Question Answering

CS4742 Natural Language Processing
Lecture 09

Jiho Noh

Department of Computer Science
Kennesaw State University

CS4742 Summer 2025 [1]

[1] This lecture is based on the slides from Dr. Hafiz Khan at KSU, which are from Prof. Lu Wang's NLP course at Northeastern Univ.

1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. Recent QA Tasks

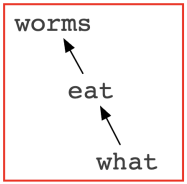
Question Answering (QA)

- **Question-Answering** is one of the *old* NLP tasks (punched card systems in 1961).
- Early 1960s, QA systems used two major paradigms:
  - *Information retrieval-based*
  - *Knowledge-based*
- QA systems are mainly designed to fill human information needs.

Question Answering

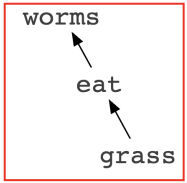
Question:

What do worms eat?

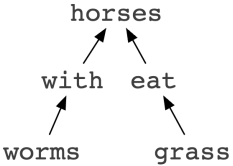


Potential Answers:

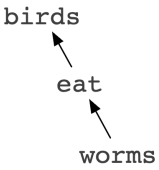
Worms eat grass



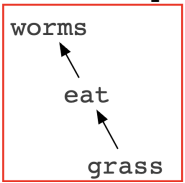
Horses with worms eat grass



Birds eat worms



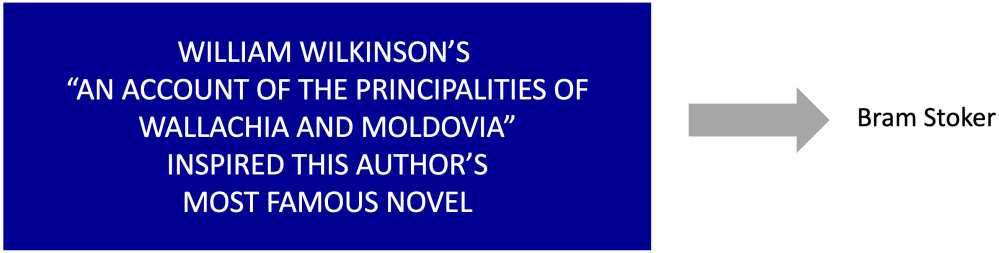
Grass is eaten by worms



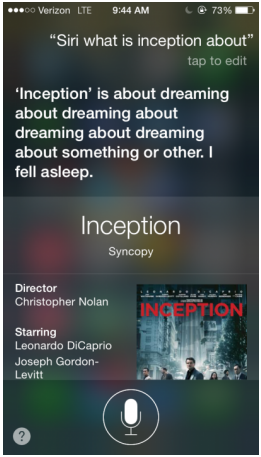
Caption

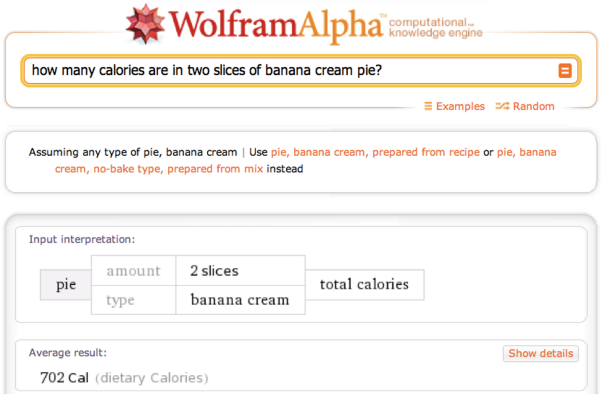
Question Answering: IBM's Watson

- **Won Jeopardy** on *February 16, 2011*!



Daily Question-Answering





Types of Questions in Modern Systems

- **Factoid questions** - questions that can be answered with simple facts expressed in short texts
  - *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

- *Where is the **Louvre Museum** located?* In Paris, France
- *What's the abbreviation for **limited partnership**?* L.P.
- *What are the names of **Odin's ravens**?* Huginn and Muninn
- *What currency is used in **China**?* The yuan
- *What kind of nuts are used in **marzipan**?* almonds
- *What instrument does **Max Roach** play?* drums
- *What is the telephone number for **Stanford University**?* 650-723-2300

Paradigms of QA Systems

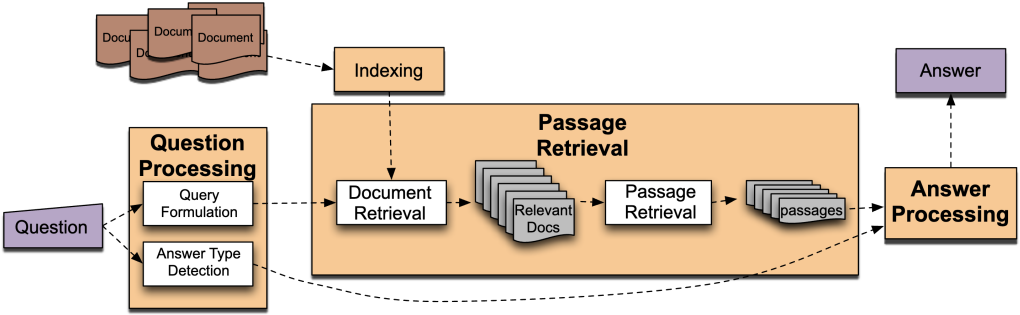
- **IR-based approaches** or open domain question QA
  - Relies on the vast amount of text on the web or in collections of scientific papers like *PubMed*.
  - Given a user question, *information retrieval* is used to find relevant passages.
  - Neural reading comprehension algorithms read these retrieved passages and draw an answer directly from spans of text.
- Examples: **TREC**, **IBM Watson**, **Google**

Paradigms of QA Systems

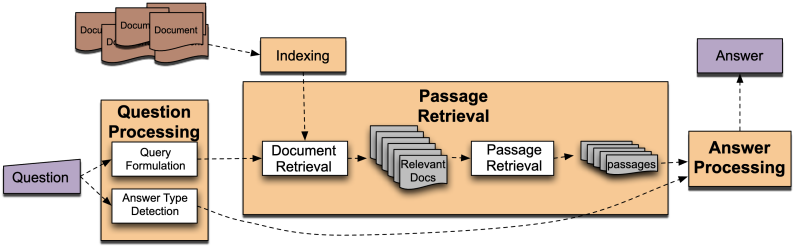
- **Knowledge-based** and **Hybrid** approaches
  - A system builds a *semantic representation* of the query
  - Example: Mapping What states border Texas? to the logical representation: $\lambda x : \text{state}(x) \wedge \text{borders}(x, \text{texas})$
  - Example: Mapping When was Ada Lovelace born? to the gapped relation: *birth-year* (Ada Lovelace,?x)
  - These *meaning representations* are used to query databases of facts
- Examples: **IBM Watson**, **Apple Siri**, **Wolfram Alpha**, **True Knowledge Evi**

1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. Recent QA Tasks

IF-based Factoid QA Pipeline

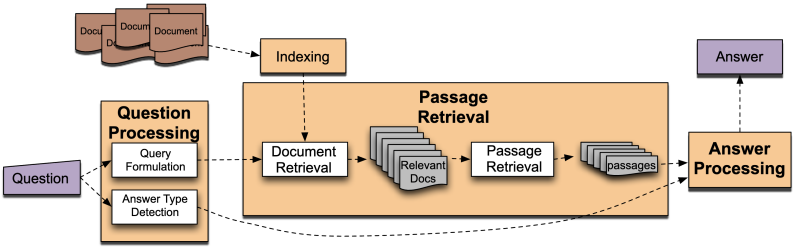


IF-based Factoid QA Pipeline — Question Processing



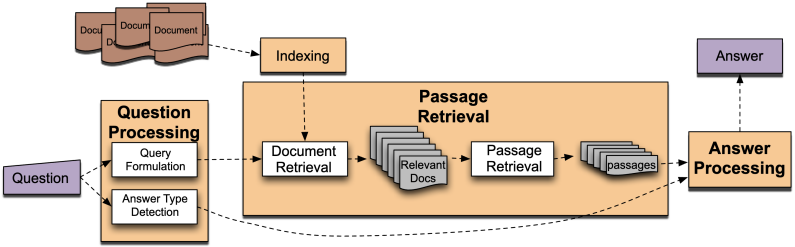
- Detect *question type*, *answer type*, *focus*, *relations*
  - "Who is the president of US?" $\rightarrow$ *person*
- Formulate queries to send to a search engine
  - "president of United States"

IF-based Factoid QA Pipeline — Passage Retrieval



- Retrieve *ranked documents*
- Break into suitable passages and **rerank**

IF-based Factoid QA Pipeline — Answer Processing



- *Extract* candidate answers
- Rank candidates
  - using **evidence** from the text and external sources

1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. Recent QA Tasks

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** (if there are more than one entities)
  - Find relations between entities in the question

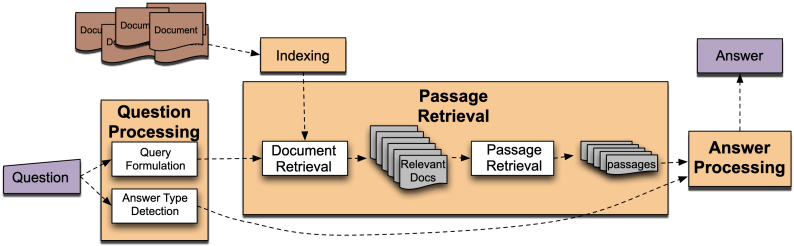
Question Processing — Example

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)

IF-based Factoid QA Pipeline — Question Processing



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

Answer Type Detection: Named Entities

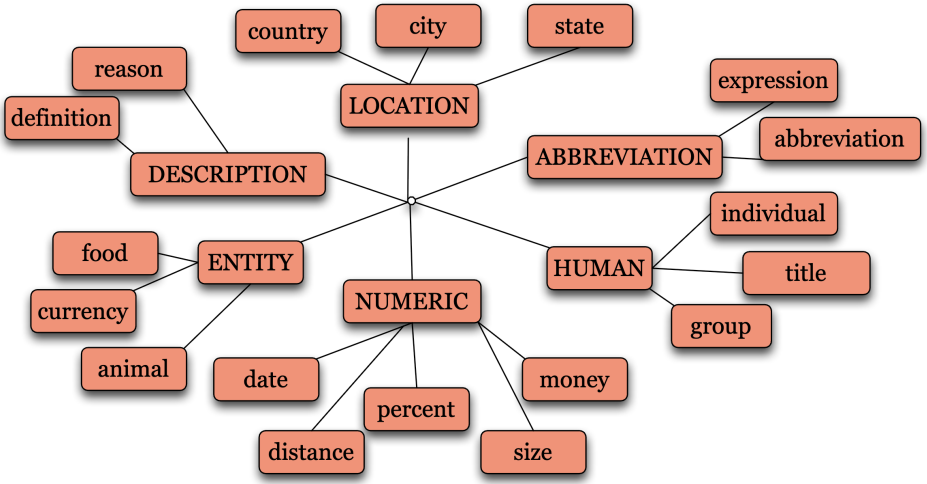
- **Who founded Virgin Airlines?**
  - *Answer Type:* PERSON
- **What Canadian city has the largest population?**
  - *Answer Type:* CITY

Answer Type Taxonomy [3]

- **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...

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

Answer Type Taxonomy (sampled) (Cont.)



Answer types in Jeopardy[4]

- **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

[4] Ferrucci et al. 2010. Building Watson: An Overview of the DeepQA Project. AI Magazine. Fall 2010. 59-79.

Answer Type Detection

- **Hand-written rules**
- **Machine Learning**

discussion

- What kind of linguistic features would you use?
- What kind of NLP preprocessing would you do?
- What kind of ML algorithms would you use?

Answer Type Detection

- **Regular expression-based rules**:
  - *Who* {is | was | are | were} PERSON
  - PERSON (YEAR – YEAR)
- Other rules use the **question headword**:
  - *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**!

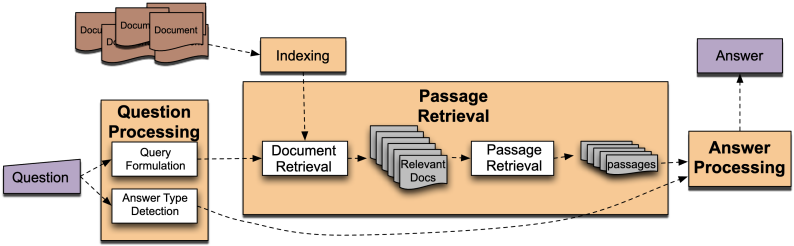
Features for Answer Type Detection

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

Q: *Which city in China has the largest number of foreign financial companies?*

Q: *What is the state flower of California?*

IF-based Factoid QA Pipeline — Question Processing



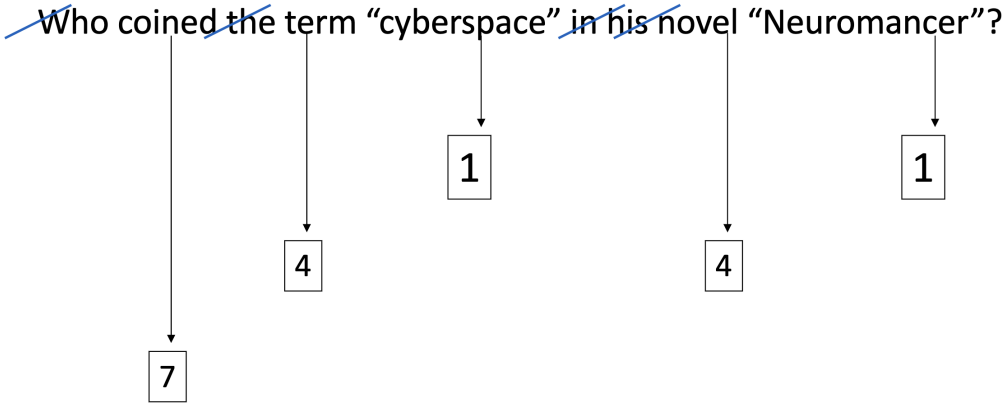
- Detect *question type*, *answer type*, *focus*, *relations*
  - "Who is the president of US?" $\rightarrow$ *person*
- **Formulate queries to send to a search engine**
  - "president of United States"

Keyword Selection Algorithm[5]

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

[5] Dan Moldovan, Sanda Harabagiu, Marius Paca, Rada Mihalcea, Richard Goodrum, Roxana Girju and Vasile Rus. 1999. Proceedings of TREC-8.

Choosing Keywords from the Query [6]

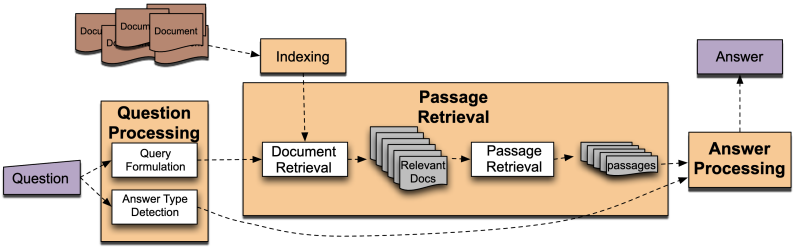


cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

[6] Slide from Mihai Surdeanu

1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. Recent QA Tasks

IF-based Factoid QA Pipeline — Passage Retrieval



- Retrieve *ranked documents*
- Break into suitable passages and **rerank**

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 from around the world.] ["That's some of the most encouraging news about the Amazon rain forest in recent years," said Thomas Lovejoy, an Amazon specialist.] ["It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon."]

[Sixty percent of the Amazon, the world's largest tropical rain forest, lies in Brazil.]

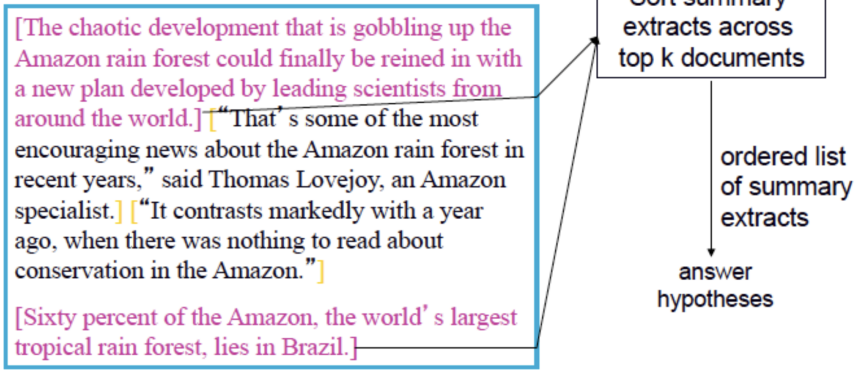
Extract passages that best summarize each document w.r.t. the query

Passage Retrieval as Query-focused Summarization

- Decide on a summary length (**10%** of document length).
- Use *standard ad-hoc retrieval algorithm* to retrieve top k documents.
- Treat each sentence/paragraph in top N documents as a document itself.
  - Use *standard document similarity equations* to assign a similarity score to the sentence/paragraph.
- Return **highest-scoring** sentences/paragraphs as the summary, subject to the length constraint.

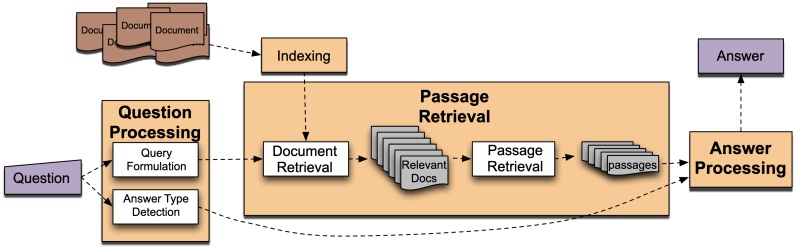
Passage Retrieval as Query-focused Summarization

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



1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. Recent QA Tasks

IF-based Factoid QA Pipeline — Answer Processing



- *Extract* **candidate answers**
- Rank candidates
  - using **evidence** from the text and external sources

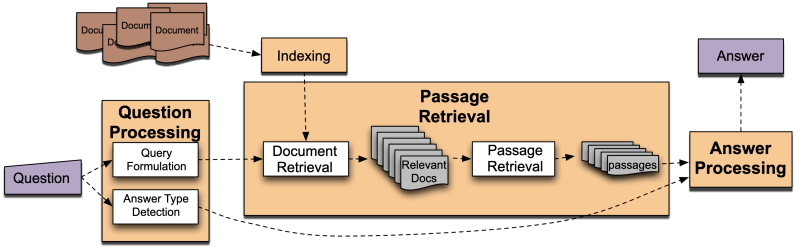
Answer Extraction

- 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:
  - Q: 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.
  - Q: 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*"

IF-based Factoid QA Pipeline — 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

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**

1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. Recent QA Tasks

Common Evaluation Metrics

- **Accuracy** (does answer match gold-labeled answer?)
- **Mean Reciprocal Rank**

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

  - 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: $\frac{1}{2}$
    - 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

1 Introduction to Question Answering

2 Information Retrieval (IR)-based (Factoid) QA
- Factoid QA pipeline
- Question Processing
- Passage Retrieval
- Answer Processing
- Factoid QA Evaluation

3 **Knowledge in QA**

4 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.

1. Introduction to Question Answering
2. Information Retrieval (IR)-based (Factoid) QA
   - Factoid QA pipeline
   - Question Processing
   - Passage Retrieval
   - Answer Processing
   - Factoid QA Evaluation
3. Knowledge in QA
4. **Recent QA Tasks**

Trends

- Reading Comprehension
- Visual Question Answering
- Conversational
- *Leveraging Large Language Models!!*

More in Advanced Natural Language Processing Courses