Asking Questions

Questions in Search

Personal Assistants

Question Answering (Some Background)

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

<table>
<thead>
<tr>
<th>Question</th>
<th>Potential Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do worms eat?</td>
<td>Worms eat grass, birds eat worms, grass is eaten by worms</td>
</tr>
<tr>
<td>Grass with worms eat grass</td>
<td></td>
</tr>
</tbody>
</table>

What do worms eat?

Worms eat grass

Birds eat worms

Grass is eaten by worms

Worms with worms eat grass

Worms with worms eat grass

Grass is eaten by worms

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

Question: What do worms eat?
Potential Answers: Worms eat grass, Horses with worms eat grass

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

Question: What do worms eat?
Potential Answers: Worms eat grass, Horses with worms eat grass

Question Answering

One milestone for 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

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?

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 (some other parts); Apple Siri; Wolfram Alpha
- Built upon the above two:
  - Data-driven, neural network-based approaches that rely on retrieval AND knowledge bases
**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? Georgia and Alabama?

• **Answer Type:** US state
• **Query Formulation:** two states, border, Florida, north
• **Focus:** the two states
• **Relations:** borders(Florida, ?x, north)

**Answer Type Detection: Named Entities**

• **Who founded Virgin Airlines?**

• **What Canadian city has the largest population?**

**Answer Type Taxonomy**

Xin U, Dan Roth. 2002. Learning Question Classifiers. COLING’02

• 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

More Answer Types

Answer Types in Jeopardy

IR-based Factoid QA

Answer Type Detection
Answer Type Detection

• Regular expression-based rules can get some cases:
  • Who is\{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
  • “Who is the president of 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

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”?
IR-based Factoid QA
- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
    - "Who is the president of 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

Features for Passage Ranking
Using 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
- Step 1: IR engine retrieves documents using query terms
- Step 2: Segment the documents into shorter units
  - E.g. paragraphs or consecutive sentences
  - Think about why? Why not single sentences?
- Step 3: Passage ranking
  - E.g. use answer type to help rerank passages

IR-based Factoid QA
- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
    - "Who is the president of 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 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
IR-based Factoid QA

- **QUESTION PROCESSING**
  - Detect question type, answer type, focus, relations
  - "Who is the president of US?" - person
- **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.
- **Feature 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 like knowledge bases)
  - 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 $1/\text{Rank of the first correct answer}$
     - If first answer is correct: 1
     - Else if second answer is correct: $\frac{1}{2}$
     - Else if third answer is correct: $\frac{1}{3}$, etc.
     - Score is 0 if none of the M answers are correct
   - Take the mean over all N queries

\[ MRR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\text{rank}_i} \]

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 obituaries, 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. → Natural language generation!

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 task
- Popular QA benchmarks
- Visual Question Answering (language + vision)

Reading Comprehension

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

Sample Reading Comprehension Test

1. What is the name of the wind? (B) "Who was the wind named after?"
   A. The wind was named after the wind god, Boreas.

2. What did the wind do? (B) "What did the wind do?"
   A. The wind carried the leaves across the field.

3. What is the title of the story? (C) "What is the title of the story?"
   A. "Leaves Blowing in the Wind"

4. What is important about the wind? (B) "What is important about the wind?"
   A. The wind can be both powerful and gentle, blowing leaves and carrying them away.

5. What is the author's purpose in writing this story? (B) "What is the author's purpose in writing this story?"
   A. To describe the beauty and power of wind.
Large Scale Reading Comprehension Data

- DeepMind’s large-scale data for reading comprehension QA (Hermann et al., 2015).
- News articles used as source documents.
- Questions constructed automatically from article summary sentences.

<table>
<thead>
<tr>
<th>Question</th>
<th>Document</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which team won Super Bowl 50?</td>
<td>Denver Broncos, Carolina Panthers</td>
<td>Denver Broncos defeated Carolina Panthers 24–10</td>
</tr>
<tr>
<td>How old would she be?</td>
<td>80</td>
<td>she was turning 80</td>
</tr>
<tr>
<td>Did she plan to have any visitors?</td>
<td>Yes</td>
<td>Her granddaughter Annie was coming over</td>
</tr>
</tbody>
</table>

Stanford Question Answering Dataset (SQuAD)

- Question: Which team won Super Bowl 50?
- Passage (context): Super Bowl 50 was an American football game to determine the champion of the National Football League (NFL) for the 2015 season. The American Football Conference (AFC) champion Denver Broncos defeated the National Football Conference (NFC) champion Carolina Panthers 24–10 to earn their third Super Bowl title.
- 100k examples
- Answer must be a span in the passage

CoQA: A Conversational Question Answering Challenge

- Input: Jessica went to sit in her rocking chair. Today was her birthday and she was turning 80. Her granddaughter Annie was coming over in the afternoon and Jessica was very excited to see her. Her daughter Melanie and Melanie’s husband Josh were coming as well. Jessica had...

  - Q1: Who had a birthday?
    - A1: Jessica
  - Q2: How old would she be?
    - A2: 80
  - Q3: Did she plan to have any visitors?
    - A3: Yes
    - R3 (rational): Her granddaughter Annie was coming over.
The NarrativeQA Reading Comprehension Challenge

- Title: Ghostbusters II
- Question: How is Oscar related to Dana?
- Answer: her son
- Summary snippet: . . . Peter’s former girlfriend Dana Barrett has had a son, Oscar . . .
- Story snippet:
  - DANA (setting the wheel brakes on the buggy) Thank you, Frank. I’ll get the hang of
    this eventually.
  - She continues digging in her purse while Frank leans over the buggy and makes
    funny faces at the baby, OSCAR, a very cute nine-month-old boy.
  - FRANK (to the baby) Hiya, Oscar. What do you say, slugger?
  - FRANK (to Dana) That’s a good-looking kid you got there, Ms. Barrett.

Visual Question Answering (VQA)

- Answer natural language questions about information in images.
- VaTech/MSR group has put together VQA dataset with ~750K questions over ~250K images (Antol et al., 2016).

VQA Examples

LSTM System for VQA