CS 6120/CS 4120: Natural Language Processing
Instructor: Prof. Lu Wang
College of Computer and Information Science
Northeastern University
Webpage: www.ccs.neu.edu/home/luwang

Question Answering

Questions in Search

Google: Where is the Louvre Museum located?

Google: Where is the Louvre Museum located?

Question Answering (Some Background)

One of the oldest NLP tasks (punched card systems in 1961)

Question: What do worms eat?  
Potential Answers:  
Worms eat grass  
Grass is eaten by worms  
Birds eat worms

Question: What do worms eat?  
Potential Answers:  
Worms eat grass  
Grass is eaten by worms  
Birds eat worms

Question Answering
What do worms eat?

Worms eat grass.

Grass is eaten by worms.

Birds eat worms.

Horses eat grass.

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

IBM’s Watson

WON JEOPARDY ON FEBRUARY 16, 2011!

Types of 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>
</tbody>
</table>
Paradigms for Factoid QA

- Information Retrieval (IR)-based approaches
  - IBM Watson (some parts); Google
- Knowledge-based and Hybrid approaches
  - IBM Watson; Apple Siri; Wolfram Alpha
- Built upon the above two:
  - Data-driven, neural network-based approaches (more recent)

Information Retrieval (IR)-based QA

- Question Processing
  - Answer Type Detection
  - Query Formulation
  - Question Type classification
  - Focus Detection
  - Relation Extraction (if there are more than one entities)
- Passage Retrieval
- Answer Processing

IR-based Factoid QA

- Question Processing: Things to extract from the question
  - Answer Type Detection
  - Query Formulation
  - Question Type classification
  - Focus Detection
  - Relation Extraction (if there are more than one entities)
- Passage Retrieval
- Answer Processing

Question Processing

- Jeopardy: They’re the two states you could be reentering if you’re crossing Florida’s northern border
  - You should answer: what are the states of Georgia and Alabama?
  - Answer Type: US state
  - Query Formulation: two states, border, Florida, north
  - Focus: the two states
  - Relations: borders(Florida, ?x, north)
IR-based Factoid QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - "Who is the president of the US?" - person
  - Formulate queries to send to a search engine
    - "president of United States"
- 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

Answer Type Detection: Named Entities

- **Who founded Virgin Airlines?**
  - PERSON

- **What Canadian city has the largest population?**
  - CITY

Answer Type Taxonomy

- 6 coarse classes
  - ABBREVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION, NUMERIC
- 50 finer classes
  - LOCATION: city, country, mountain...
  - HUMAN: group, individual, title, description...
  - ENTITY: animal, body, color, currency...

Part of Li & Roth’s Answer Type Taxonomy
Answer Types

<table>
<thead>
<tr>
<th>Entity</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>actor</td>
<td>Who was the star of Captain Kirk’s ship?</td>
</tr>
<tr>
<td>date</td>
<td>When is the award ceremony held?</td>
</tr>
<tr>
<td>city</td>
<td>Which city in China has the largest number of foreign financial companies?</td>
</tr>
<tr>
<td>country</td>
<td>What is the capital of France?</td>
</tr>
<tr>
<td>group</td>
<td>What was the movie “The Godfather” about?</td>
</tr>
<tr>
<td>island</td>
<td>What are the largest islands of the world?</td>
</tr>
<tr>
<td>name</td>
<td>What is the name of Captain Kirk’s ship?</td>
</tr>
<tr>
<td>ocean</td>
<td>What is the largest ocean on Earth?</td>
</tr>
<tr>
<td>person</td>
<td>What is the name of Captain Kirk’s ship?</td>
</tr>
<tr>
<td>plant</td>
<td>What is the state flower of California?</td>
</tr>
<tr>
<td>region</td>
<td>What is the capital of France?</td>
</tr>
<tr>
<td>river</td>
<td>What is the longest river in the world?</td>
</tr>
<tr>
<td>state</td>
<td>What is the capital of France?</td>
</tr>
<tr>
<td>song</td>
<td>What is the most popular song of all time?</td>
</tr>
<tr>
<td>sport</td>
<td>What is the most popular sport in the world?</td>
</tr>
<tr>
<td>team</td>
<td>What is the name of Captain Kirk’s ship?</td>
</tr>
<tr>
<td>show</td>
<td>What is the name of Captain Kirk’s ship?</td>
</tr>
</tbody>
</table>

More Answer Types

<table>
<thead>
<tr>
<th>Type</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>Who is the CEO of Google?</td>
</tr>
<tr>
<td>Location</td>
<td>What is the capital of France?</td>
</tr>
<tr>
<td>Date</td>
<td>What is the date of the award ceremony?</td>
</tr>
<tr>
<td>Number</td>
<td>How many CDs did Metallica sell in 1997?</td>
</tr>
<tr>
<td>Currency</td>
<td>What is the currency of Japan?</td>
</tr>
<tr>
<td>Name</td>
<td>What is the name of Captain Kirk’s ship?</td>
</tr>
<tr>
<td>Title</td>
<td>What is the title of the movie “The Godfather”?</td>
</tr>
</tbody>
</table>

Answer types in Jeopardy

• 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
  country, city, man, film, state, 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

• Hand-written rules
• Machine Learning

Answer Type Detection

• Regular expression-based rules can get some cases:
  • [Who is|was|are|were] PERSON
  • PERSON/YEAR—YEAR
• Other rules use the question headword:
  (the headword of the first noun phrase after the wh-word)
  • Which city in China has the largest number of foreign financial companies?
  • What is the state flower of California?
Features for Answer Type Detection

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

Which city in China has the largest number of foreign financial companies?
What is the state flower of California?

Query Formulation

• QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine
  - “predict of united states”

• 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

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 question focus word (skipped in all previous steps)
10. Select all other words

Choosing keywords from the query

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

cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

IR-based Factoid QA

• QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - “predict of united states”
  - Formulate queries to send to a search engine
  - “predict of united states”

• 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

Passage Retrieval

• Step 1: IR engine retrieves documents using query terms
• Step 2: Segment the documents into shorter units
  - E.g. paragraphs or consecutive sentences
• 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

Passage Retrieval as Query-focused Summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists around the world.] "This is one of the most compelling stories about the Amazon rain forest in recent years," said Thomas Lovejoy, an Amazon specialist. "In contrast, practically a year ago, when there was nothing to read about conservation in the Amazon."

[Print page of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]

Passage Retrieval as Query-focused Summarization

Which country has the largest part of the Amazon rain forest?

[The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by leading scientists around the world.] "This is one of the most compelling stories about the Amazon rain forest in recent years," said Thomas Lovejoy, an Amazon specialist. "In contrast, practically a year ago, when there was nothing to read about conservation in the Amazon."

[Print page of the Amazon, the world’s largest tropical rain forest, lies in Brazil.]

IR-based Factoid QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - "Who is the president of the USA?" - person
  - Formulate queries to send to a search engine
  - "president of the USA" - "president of United States"

- 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
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 29,035 feet

Adding Analysis Patterns

• "Who is Elvis?"
  • Question type: "who"
  • Named-entity tagging: "Who is <person-name> Elvis</person-name>"
  • Analysis pattern: if question type = "who" and question contains <person-name> then
  • Desired answer probably is a description
  • Likely answer extraction patterns
    • "Elvis, the X", e.g., "Elvis, the king of rock and roll!"
    • "the X Elvis", e.g., "the legendary entertainer Elvis"

IR-based Factoid QA

• QUESTION PROCESSING
  • Detect question type, answer type, focus, relations
    • "Who is the president of US?" - person
  • Formulate query to send to a search engine
    • "president of united states"
• 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

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

Use machine learning:
Features for ranking candidate answers

  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.
Candidate Answer scoring in IBM 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

Information Retrieval (IR)-based QA

• Factoid QA pipeline
• Factoid QA evaluation
• Common Knowledge used in QA
• Recent QA tasks

Common Evaluation Metrics

1. **Accuracy** (does answer match gold-labeled answer?)
2. **Mean Reciprocal Rank**
   • For each query return a ranked list of M candidate answers.
   • Query score is $1/rank$ of the first correct answer
   • If first answer is correct: 1
   • Else if second answer is correct: ½
   • Else if third answer is correct: ⅓, etc.
   • Score is 0 if none of the M answers are correct
   • Take the mean over all N queries

Information Retrieval (IR)-based QA

• Factoid QA pipeline
• Factoid QA evaluation
• Common Knowledge used in QA
• Recent QA tasks

Knowledge in QA

• What are other types of knowledge useful for a QA system?
  • Relations
  • Temporal information
  • Dialogue context

Relation Extraction

• **Answers**: Databases of Relations
  • born-in(“Emma Goldman”, “June 27 1869”)
  • author-of(“Cao Xue Qin”, “Dream of the Red Chamber”)
  • Draw from Wikipedia infoboxes, DBpedia, FreeBase, etc.
• **Questions**: Extracting Relations in Questions
  Whose granddaughter starred in E.T.?
  (acted-in ?x “E.T.”)
  (granddaughter-of ?x ?y)
Temporal Reasoning

- **Relation databases**
  
  - (and dictionaries, biographical dictionaries, etc.)

- IBM Watson

  "In 1594 he took a job as a tax collector in Andalusia"

Candidates:

- Thoreau is a bad answer (born in 1817)
- Cervantes is possible (was alive in 1594)

Context and Conversation in Virtual Assistants like Siri

- Coreference helps resolve ambiguities

  U: "Book a table at Il Fornaio at 7:00 with my mom"
  U: "Also send her an email reminder"

- Clarification questions:

  U: "Chicago pizza"
  S: "Did you mean pizza restaurants in Chicago or Chicago-style pizza?"

Limitations of Factoid Q/A

- Question must query a specific fact that is explicitly stated somewhere in the document corpus.
- Does not allow aggregating or accumulating information across multiple information sources.
- Does not require "deep compositional" semantics, nor inferential reasoning to generate answer.

Information Retrieval (IR)-based QA

- Factoid QA pipeline
- Factoid QA evaluation
- Common Knowledge used in QA

Recent QA tasks

What are recent tasks for QA?

- Reading comprehension (machine reading)
- Visual Question Answering

Reading Comprehension Q/A

- Answer questions that test comprehension of a specific document.
- Use standardized tests of reading comprehension to evaluate performance (Hirschman et al. 1999; Rito & Thelen, 2000; Ng et al. 2000; Charniak et al. 2000).
Sample Reading Comprehension Test

Large Scale Reading Comprehension Data
• DeepMind's large-scale data for reading comprehension Q/A (Hermann et al., 2015).
• News articles used as source documents.
• Questions constructed automatically from article summary sentences.

Sample DeepMind Reading Comprehension Test

Deep LSTM Reader
• DeepMind uses LSTM recurrent neural net (RNN) to encode document and query into a vector that is then used to predict the answer.

Visual Question Answering (VQA)
• Answer natural language questions about information in images.
• ViTech/MSR group has put together VQA dataset with ~750K questions over ~210K images (Antol et al., 2016).
LSTM System for VQA