352.7734... sound file Prime Minister BG Menzies opens "Australia Calling". 1939, ... 1950 BACK. Image file Ian McLachlan, presenter of Australia All Over, c. 1990. (http://www.abc.net.au/gallery/gallery.htm)
3. 352.0589 John Howard is the prime minister of Australia. Niall is something he unfortunately stepped on, though he managed to scope most of it off on the curb. (http://spiegel.de/-/dfc/2004/01/20/comments.php?entry_id=9135)
4. 352.0589 Mr. Howard is the prime minister of Australia. (http://www.politician.com/Politics/Feature.html?FeatID=11000323)
5. 347.5493 Visit to Japan by Prime Minister John Winston Howard of Australia (July 2005) (http://www.mofa.go.jp/region/asia-pacific/australia/)
7. 396.8627 Prime Minister Junichiro Koizumi held talks with Mr. John Howard, the Prime Minister of Australia, at the Prime Minister's Official Residence. (http://www.kantei.go.jp/foreign/koizumimophoto/2002/07/16/20020716_04167.html)
8. 394.0130 Summit Meeting Between Prime Minister Junichiro Koizumi and Prime Minister John Winston Howard of Australia (Summary) (May 2002) (http://www.mofa.go.jp/region/asia-pacific/australia/)
9. 394.0130 Summit Meeting Between Prime Minister Junichiro Koizumi and Prime Minister John Winston Howard of Australia (Summary) (May 2002) (http://www.mofa.go.jp/region/asia-pacific/australia/)
10. 394.0130 Summit Meeting Between Prime Minister Junichiro Koizumi and Prime Minister John Winston Howard of Australia (Summary) (May 2002) (http://www.mofa.go.jp/region/asia-pacific/australia/)
Searching for answer sentences on the subject: gum arabic

Expecting an answer of type: Q-COMPONENTS, S-NP (0.8)

View analysis of question:

3. 37 0068 From a North African coffee a sap was extracted and boiled to make an alcohol-based gum arabic. (about 2 tablespoons) 1/2 cup water 2 teaspoons honey. (http://www.corn.com/char Elias/solutions/home/1579)

4. 34 2060 You have your gum arabic solution made in five minutes of not so hard work, and all you need now is to allow some time for all the durt to precipitate to the bottom. (http://www.micropcap-uk.org.uk/image/straw03/3wpdpart3c.html)

5. 34 2060 You have your gum arabic solution made in five minutes of not so hard work, and all you need now is to allow some time for all the durt to precipitate to the bottom (perhaps overnight). 3 Decant the supernatant liquid and adjust the density to suit your preferences. (http://www.micropcap-uk.org.uk/image/straw03/3wpdpart3c.html)

6. 34 0066 Unfortunately, dried (crystallized) pure gum arabic is very brittle. prints made with a high proportion of gum binder often lift easily (redissolve) from the paper, but also will bronze (appear shiny or leathery) and crack or flake if applied in a thick or massivene layer. (http://www.newssl.com/gumabrac/colour_water_gelatine/index.html)

7. 34 0219 Made from the dried sap of the Sahelian acacia tree, gum arabic sales totaled two-thirds of all U.S. imports from Sudan in 1996. (http://www.pcp.org/oci.cifm?ImageAndID=1084&sectionID=8900038&contentID=251S0)

8. 33 3265 gum arabic or gum acacia. - Hardened sap... gyaotica. In Japanese tradition, a relief print made from an actual fish, and sometimes from a shell leaf, or other... (http://www.saltex.com/ArtFo/On.html)

9. 33 1687 gum arabic n. gum from the acacia tree, used as a thickeners (especially in candies and pharmaceuticals) [syn: gum acacia] (http://bites.torrent.com/gum%20arabic/)

10. 22 4889 Sugro-coated confections made by the paning process employ gum arabic to provide an adhesive and film coating for nuts, candy corn, jelly beans, bridge mix, and others (http://www.jumbo.com/application.html)
Ravichandran and Hovy 2002
Learning Surface Patterns

- 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>-"
  - Regular expressions
Use Pattern Learning

- **Example: Start with “Mozart 1756”**
  - **Results:**
    - “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> - )
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
Pattern Precision

- **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 the background, continues to lie empty"
  - <NAME> in <ANSWER>

- **Long distance dependencies**
  - "Where is London?"
  - "London, which has one of the busiest airports in the world, lies on the banks of the river Thames"
  - would require pattern like:
    <QUESTION>, (<any_word>)*, lies on <ANSWER>
  - But: abundance of Web data compensates
Aggregation: 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
- Don’t just want URLs
AskMSR: Shallow approach

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

1. Rewrite Query
2. <Search Engine>
3. Collect Summaries, Mine N-grams
4. Filter N-Grams
5. Tile N-Grams

Question
Where is the Louvre Museum located?

in Paris France 59%
museums 12%
hostels 10%

N-Best Answers
Step 1: Rewrite queries

- Intuition: The user’s question is often syntactically quite close to sentences that contain the answer
  
  - Where is the Louvre Museum located?
  
  - The Louvre Museum is located in *Paris*
  
  - Who created the character of Scrooge?
  
  - *Charles Dickens* created the character of Scrooge.
Query Rewriting: Variations

- Classify question into seven categories
  - **Who** is/was/are/were...?
  - **When** is/did/will/are/were ...
  - **Where** is/are/were ...

  a. Category-specific transformation rules
     eg “For Where questions, move ‘is’ to all possible locations”
     “Where is the Louvre Museum located”
     → “is the Louvre Museum located”
     → “the is Louvre Museum located”
     → “the Louvre is Museum located”
     → “the Louvre Museum is located”
     → “the Louvre Museum located is”

  b. Expected answer “Datatype” (eg, Date, Person, Location, ...)
     **When** was the French Revolution? → DATE

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

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

- Simple: Enumerate all N-grams (N=1,2,3 say) in all retrieved snippets

- 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... → Date
  - Where... → Location
  - What ...
  - Who ...

- 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

Scores

20  Charles  Dickens
15  Dickens
10  Mr Charles

merged, discard old n-grams

Score 45  Mr Charles  Dickens

N-Grams  tile highest-scoring n-gram  N-Grams

Repeat, until no more overlap
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 redundancy 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., an individual’s email...) we only have a limited 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
Abduction: LCC
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
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`
How hot does the inside of an active volcano get?
“lava fragments belched out of the mountain were as hot as 300 degrees Fahrenheit”

- volcano ISA mountain
- lava ISPARTOF volcano ■ lava IN 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’
Watson

Slides from Ferrucci et al, AI Magazine, 2010
Jeopardy...

Category: General Science
Clue: When hit by electrons, a phosphor gives off electromagnetic energy in this form.
Answer: Light (or Photons)

Category: Lincoln Blogs
Clue: Secretary Chase just submitted this to me for the third time; guess what, pal. This time I’m accepting it.
Answer: his resignation

Category: Head North
Clue: They’re the two states you could be reentering if you’re crossing Florida’s northern border.
Answer: Georgia and Alabama

Category: Decorating
Clue: Though it sounds “harsh,” it’s just embroidery, often in a floral pattern, done with yarn on cotton cloth.
Answer: crewel

Category: “Rap” Sheet
Clue: This archaic term for a mischievous or annoying child can also mean a rogue or scamp.
Subclue 1: This archaic term for a mischievous or annoying child.
Subclue 2: This term can also mean a rogue or scamp.
Answer: Rapscallion

Category: Before and After Goes to the Movies
Clue: Film of a typical day in the life of the Beatles, which includes running from bloodthirsty zombie fans in a Romero classic.
Subclue 2: Film of a typical day in the life of the Beatles.
Answer 1: (A Hard Day’s Night)
Subclue 2: Running from bloodthirsty zombie fans in a Romero classic.
Answer 2: (Night of the Living Dead)
Answer: A Hard Day’s Night of the Living Dead
Watson on TREC

- DeepQA performs on BOTH tasks 9/2008
- CMU’s 2007 TREC QA
- IBM’s 2005 TREC QA
- DeepQA prior to Adaptation

- Jeopardy!
- TREC

- PIQUANT
- Ephyra
- DeepQA
Complex QA
Example of Complex Questions

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?