Question Answering (and Textual Entailment)

Prof. Sameer Singh

CS 295: STATISTICAL NLP
WINTER 2017

March 14, 2017

Based on slides from Dan Jurafsky, Yejin Choi, Stephen Clark, Dan Klein, Niranjan Balasubramanian, and everyone else they copied from.
Upcoming...

Homework
- Homework 4 was due on last night
- Lowest grade of the homeworks will be dropped

Summaries
- Paper summaries due tonight
- Summary 2 graded

Project
- Final report due in a week: March 20, 2017
- Instructions coming soon: ACL style, 5 pages (+references)

TA/Instructor Evaluations are available!
Outline

Question Answering
IR-Based QA System
Other Extensions
Textual Entailment
Outline

- Question Answering
- IR-Based QA System
- Other Extensions
- Textual Entailment
Questions in Modern Systems

Factoid questions
- Who wrote “The Universal Declaration of Human Rights”?
- How many calories are there in two slices of apple pie?
- What is the average age of the onset of autism?
- Where is Apple Computer based?

Complex (narrative) questions:
- In children with an acute febrile illness, what is the efficacy of acetaminophen in reducing fever?
- What do scholars think about Jefferson’s position on dealing with pirates?
## Commercial systems: mainly factoid questions

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the Louvre Museum located?</td>
<td>In Paris, France</td>
</tr>
<tr>
<td>What’s the abbreviation for limited partnership?</td>
<td>L.P.</td>
</tr>
<tr>
<td>What are the names of Odin’s ravens?</td>
<td>Huginn and Muninn</td>
</tr>
<tr>
<td>What currency is used in China?</td>
<td>The yuan</td>
</tr>
<tr>
<td>What kind of nuts are used in marzipan?</td>
<td>almonds</td>
</tr>
<tr>
<td>What instrument does Max Roach play?</td>
<td>drums</td>
</tr>
<tr>
<td>What is the telephone number for Stanford University?</td>
<td>650-723-2300</td>
</tr>
</tbody>
</table>
Paradigms for QA

IR-based approaches
  ◦ TREC; IBM Watson; Google

Knowledge-based and Hybrid approaches
  ◦ IBM Watson; Apple Siri; Wolfram Alpha; True Knowledge Evi
Many questions can already be answered by web search

Google

What are the names of Odin’s ravens?

About 214,000 results (0.38 seconds)

Huginn and Muninn - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Huginn_and_Muninn

The names of the ravens are sometimes modernly anglicized as Hugin and Munin. In the Poetic Edda, a disguised Odin expresses that he fears that they may ...

Attestations - Archaeological record - Theories - See also
IR-based Question Answering

Google

Where is the Louvre Museum located?

Search

About 904,000 results (0.30 seconds)

Everything

Best guess for Louvre Museum Location is Paris, France
Mentioned on at least 7 websites including wikipedia.org, answers.com and east-buc.k12.ia.us - Show sources - Feedback

Images

Maps

Musée du Louvre - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Musée_du_Louvre
Musée du Louvre is located in Paris. Location within Paris. Established, 1793. Location, Palais Royal, Musée du Louvre, 75001 Paris, France. Type, Art museum ...

Videos

News

Louvre Palace - List of works in the Louvre - Category:Musée du Louvre
IR-based Factoid QA

Question Processing
- Query Formulation
- Answer Type Detection

Passage Retrieval
- Document Retrieval
- Relevant Docs
- Passage Retrieval
- Relevant passages

Answer Processing

Indexing

Answer
Knowledge-based QA (Siri)

Build a semantic representation of the query

- Times, dates, locations, entities, numeric quantities

Map from this semantics to query structured data or resources

- Geospatial databases
- Ontologies (Wikipedia infoboxes, dbPedia, WordNet, Yago)
- Restaurant review sources and reservation services
- Scientific databases
Hybrid approaches (Watson)

Build a shallow semantic representation of the query

Generate answer candidates using IR methods
- Augmented with ontologies and semi-structured data

Score each candidate using richer knowledge sources
- Geospatial databases
- Temporal reasoning
- Taxonomical classification
IBM’s Watson

WILLIAM WILKINSON’S “AN ACCOUNT OF THE PRINCIPALITIES OF WALLACHIA AND MOLDOVIA” INSPIRED THIS AUTHOR’S MOST FAMOUS NOVEL

Bram Stoker

Won Jeopardy on February 16, 2011!
Motivation for Watson

Top human players are remarkably good.

Each dot – actual historical human Jeopardy! games

Winning Human Performance

Grand Champion Human Performance

2007 QA System

More Confident

Less Confident
Single Source is not Sufficient

Structured KB approach delivers high confidence if questions can be precisely mapped to existing & reliable sources. Turns out to be rarely the case and confidence and accuracy drop off quickly.

Basic Text Search approach never delivers high confidences but quickly reaches and maintains a peak accuracy at about 30%.

Must combine deep and shallow semantic analysis over structured & unstructured content to drive up precision, recall and confidence.
Watson Architecture

Learned Models help combine and weigh the Evidence

Answer & Confidence

Question & Topic Analysis

Primary Search → Candidate Answer Generation

Answer Sources

Question Decomposition → Hypothesis Generation

Hypothesis Generation

Hypothesis and Evidence Scoring

Hypothesis and Evidence Scoring

Merging & Ranking

Synthesis

Final Confidence Merging & Ranking
Watson Performance
Outline

- Question Answering
- IR-Based QA System
- Other Extensions
- Textual Entailment
IR-based Factoid QA

Question Processing
- Query Formulation
- Answer Type Detection

Document Retrieval

Passage Retrieval
- Relevant Docs

Answer Type Detection

Indexing

Answer Processing

Answer
IR-based Factoid QA

QUESTION PROCESSING
- Detect question type, answer type, focus, relations
- Formulate queries to send to a search engine

PASSAGE RETRIEVAL
- Retrieve ranked documents
- Break into suitable passages and rerank

ANSWER PROCESSING
- Extract candidate answers
- Rank candidates
  - using evidence from the text and external sources
Factoid Q/A
Question Processing

Answer Type Detection
- Decide the named entity type (person, place) of the answer

Query Formulation
- Choose query keywords for the IR system

Question Type classification
- Is this a definition question, a math question, a list question?

Focus Detection
- Find the question words that are replaced by the answer

Relation Extraction
- Find relations between entities in the question
They’re the two states you could be reentering if you’re crossing Florida’s northern border

Answer Type: US state
Query: two states, border, Florida, north
Focus: the two states
Relations: borders(Florida, ?x, north)
Answer Types: Named Entities

Who founded Virgin Airlines?
  PERSON

What Canadian city has the largest population?
  CITY
Part of Answer Type Taxonomy

- LOCATION
  - country
  - city
  - state
- DESCRIPTION
  - reason
  - definition
  - food
  - currency
  - animal
- ENTITY
  - date
  - percent
  - distance
- NUMERIC
  - money
  - size
- ABBREVIATION
  - expression
  - abbreviation
- HUMAN
  - individual
  - title
  - group

Li, Roth. Learning Question Classifiers. COLING (2002)
# Answer Types

<table>
<thead>
<tr>
<th>ENTITY</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>What are the names of Odin’s ravens?</td>
</tr>
<tr>
<td>body</td>
<td>What part of your body contains the corpus callosum?</td>
</tr>
<tr>
<td>color</td>
<td>What colors make up a rainbow?</td>
</tr>
<tr>
<td>creative</td>
<td>In what book can I find the story of Aladdin?</td>
</tr>
<tr>
<td>currency</td>
<td>What currency is used in China?</td>
</tr>
<tr>
<td>disease/medicine</td>
<td>What does Salk vaccine prevent?</td>
</tr>
<tr>
<td>event</td>
<td>What war involved the battle of Chapultepec?</td>
</tr>
<tr>
<td>food</td>
<td>What kind of nuts are used in marzipan?</td>
</tr>
<tr>
<td>instrument</td>
<td>What instrument does Max Roach play?</td>
</tr>
<tr>
<td>lang</td>
<td>What’s the official language of Algeria?</td>
</tr>
<tr>
<td>letter</td>
<td>What letter appears on the cold-water tap in Spain?</td>
</tr>
<tr>
<td>other</td>
<td>What is the name of King Arthur’s sword?</td>
</tr>
<tr>
<td>plant</td>
<td>What are some fragrant white climbing roses?</td>
</tr>
<tr>
<td>product</td>
<td>What is the fastest computer?</td>
</tr>
<tr>
<td>religion</td>
<td>What religion has the most members?</td>
</tr>
<tr>
<td>sport</td>
<td>What was the name of the ball game played by the Mayans?</td>
</tr>
<tr>
<td>substance</td>
<td>What fuel do airplanes use?</td>
</tr>
<tr>
<td>symbol</td>
<td>What is the chemical symbol for nitrogen?</td>
</tr>
<tr>
<td>technique</td>
<td>What is the best way to remove wallpaper?</td>
</tr>
<tr>
<td>term</td>
<td>How do you say “Grandma” in Irish?</td>
</tr>
<tr>
<td>vehicle</td>
<td>What was the name of Captain Bligh’s ship?</td>
</tr>
<tr>
<td>word</td>
<td>What’s the singular of dice?</td>
</tr>
</tbody>
</table>
More Answer Types

<table>
<thead>
<tr>
<th>HUMAN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>description</strong></td>
</tr>
<tr>
<td>Who was Confucius?</td>
</tr>
<tr>
<td>Who were the major companies that are part of Dow Jones?</td>
</tr>
<tr>
<td>Who was the first Russian astronaut to do a spacewalk?</td>
</tr>
<tr>
<td>What was Queen Victoria’s title regarding India?</td>
</tr>
<tr>
<td><strong>group</strong></td>
</tr>
<tr>
<td><strong>ind</strong></td>
</tr>
<tr>
<td><strong>title</strong></td>
</tr>
<tr>
<td>What’s the oldest capital city in the Americas?</td>
</tr>
<tr>
<td>What country borders the most others?</td>
</tr>
<tr>
<td>What is the highest peak in Africa?</td>
</tr>
<tr>
<td>What river runs through Liverpool?</td>
</tr>
<tr>
<td>What states do not have state income tax?</td>
</tr>
<tr>
<td><strong>LOCATION</strong></td>
</tr>
<tr>
<td><strong>city</strong></td>
</tr>
<tr>
<td><strong>country</strong></td>
</tr>
<tr>
<td><strong>mountain</strong></td>
</tr>
<tr>
<td><strong>other</strong></td>
</tr>
<tr>
<td><strong>state</strong></td>
</tr>
<tr>
<td><strong>NUMERIC</strong></td>
</tr>
<tr>
<td><strong>code</strong></td>
</tr>
<tr>
<td>About how many soldiers died in World War II?</td>
</tr>
<tr>
<td>What is the date of Boxing Day?</td>
</tr>
<tr>
<td>How long was Mao’s 1930s Long March?</td>
</tr>
<tr>
<td>How much did a McDonald’s hamburger cost in 1963?</td>
</tr>
<tr>
<td>Where does Shanghai rank among world cities in population?</td>
</tr>
<tr>
<td>What is the population of Mexico?</td>
</tr>
<tr>
<td>What was the average life expectancy during the Stone Age?</td>
</tr>
<tr>
<td>What fraction of a beaver’s life is spent swimming?</td>
</tr>
<tr>
<td>What is the speed of the Mississippi River?</td>
</tr>
<tr>
<td>How fast must a spacecraft travel to escape Earth’s gravity?</td>
</tr>
<tr>
<td>What is the size of Argentina?</td>
</tr>
<tr>
<td>How many pounds are there in a stone?</td>
</tr>
</tbody>
</table>
Answer types in Watson

2500 answer types in 20,000 Jeopardy question sample
  ◦ The most frequent 200 answer types cover < 50% of data

The 40 most frequent Jeopardy answer types

he, country, city, man, film, state, she, author, group, here, company, president, capital, star, novel, character, woman, river, island, king, song, part, series, sport, singer, actor, play, team, show, actress, animal, presidential, composer, musical, nation, book, title, leader, game
Answer Type Detection

Regular expression-based rules can get some cases:
- Who {is|was|are|were} PERSON

Other rules use the question headword:
- (the headword of the first noun phrase after the wh-word)
- Which city in China has most foreign financial companies?
- What is the state flower of California?

Question words and phrases
- Part-of-speech tags
- Parse features (headwords)
- Named Entities
- Semantically related words

+ Rules
Factoid Q/A
Keyword Selection Algorithm

1. Select all non-stop words 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 their adjectival modifiers
6. Select all other nouns
7. Select all verbs
8. Select all adverbs
9. Select the QFW word (skipped in all previous steps)
10. Select all other words

\[ \cos(q, D) \]

weighted cosine
Who coined the term “cyberspace” in his novel “Neuromancer”? 

cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7
Factoid Q/A
Passage Retrieval

Step 1
- Retrieve documents using IR
  - query terms as keywords

Step 2
- Segment the documents into shorter units
  - something like paragraphs

Step 3
- Passage ranking
  - Use answer type to help rerank passages
Features for Passage Ranking

Either in rule-based classifiers or with supervised machine learning

Number of Named Entities of the right type in passage
Number of query words in passage
Number of question N-grams also in passage
Proximity of query keywords to each other in passage
Longest sequence of question words
Rank of the document containing passage
Factoid Q/A
Answer Extraction

Run an answer-type named-entity tagger on the passages

- Each answer type requires a named-entity tagger that detects it
- If answer type is CITY, tagger has to tag CITY
  - Can be full NER, simple regular expressions, or hybrid

Return the string with the right type:

- Who is the prime minister of India (PERSON)
- Manmohan Singh, Prime Minister of India, had told left leaders that the deal would not be renegotiated.
- How tall is Mt. Everest? (LENGTH)
- The official height of Mount Everest is 29035 feet
Ranking Candidate Answers

But what if there are multiple candidate answers!

Q: Who was Queen Victoria’s second son?
Answer Type: Person

• Passage:
The Marie biscuit is named after Marie Alexandrovna, the daughter of Czar Alexander II of Russia and wife of Alfred, the second son of Queen Victoria and Prince Albert
Ranking Candidate Answers

But what if there are multiple candidate answers!

Q: Who was Queen Victoria’s second son?

Answer Type: Person

• Passage:

The Marie biscuit is named after Marie Alexandrovna, the daughter of Czar Alexander II of Russia and wife of Alfred, the second son of Queen Victoria and Prince Albert
Features for ML

**Answer type match**: Candidate contains a phrase with the correct answer type.

**Pattern match**: Regular expression pattern matches the candidate.

**Question keywords**: # of question keywords in the candidate.

**Keyword distance**: Distance in words between the candidate and query keywords.

**Novelty factor**: A word in the candidate is not in the query.

**Apposition features**: The candidate is an appositive to question terms.

**Punctuation location**: The candidate is immediately followed by a comma, period, quotation marks, semicolon, or exclamation mark.

**Sequences of question terms**: The length of the longest sequence of question terms that occurs in the candidate answer.
Scoring Candidates in Watson

Each candidate answer gets scores from >50 components

- (from unstructured text, semi-structured text, triple stores)
- logical form (parse) match between question and candidate
- passage source reliability
- geospatial location
  - California is “southwest of Montana”
- temporal relationships
- taxonomic classification
Outline

- Question Answering
- IR-Based QA System
- Other Extensions
- Textual Entailment
AskMSR

Question Processing

Where is the Louvre Museum located?

in Paris France 59%, museums 12%, hostels 10%

N-Best Answers

Answer Scoring

Question

Rewrite Query

<Search Engine>

Collect Summaries, Mine N-grams

Tile N-Grams

Filter N-Grams

Answer Extraction

Search

Dumais, Banko, Brill, Lin, Ng, SIGIR (2002)
Step 1: Rewrite Queries

**Intuition:** Questions are often syntactically quite close to sentences with 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

“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”

Nonsense, but who cares? It's only a few more queries
Feedback Loops: FALCON
Which object is the best conductor of electricity?
(A) a wax crayon  
(B) a plastic spoon  
(C) a rubber eraser  
(D) an iron nail
Which object is the best conductor of electricity?
(A) a wax crayon  (B) a plastic spoon  
(C) a rubber eraser  (D) an iron nail
Fourth graders are planning a roller-skate race. Which surface would be the best for this race? (A) gravel (B) sand (C) blacktop (D) grass
Fourth graders are planning a roller-skate race. Which surface would be the best for this race? (A) gravel  (B) sand  (C) blacktop  (D) grass

- Information retrieval methods fail
- Word co-occurrence methods struggle

Graders are commonly used in the construction and maintenance of dirt roads and gravel roads

Also strong correlations between:
- grass ↔ race
- gravel ↔ surface
A student puts two identical plants in the same type and amount of soil. She gives them the same amount of water. She puts one of these plants near a sunny window and the other in a dark room. This experiment tests how the plants respond to (A) light (B) air (C) water (D) soil
A student puts two identical plants in the same type and amount of soil. She gives them the same amount of water. She puts one of these plants near a sunny window and the other in a dark room. This experiment tests how the plants respond to 

(A) **light**  (B) air  (C) water  (D) soil

**Knowledge needed (for example):**
- near sunny window → receive light
- in a dark room → no light
- test X’s response to Y → compare X+Y with X+not(Y)
Outline

- Question Answering
- IR-Based QA System
- Other Extensions
- Textual Entailment
Natural Language & Meaning

- **Meaning**
  - Variability
  - Ambiguity

- **Language**

- **Interpretation**

- **Expression**
Inference vs Entailment

- symbolic – logical forms
- statistical – embeddings

Meaning Representation

Natural Language

Inference

Textual Entailment
Textual Entailment

• A directional relation between two text fragments: *Text (t)* and *Hypothesis (h)*:

\[ t \text{ entails } h \ (t \Rightarrow h) \text{ if } \]

**humans reading t will infer that h is most likely true**

• Assuming “common background knowledge” – which is indeed expected from applications
Example

Text: The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

Hyp 1: BMI acquired an American company. ✓
Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn. ✗
Hyp 3: BMI is an employee-owned concern. ✗
More Sentence Pairs

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Some students came to school by car.</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>Some students came to school.</td>
<td>✔</td>
</tr>
<tr>
<td>2.</td>
<td>No students came to school by car.</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>Some students came to school.</td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>John drove legally.</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>John drove.</td>
<td>✔</td>
</tr>
<tr>
<td>4.</td>
<td>John drove predictably.</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>John drove.</td>
<td>✔</td>
</tr>
<tr>
<td>5.</td>
<td>Legally, John could drive.</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>John drove.</td>
<td>✗</td>
</tr>
</tbody>
</table>
Entailment with Knowledge

$t$ entails $h$ ($t \Rightarrow h$) if

**humans** reading $t$ will infer that $h$ is *most likely* true

For textual entailment to hold we require:

- $text$ AND $knowledge \Rightarrow h$, but
- *knowledge* should not entail $h$ alone

Systems are **not** supposed to validate $h$’s truth regardless of $t$
(e.g. by searching $h$ on the web)
While no one accuses Madonna of doing anything illegal in adopting the 4-year-old girl, reportedly named Mercy, there are questions nonetheless about how Madonna is able to navigate Malawi's 18-to-24 month vetting period in just a matter of days or weeks...

HYPOTHESIS:
Madonna is 50 years old.
Third Label: Contradictions

\[ t \text{ contradicts } h \quad (t \perp h) \text{ if } \]

**humans** reading \( t \) will find the relations/events described by \( h \) to be **highly unlikely** given \( t \).
Example

Text: The purchase of Houston-based LexCorp by BMI for $2Bn prompted widespread sell-offs by traders as they sought to minimize exposure. LexCorp had been an employee-owned concern since 2008.

Hyp 1: BMI acquired an American company. ✓
Hyp 2: BMI bought employee-owned LexCorp for $3.4Bn. ❌
Hyp 3: BMI is an employee-owned concern. ✗
More Examples

<table>
<thead>
<tr>
<th>Text</th>
<th>Hypothesis</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>John Smith bought a Honda Civic.</td>
<td>John owns a car.</td>
<td>entails</td>
</tr>
<tr>
<td>John Smith bought a Honda Civic.</td>
<td>John sold a car.</td>
<td>does not entail</td>
</tr>
<tr>
<td>John’s June 5th purchase of a Honda Civic gave him an alibi.</td>
<td>John did not buy a car.</td>
<td>contradicts</td>
</tr>
<tr>
<td>John Smith bought a Honda Civic for $14,000.</td>
<td>John bought a car for $12,500.</td>
<td>contradicts</td>
</tr>
<tr>
<td>John Smith bought a Honda Civic in Paris.</td>
<td>John bought a car in France.</td>
<td>entails</td>
</tr>
<tr>
<td>John Smith bought a Honda Civic in Paris.</td>
<td>Paris is in France.</td>
<td>does not entail</td>
</tr>
</tbody>
</table>
Applications

Question Answering

**Question**

Who bought Overture?  

**Expected answer form**

X bought Overture

Information Extraction

Overture’s acquisition by Yahoo entails Yahoo bought Overture

Similarity based on whether document entails the query.

Machine Translation

He abhorred the men’s unctuous ways.

He disliked the men’s flattering ways.

Information Retrieval