What is Question Answering?
Question Answering

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

Birds eat worms
Grass is eaten by worms

Simmons, Klein, McConlogue. 1964. Indexing and Dependency Logic for Answering English Questions. American Documentation 15:30, 196-204
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
Apple’s Siri

“Do I need an umbrella tomorrow in San Francisco”

Yes, San Francisco should get rain tomorrow:

Weekly Forecast

<table>
<thead>
<tr>
<th>TUES</th>
<th>WED</th>
<th>THU</th>
<th>FRI</th>
<th>SAT</th>
<th>SUN</th>
</tr>
</thead>
<tbody>
<tr>
<td>14°</td>
<td>16°</td>
<td>17°</td>
<td>17°</td>
<td>18°</td>
<td>18°</td>
</tr>
<tr>
<td>9°</td>
<td>7°</td>
<td>6°</td>
<td>6°</td>
<td>7°</td>
<td>8°</td>
</tr>
</tbody>
</table>
Dan Jurafsky

Wolfram Alpha

how many calories are in two slices of banana cream pie?

Assuming any type of pie, banana cream | Use pie, banana cream, prepared from recipe or pie, banana cream, no-bake type, prepared from mix instead

Input interpretation:

| pie | amount | 2 slices | type | banana cream | total calories |

Average result:

702 Cal (dietary Calories)
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>
<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

What are the names of Odin’s ravens?

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

Where is the Louvre Museum located?

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

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 ...
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
- Relevant Docs
- Passage Retrieval

Answer Processing

Indexing

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
Knowledge-based approaches (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 (IBM 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
What is Question Answering?
Question Answering

Answer Types and Query Formulation
Factoid Q/A

1. Question Processing
   - Query Formulation
   - Answer Type Detection

2. Passage Retrieval
   - Document Retrieval
   - Relevant Docs
   - Passage Retrieval
   - passages

3. Answer Processing
   - Answer
Question Processing
Things to extract from the question

• 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 Type Detection: Named Entities

• **Who founded Virgin Airlines?**
  • PERSON

• **What Canadian city has the largest population?**
  • CITY.
Answer Type Taxonomy

Xin Li, Dan Roth. 2002. Learning Question Classifiers. COLING'02

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

- Hand-written rules
- Machine Learning
- Hybrids
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?
Answer Type Detection

- Most often, we treat the problem as machine learning classification
  - **Define** a taxonomy of question types
  - **Annotate** training data for each question type
  - **Train** classifiers for each question class using a rich set of features.
    - features include those hand-written rules!
Features for Answer Type Detection

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

Question Processing
- Query Formulation
- Answer Type Detection

Answer Type Detection -> Question

Indexing

Document Retrieval

Passage Retrieval

Relevant Docs

Passage Retrieval

Answer

Document Retrieval

Answer Processing

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

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

Answer Types and Query Formulation
Passage Retrieval and Answer Extraction
Factoid Q/A

Question Processing
- Query Formulation
- Answer Type Detection

Passage Retrieval
- Document Retrieval
- Relevant Docs
- Passage Retrieval

Answer Processing

Indexing

Answer
Passage Retrieval

• Step 1: IR engine retrieves documents using query terms
• 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
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
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

$$MRR = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{\text{rank}_i}$$
Question Answering

Passage Retrieval and Answer Extraction
Question Answering

Using Knowledge in QA
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)
Geospatial knowledge (containment, directionality, borders)

- Beijing is a good answer for "Asian city"
- California is "southwest of Montana"
- geonames.org:
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?”
Question Answering

Using Knowledge in QA
Question Answering

Question Answering in Watson (Deep QA)
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
The Architecture of Watson

(1) Question Processing
- Focus Detection
- Lexical Answer Type Detection
- Question Classification
- Parsing
- Named Entity Tagging
- Relation Extraction
- Coreference

(2) Candidate Answer Generation
- From Text Resources
  - Document and Passage Retrieval
  - Answer Extraction
    - Document titles
    - Anchor text
- From Structured Data
  - Relation Retrieval
    - DBpedia
    - Freebase

(3) Candidate Answer Scoring
- Evidence Retrieval and scoring
  - Text Evidence Sources
  - Time from DBpedia
  - Answer Type
  - Space from Facebook

(4) Confidence Merging and Ranking
- Merge Equivalent Answers
- Logistic Regression
- Answer Type
- Confidence

Answer and Confidence
Stage 1: Question Processing

- Parsing
- Named Entity Tagging
- Relation Extraction
- Focus
- Answer Type
- Question Classification
Poets and Poetry: He was a bank clerk in the Yukon before he published “Songs of a Sourdough” in 1907.


Author of (“Songs of a sourdough”) publish (e1, he, “Songs of a sourdough”) in (e2, e1, 1907) temporallink(publish(...), 1907)
Focus extraction

- **Focus**: the part of the question that co-refers with the answer
- Replace it with answer to find a supporting passage.
- Extracted by hand-written rules
  - "Extract any noun phrase with determiner this"
  - “Extracting pronouns she, he, hers, him, “
Lexical Answer Type

- The semantic class of the answer
- But for Jeopardy the TREC answer type taxonomy is insufficient
- DeepQA team investigated 20,000 questions
- 100 named entities only covered <50% of the questions!
- Instead: Extract lots of words: 5,000 for those 20,000 questions
Lexical Answer Type

- Answer types extracted by hand-written rules
  - Syntactic headword of the focus.
  - Words that are coreferent with the focus
  - Jeopardy! category, if refers to compatible entity.

Poets and Poetry: He was a bank clerk in the Yukon before he published “Songs of a Sourdough” in 1907.
Relation Extraction in DeepQA

• For the most frequent 30 relations:
  • Hand-written regular expressions
  • AuthorOf:
    • Many patterns such as one to deal with:
      • a Mary Shelley tale, the Saroyan novel, Twain’s travel books, a 1984 Tom Clancy thriller
    • [Author] [Prose]

• For the rest: distant supervision
Stage 2: Candidate Answer Generation
Extracting candidate answers from triple stores

• If we extracted a relation from the question
  
  ... he published “Songs of a sourdough”
  
  (author-of ?x “Songs of a sourdough”)

• We just query a triple store
  
  • Wikipedia infoboxes, DBpedia, FreeBase, etc.
  • born-in(“Emma Goldman”, “June 27 1869”)
  • author-of(“Cao Xue Qin”, “Dream of the Red Chamber”)
  • author-of(“Songs of a sourdough”, “Robert Service”)
Extracting candidate answers from text: get documents/passages

1. Do standard IR-based QA to get documents

Robert Redford and Paul Newman starred in this depression-era grifter flick.

(2.0 Robert Redford) (2.0 Paul Newman) star depression era grifter (1.5 flick)
Extracting answers from documents / passages

• Useful fact: Jeopardy! answers are mostly the title of a Wikipedia document
  • If the document is a Wikipedia article, just take the title
  • If not, extract all noun phrases in the passage that are Wikipedia document titles
  • Or extract all anchor text <a>The Sting</a>
Stage 3: Candidate Answer Scoring

- Use lots of sources of evidence to score an answer
  - more than 50 scorers

- **Lexical answer type** is a big one
  - Different in DeepQA than in pure IR factoid QA
  - In pure IR factoid QA, answer type is used to strictly filter answers
  - In DeepQA, answer type is just one of many pieces of evidence
Lexical Answer Type (LAT) for Scoring Candidates

• Given:
  • candidate answer & lexical answer type

• Return a score: can answer can be a subclass of this answer type?

• Candidate: “difficulty swallowing” & LAT “condition”

1. Check DBPedia, WordNet, etc
   • difficulty swallowing -> Dbpedia Dysphagia -> WordNet Dysphagia
   • condition -> WordNet Condition

2. Check if “Dysphagia” IS-A “Condition” in WordNet
   • Wordnet for dysphagia
Relations for scoring

- **Q**: This hockey defenseman ended his career on June 5, 2008
- **Passage**: On June 5, 2008, Wesley announced his retirement after his 20th NHL season

- Question and passage have very few keywords in common
- But both have the Dbpedia relation `ActiveYearsEndDate`
Temporal Reasoning for Scoring Candidates

• 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)
Geospatial knowledge (containment, directionality, borders)

- Beijing is a good answer for “Asian city”
- California is “southwest of Montana”
- geonames.org:
Text-retrieval-based answer scorer

- Generate a query from the question and retrieve passages
- Replace the focus in the question with the candidate answer
- See how well it fits the passages.
- Robert Redford and Paul Newman starred in this depression-era grifter flick
- Robert Redford and Paul Newman starred in The Sting

Robert Redford - Wikipedia, the free encyclopedia

Redford starred in Sydney Pollack's Out of Africa (1985), which was an .... by William Goldman, in which he was paired for the first time with Paul Newman. ... the blockbuster crime caper The Sting (1973), which became one of the top 20 ...
Stage 4: Answer Merging and Scoring

• Now we have a list candidate answers each with a score vector
  • J.F.K  [.5 .4 1.2 33 .35 ... ]
  • John F. Kennedy [.2 .56 5.3 2 ...]

• Merge equivalent answers: J.F.K. and John F. Kennedy
  • Use Wikipedia dictionaries that list synonyms:
    • JFK, John F. Kennedy, John Fitzgerald Kennedy, Senator John F. Kennedy, President Kennedy, Jack Kennedy
  • Use stemming and other morphology
Stage 4: Answer Scoring

- Build a classifier to take answers and a score vector and assign a probability
- Train on datasets of hand-labeled correct and incorrect answers.
Question Answering

Question Answering in Watson (Deep QA)