
Question Answering Systems

Question Answering (QA)

- IR assumes that the user wants an entire document, QA assumes that the user wants a small, focused text result that answers a question
- Answers may not only be phrased with different vocabulary, but they may be implied by other statements or facts
Q: *What year did Marco Polo travel to Asia?*
A: *Marco Polo divulged the truth after returning in 1292 from his travels, which included several months on Sumatra.*
- QA systems must apply more NLP analysis to text in order to find the answers to questions
- The document collections of QA systems range in size from the Web to a targeted collection of documents such as a company's product documents

Typical Traditional QA System

- Traditionally, QA systems applied a two-step strategy
 - They assumed the document collection of a QA system is too large to apply the (time-consuming) NLP processing to all the documents
 - First, use IR to retrieve a set of relevant documents
 - Then process those documents with NLP techniques
 - Identify answers in the documents

Factoid Questions (from TREC)

- *Where is Belize located?*
- *What type of bridge is the Golden Gate Bridge?*
- *What is the population of the Bahamas?*
- *How far away is the moon?*
- *What is Francis Scott Key best known for?*
- *What state has the most Indians?*
- *Who invented the paper clip?*
- *How many dogs pull a sled in the Iditarod?*
- *Where did bocci originate?*
- *Who invented the electric guitar?*
- *Name a flying mammal?*
- *How many hexagons are on a soccer ball?*
- *Who is the leader of India?*

TREC Question Answering Track

- Goal: encourage research into systems that return actual answers
 - From 1999 – 2004 in various forms
- Questions were short and fact based
 - From Encarta and Excite search logs
- Extract or construct answer from set of documents
- For each question
 - Supply up to five ranked answer submissions
 - With the most likely answer ranked first
 - Answer strings were evaluated by NIST's human assessors for correctness
- Evaluation: Mean-Reciprocal Rank (MRR) – score given is the reciprocal of the rank of the first correct answer

Evaluation of Answers

- How much of an answer is enough?

Q: Where will the 2002 winter Olympics be held?

Text: The 2002 Winter Olympics will be held in beautiful Salt Lake City, Utah.

A1: beautiful Salt Lake City, Utah

A2: Salt Lake City, Utah

A3: Salt Lake City

A4: Salt Lake

A5: Utah

- The questioners would probably prefer an entire sentence that established an authority for the answer, perhaps with the answer text highlighted.

Question Classification

- Questions are classified by an answer type taxonomy or question ontology
 - Usually hand-built, but sometimes automatically learned from hand annotated questions
 - Includes types of entities to be found as an answer
 - “Who founded Virgin Airlines?” Expected type: Person
 - “What Canadian city has the largest population?” City
 - Also Location, Time, Quantities etc. (Where, How long, How much)
 - Or more specifically for Flower, Human Description, Human Group, Human Individual, etc.
 - See Figure 23.9 in J&M
 - Other Question types
 - Definition questions, reasons (to answer why questions)
 - Special forms such as birth and death dates
 - “When did George Washington die?”

Answer Processing

- Approaches typically have a suite of answer processing modules that use different techniques to find the answer
- Some are specific to particular question types and expected answer types
- For questions that are looking for particular entity types, find sentences with those types
 - For example, if question is looking for Person, return sentences with named entities of type Person
 - Rank those sentences according to presence of question words
Q: “Who is the prime minister of India?”
Answer sentence: “Monmohan Singh, Prime Minister of India, has told leaders . . . “

Specialty Answer Processing

- Some question types, such as those querying when someone is born or died, are often answered with a fairly small set of sentence constructions
- A special purpose module may be built that contains a number of the patterns of those sentences
 - Often hand-built regular expression patterns
 - Define patterns of sentences to answer the question
 - Use parsed sentences to match those patterns
 - Patterns include subject/verb forms
 - Appositions
 - Known relations for that type of question
 - Find additional forms by using bootstrapping similar to relation extraction

Answer patterns

- Query Reformulation
 - Rephrase query as a partial answer sentence “Virgin Airlines was founded by X” and try to find semantically similar sentences
- The Web approach
 - Since there are potentially millions of documents on the Web that could answer this question, assume that there is one that answers the question in the same form and vocabulary as the question
 - Search the web to find the answer in the right form

Deeper semantic reasoning

- Query Reformulation
 - Rephrase query as a partial semantic relation
Founded (X, Virgin Airlines)
and try to find or infer instances of this relation from the text
- Semantic processing of the text may have
 - Relation extraction
 - Parsing and semantic role labeling, using the semantic roles as relations (and identifying semantically similar verbs)

“Richard Branson formed Virgin Atlantic Airways in 1984, launched Virgin Mobile in 1999, . . . “

 - Will also need to use knowledge sources such as gazeteers, and some relations such as “near”
 - Can use other forms of real-world knowledge, such as from Wikipedia

Selecting the final answer(s)

- In approaches with multiple answer modules, the results must be combined
 - Answer ranking
 - Confidence levels
- Often difficult to produce comparable confidence scores from different modules

IBM's Watson

- Deep QA project, headed by David Ferrucci *
 - Similar system architecture to typical systems but bigger and faster
- Watson interface allowed the Deep QA system to be used to play Jeopardy by selecting questions, deciding when to “ring in” with a confident answers and how to bet



* Ferrucci: Introduction to “This is Watson”, IBM Journal of Research, 2012

IBM's Watson

- Real time challenge
 - Need answers in under 3 seconds
 - Optimized hardware, based on IBM Power 7, and software
 - Distributed parallel computing for answer modules
- Incorporated NLP, Machine Learning, Knowledge representation
- Question categories contribute to expected answer types
 - Some are entity types:
 - Vice presidents, Canadian cities, Characters in classic literature,
 - Some are more complex:
 - Rhyme time, Starts with “A”

Jeopardy Example Topics and Questions

Recent History:

President under whom the U.S. gave full recognition to Communist China.
(Answer: Jimmy Carter)

Pop Music:

Their grandfather had a number 1 record in 1935; their father, Number 1's in 1958 & 1961; and they hit number 1 in 1990.
(Answer: Gunnar & Matthew Nelson)

Before & After:

The “Jerry Maguire” star who automatically maintains your vehicle’s speed.
(Answer: Tom Cruise control)

Art History:

Unique quality of “First Communion of Anemic Young Girls in the Snow” shown at the 1883 Arts Incoherents Exhibit.
(Answer: all white)

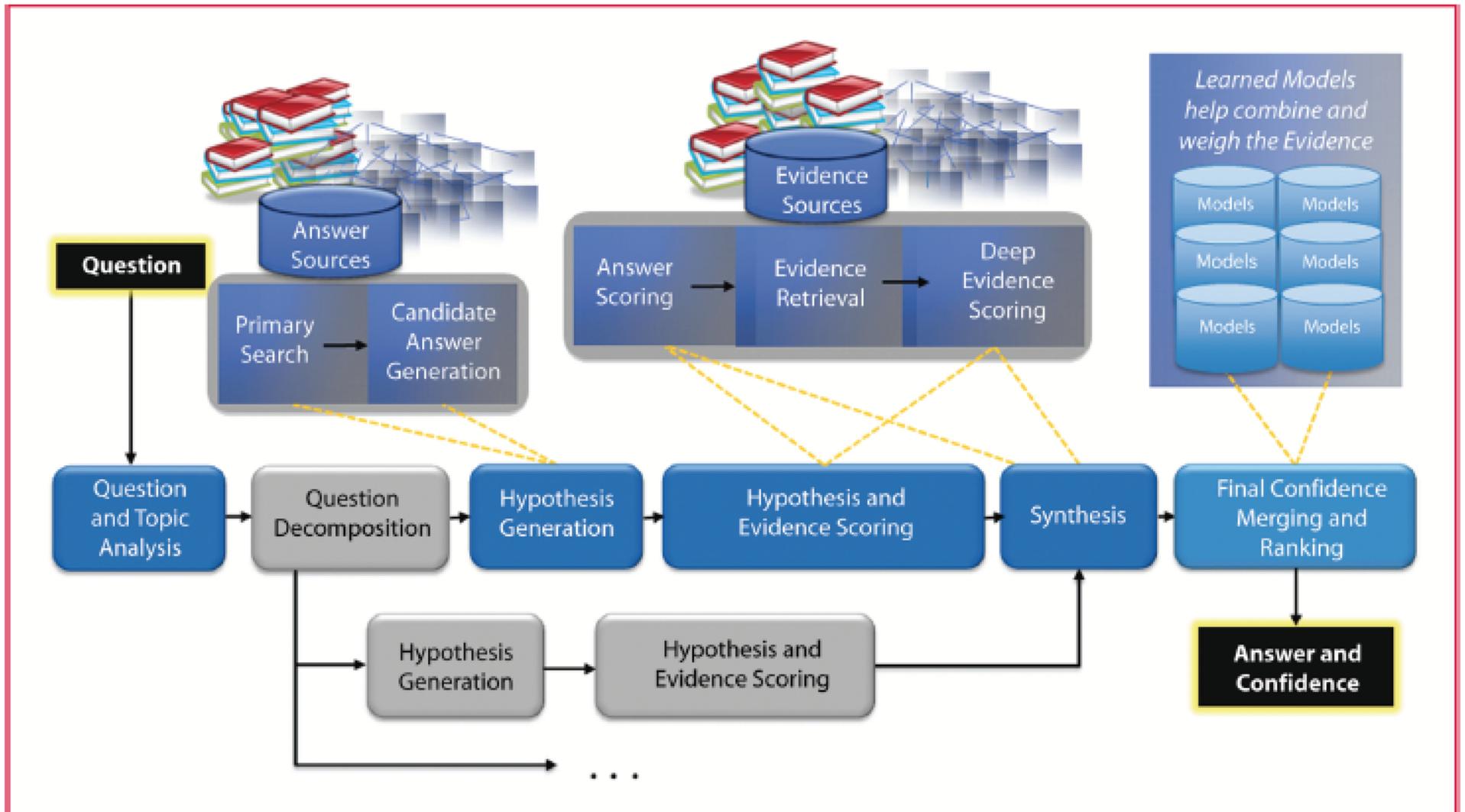
Common Bonds:

Feet, eyebrows and McDonald’s.
(Answer: arches)

Language:

The lead singer of the band Dengue Fever is from this country & often sings in Khmer.
(Answer: Cambodia)

Watson Deep QA Architecture



Watson's Data

- Not connected to the Internet during the game
- Stored approximately 4 terrabytes of data
- “Deep analytics” of text to extract relations and facts
 - Knowledge representation, primarily in the form of relations, both from relation extraction and semantic role labeling
- Processed 200 million pages of structured and unstructured data, including the entire Wikipedia
- Sources included encyclopedias, dictionaries, thesauri, newswire articles, and literary works
 - But only 2% of questions answered directly from structured data; all others used at least partly information from text

Watson's Steps

- Question Analysis: type of question and what the question is asking for
- Hypothesis generation by “thousands” of tasks
 - Quantity over accuracy
- Hypothesis and evidence scoring
 - Collects positive and negative evidence from passages
 - Uses relations between concepts learned from text:
 - Books have authors
 - Authors create characters
- Final merging and ranking
 - Uses past experience to weight types of evidence for different questions
 - Uses confidence of top answer to decide whether to “buzz in”
 - Wagering strategy

Another type of information search: Complex Questions

- Many questions go beyond asking for simple facts
 - Comparison questions
 - Questions that involve deeper understanding of issues
- These questions cannot typically be answered by a simple phrase or sentence
- Approaches focus on identifying an answer passage

Example Passage Retrieval System

- CNLP built a domain-specific QA system to support on-line aeronautics class
 - National Aeronautic & Space Agency funding
- Collection
 - **Textbooks, technical papers / reports, websites**
 - **Pre-selected for relevance and pedagogical value**
- Community of users focused on specific tasks
 - For undergrad students from 2 universities majoring in Aerospace Engineering
 - Students using system for a course can ask questions while working in teams or on own
- Where QA system must function:
 - In **real** time, not batch mode
 - On **real** users' **real** world questions
 - With **real**, not surrogate assessments of relevance
- Used in a collaborative learning environment

Sample Questions from Real Users

- *How difficult is it to mold and shape graphite-epoxies compared with alloys or ceramics that may be used for thermal protective applications?*
- *How doose the shuttle fly?*
- *Do welding sites yield any structural weaknesses that could be a threat for failure?*
- *Are Thermal Protection systems of spacecrafts commonly composed of one panel or a collection of smaller tiles?*
- *How can the aerogels be used in insulation of holes in TPS?*
- **Two-stage QA model**
 - **1st – passage retrieval using expanded query representation**
 - **2nd – selection of answer-providing passages based on generic + specialized entities & relations**