Question-Answering: Overview

Ling573
Systems & Applications
March 31, 2011
Quick Schedule Notes

- No Treehouse this week!

- CS Seminar: Retrieval from Microblogs (Metzler)
  - April 8, 3:30pm; CSE 609
Roadmap

- Dimensions of the problem
- A (very) brief history
- Architecture of a QA system
- QA and resources
- Evaluation
- Challenges
- Logistics Check-in
Dimensions of QA

- Basic structure:
  - Question analysis
  - Answer search
  - Answer selection and presentation
Dimensions of QA

- Basic structure:
  - Question analysis
  - Answer search
  - Answer selection and presentation

- Rich problem domain: Tasks vary on
  - Applications
  - Users
  - Question types
  - Answer types
  - Evaluation
  - Presentation
Applications

- Applications vary by:
  - Answer sources
    - Structured: e.g., database fields
    - Semi-structured: e.g., database with comments
    - Free text
Applications

- Applications vary by:
  - Answer sources
    - Structured: e.g., database fields
    - Semi-structured: e.g., database with comments
    - Free text
      - Web
      - Fixed document collection (Typical TREC QA)
Applications

- Applications vary by:
  - Answer sources
    - Structured: e.g., database fields
    - Semi-structured: e.g., database with comments
    - Free text
      - Web
      - Fixed document collection (Typical TREC QA)
Applications

- Applications vary by:
  - Answer sources
    - Structured: e.g., database fields
    - Semi-structured: e.g., database with comments
    - Free text
      - Web
      - Fixed document collection (Typical TREC QA)
      - Book or encyclopedia
      - Specific passage/article (reading comprehension)
Applications

- Applications vary by:
  - Answer sources
    - Structured: e.g., database fields
    - Semi-structured: e.g., database with comments
    - Free text
      - Web
      - Fixed document collection (Typical TREC QA)
      - Book or encyclopedia
      - Specific passage/article (reading comprehension)
  - Media and modality:
    - Within or cross-language; video/images/speech
Users

- Novice
  - Understand capabilities/limitations of system
Users

- Novice
  - Understand capabilities/limitations of system

- Expert
  - Assume familiar with capabilities
  - Wants efficient information access
  - Maybe desirable/willing to set up profile
Question Types

- Could be factual vs opinion vs summary
Question Types

- Could be factual vs opinion vs summary
- Factual questions:
  - Yes/no; wh-questions
Question Types

- Could be factual vs opinion vs summary
- Factual questions:
  - Yes/no; wh-questions
  - Vary dramatically in difficulty
    - Factoid, List
Question Types

- Could be factual vs opinion vs summary

- Factual questions:
  - Yes/no; wh-questions
  - Vary dramatically in difficulty
    - Factoid, List
    - Definitions
    - Why/how..
Question Types

- Could be factual vs opinion vs summary

- Factual questions:
  - Yes/no; wh-questions
  - Vary dramatically in difficulty
    - Factoid, List
    - Definitions
    - Why/how..
    - Open ended: ‘What happened?’
Question Types

• Could be factual vs opinion vs summary

• Factual questions:
  • Yes/no; wh-questions
  • Vary dramatically in difficulty
    • Factoid, List
    • Definitions
    • Why/how..
    • Open ended: ‘What happened?’

• Affected by form
  • Who was the first president?
Question Types

- Could be factual vs opinion vs summary

- Factual questions:
  - Yes/no; wh-questions
  - Vary dramatically in difficulty
    - Factoid, List
    - Definitions
    - Why/how..
    - Open ended: ‘What happened?’

- Affected by form
  - Who was the first president? Vs Name the first president
Answers

- Like tests!
Answers

- Like tests!
- Form:
  - Short answer
  - Long answer
  - Narrative
Answers

• Like tests!
  • Form:
    • Short answer
    • Long answer
    • Narrative

• Processing:
  • Extractive vs generated vs synthetic
Answers

- Like tests!
  - Form:
    - Short answer
    - Long answer
    - Narrative

- Processing:
  - Extractive vs generated vs synthetic

- In the limit -> summarization
  - What is the book about?
Evaluation & Presentation

- What makes an answer good?
Evaluation & Presentation

- What makes an answer good?
  - Bare answer
Evaluation & Presentation

• What makes an answer good?
  • Bare answer
  • Longer with justification
Evaluation & Presentation

- What makes an answer good?
  - Bare answer
  - Longer with justification
    - Implementation vs Usability

- QA interfaces still rudimentary
  - Ideally should be
Evaluation & Presentation

- What makes an answer good?
  - Bare answer
  - Longer with justification
    - Implementation vs Usability

- QA interfaces still rudimentary
  - Ideally should be
    - Interactive, support refinement, dialogic
(Very) Brief History

- Earliest systems: NL queries to databases (60-s-70s)
- BASEBALL, LUNAR
(Very) Brief History

- Earliest systems: NL queries to databases (60-s-70s)
  - BASEBALL, LUNAR
  - Linguistically sophisticated:
    - Syntax, semantics, quantification, ...
(Very) Brief History

- Earliest systems: NL queries to databases (60-s-70s)
  - BASEBALL, LUNAR
- Linguistically sophisticated:
  - Syntax, semantics, quantification, ,,,
- Restricted domain!
(Very) Brief History

- Earliest systems: NL queries to databases (60-s-70s)
  - BASEBALL, LUNAR
  - Linguistically sophisticated:
    - Syntax, semantics, quantification, ,,,
  - Restricted domain!

- Spoken dialogue systems (Turing!, 70s-current)
  - SHRDLU (blocks world), MIT’s Jupiter, lots more
(Very) Brief History

- Earliest systems: NL queries to databases (60-s-70s)
  - BASEBALL, LUNAR
  - Linguistically sophisticated:
    - Syntax, semantics, quantification, ,,,
  - Restricted domain!
- Spoken dialogue systems (Turing!, 70s-current)
  - SHRDLU (blocks world), MIT’s Jupiter , lots more
- Reading comprehension: (~2000)
(Very) Brief History

- Earliest systems: NL queries to databases (60s-70s)
  - BASEBALL, LUNAR
  - Linguistically sophisticated:
    - Syntax, semantics, quantification, ,,,
  - Restricted domain!

- Spoken dialogue systems (Turing!, 70s-current)
  - SHRDLU (blocks world), MIT’s Jupiter , lots more

- Reading comprehension: (~2000)

- Information retrieval (TREC); Information extraction (MUC)
General Architecture
Basic Strategy

• Given a document collection and a query:
Basic Strategy

- Given a document collection and a query:
- Execute the following steps:
Basic Strategy

- Given a document collection and a query:
- Execute the following steps:
  - Question processing
  - Document collection processing
  - Passage retrieval
  - Answer processing and presentation
  - Evaluation
Basic Strategy

- Given a document collection and a query:
- Execute the following steps:
  - Question processing
  - Document collection processing
  - Passage retrieval
  - Answer processing and presentation
  - Evaluation
- Systems vary in detailed structure, and complexity
AskMSR

- Shallow Processing for QA
Deep Processing Technique for QA

- LCC (Moldovan, Harabagiu, et al)
Query Formulation

- Convert question suitable form for IR
- Strategy depends on document collection
  - Web (or similar large collection):
Query Formulation

- Convert question suitable form for IR
- Strategy depends on document collection
  - Web (or similar large collection):
    - ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
  - Corporate sites (or similar smaller collection):
Query Formulation

- Convert question suitable form for IR
- Strategy depends on document collection
  - Web (or similar large collection):
    - ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
  - Corporate sites (or similar smaller collection):
    - Query expansion
      - Can’t count on document diversity to recover word variation
Query Formulation

- Convert question suitable form for IR
- Strategy depends on document collection
  - Web (or similar large collection):
    - ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
  - Corporate sites (or similar smaller collection):
    - Query expansion
      - Can’t count on document diversity to recover word variation
      - Add morphological variants, WordNet as thesaurus
Query Formulation

- Convert question suitable form for IR

- Strategy depends on document collection
  - Web (or similar large collection):
    - ‘stop structure’ removal:
      - Delete function words, q-words, even low content verbs
  - Corporate sites (or similar smaller collection):
    - Query expansion
      - Can’t count on document diversity to recover word variation
      - Add morphological variants, WordNet as thesaurus
      - Reformulate as declarative: rule-based
        - Where is X located -> X is located in
Question Classification

- Answer type recognition
  - Who
Question Classification

- Answer type recognition
  - Who -> Person
  - What Canadian city ->
Question Classification

- Answer type recognition
  - Who -> Person
  - What Canadian city -> City
  - What is surf music -> Definition

- Identifies type of entity (e.g. Named Entity) or form (biography, definition) to return as answer
Question Classification

• Answer type recognition
  • Who -> Person
  • What Canadian city -> City
  • What is surf music -> Definition

• Identifies type of entity (e.g. Named Entity) or form (biography, definition) to return as answer
  • Build ontology of answer types (by hand)

• Train classifiers to recognize
Question Classification

- Answer type recognition
  - Who --> Person
  - What Canadian city --> City
  - What is surf music --> Definition

- Identifies type of entity (e.g. Named Entity) or form (biography, definition) to return as answer
  - Build ontology of answer types (by hand)

- Train classifiers to recognize
  - Using POS, NE, words
Question Classification

- Answer type recognition
  - Who -> Person
  - What Canadian city -> City
  - What is surf music -> Definition

- Identifies type of entity (e.g. Named Entity) or form (biography, definition) to return as answer
  - Build ontology of answer types (by hand)

- Train classifiers to recognize
  - Using POS, NE, words
  - Synsets, hyper/hypo-nyms
<table>
<thead>
<tr>
<th>Tag</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABBREVIATION</td>
<td></td>
</tr>
<tr>
<td>abb</td>
<td>What’s the abbreviation for limited partnership?</td>
</tr>
<tr>
<td>exp</td>
<td>What does the “e” stand for in the equation E=mc2?</td>
</tr>
<tr>
<td>DESCRIPTION</td>
<td></td>
</tr>
<tr>
<td>definition</td>
<td>What are tammys?</td>
</tr>
<tr>
<td>description</td>
<td>What are the words to the Canadian National anthem?</td>
</tr>
<tr>
<td>manner</td>
<td>How can you get rust stains out of clothing?</td>
</tr>
<tr>
<td>reason</td>
<td>What caused the Titanic to sink?</td>
</tr>
<tr>
<td>ENTITY</td>
<td></td>
</tr>
<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>
<tr>
<td>HUMAN</td>
<td></td>
</tr>
<tr>
<td>----------------------------</td>
<td>-----------------------------------------------------------------</td>
</tr>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LOCATION</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NUMERIC</th>
<th></th>
</tr>
</thead>
<tbody>
<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>
Passage Retrieval

- Why not just perform general information retrieval?
Passage Retrieval

- Why not just perform general information retrieval?
  - Documents too big, non-specific for answers
- Identify shorter, focused spans (e.g., sentences)
Passage Retrieval

- Why not just perform general information retrieval?
  - Documents too big, non-specific for answers

- Identify shorter, focused spans (e.g., sentences)
  - Filter for correct type: answer type classification
  - Rank passages based on a trained classifier
    - Features:
      - Question keywords, Named Entities
      - Longest overlapping sequence,
      - Shortest keyword-covering span
      - N-gram overlap b/t question and passage
Passage Retrieval

- Why not just perform general information retrieval?
  - Documents too big, non-specific for answers

- Identify shorter, focused spans (e.g., sentences)
  - Filter for correct type: answer type classification
  - Rank passages based on a trained classifier
    - Features:
      - Question keywords, Named Entities
      - Longest overlapping sequence,
      - Shortest keyword-covering span
      - N-gram overlap b/t question and passage

- For web search, use result snippets
Answer Processing

- Find the specific answer in the passage
Answer Processing

- Find the specific answer in the passage

- Pattern extraction-based:
  - Include answer types, regular expressions

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;AP&gt; such as &lt;QP&gt;</td>
<td>What is autism?</td>
<td>“, developmental disorders such as autism”</td>
</tr>
<tr>
<td>&lt;QP&gt;, a &lt;AP&gt;</td>
<td>What is a caldera?</td>
<td>“the Long Valley caldera, a volcanic crater 19 miles long”</td>
</tr>
</tbody>
</table>

- Similar to relation extraction:
  - Learn relation b/t answer type and aspect of question
Answer Processing

- Find the specific answer in the passage

- Pattern extraction-based:
  - Include answer types, regular expressions

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;AP&gt; such as &lt;QP&gt;</td>
<td>What is autism?</td>
<td>“developmental disorders such as autism”</td>
</tr>
<tr>
<td>&lt;QP&gt;, a &lt;AP&gt;</td>
<td>What is a caldera?</td>
<td>“the Long Valley caldera, a volcanic crater 19 miles long”</td>
</tr>
</tbody>
</table>

- Similar to relation extraction:
  - Learn relation b/t answer type and aspect of question
  - E.g. date-of-birth/person name; term/definition
  - Can use bootstrap strategy for contexts, like Yarowsky
  - <NAME> (<BD>-<DD>) or <NAME> was born on <BD>
Answer Processing

- Find the specific answer in the passage

- Pattern extraction-based:
  - Include answer types, regular expressions

<table>
<thead>
<tr>
<th>Pattern</th>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;AP&gt;</code> such as <code>&lt;QP&gt;</code></td>
<td>What is autism?</td>
<td>“, developmental disorders such as autism”</td>
</tr>
<tr>
<td><code>&lt;QP&gt;</code>, a <code>&lt;AP&gt;</code></td>
<td>What is a caldera?</td>
<td>“the Long Valley caldera, a volcanic crater 19 miles long”</td>
</tr>
</tbody>
</table>

- Similar to relation extraction:
  - Learn relation b/t answer type and aspect of question
    - E.g. date-of-birth/person name; term/definition
  - Can use bootstrap strategy for contexts
  - `<NAME>` (<BD>-<DD>) or `<NAME>` was born on `<BD>`
Resources

- System development requires resources
- Especially true of data-driven machine learning
Resources

- System development requires resources
  - Especially true of data-driven machine learning

- QA resources:
  - Sets of questions with answers for development/test
Resources

- System development requires resources
  - Especially true of data-driven machine learning

- QA resources:
  - Sets of questions with answers for development/test
    - Specifically manually constructed/manually annotated
Resources

- System development requires resources
  - Especially true of data-driven machine learning

- QA resources:
  - Sets of questions with answers for development/test
    - Specifically manually constructed/manually annotated
    - ‘Found data’
Resources

- System development requires resources
  - Especially true of data-driven machine learning

- QA resources:
  - Sets of questions with answers for development/test
    - Specifically manually constructed/manually annotated
    - ‘Found data’
    - Trivia games!!!, FAQs, Answer Sites, etc
Resources

- System development requires resources
  - Especially true of data-driven machine learning

- QA resources:
  - Sets of questions with answers for development/test
    - Specifically manually constructed/manually annotated
    - ‘Found data’
      - Trivia games!!!, FAQs, Answer Sites, etc
      - Multiple choice tests (IP???)
Resources

- System development requires resources
  - Especially true of data-driven machine learning

- QA resources:
  - Sets of questions with answers for development/test
    - Specifically manually constructed/manually annotated
    - ‘Found data’
      - Trivia games!!!, FAQs, Answer Sites, etc
      - Multiple choice tests (IP???)
      - Partial data: Web logs – queries and click-throughs
Information Resources

- Proxies for world knowledge:
  - WordNet: Synonymy; IS-A hierarchy
Information Resources

- Proxies for world knowledge:
  - WordNet: Synonymy; IS-A hierarchy
  - Wikipedia
Information Resources

- Proxies for world knowledge:
  - WordNet: Synonymy; IS-A hierarchy
  - Wikipedia
  - Web itself
  - ....

- Term management:
  - Acronym lists
  - Gazetteers
  - ....
Software Resources

- General: Machine learning tools
Software Resources

- General: Machine learning tools
- Passage/Document retrieval:
  - Information retrieval engine:
    - Lucene, Indri/lemur, MG
  - Sentence breaking, etc..
Software Resources

- General: Machine learning tools

- Passage/Document retrieval:
  - Information retrieval engine:
    - Lucene, Indri/lemur, MG
    - Sentence breaking, etc..

- Query processing:
  - Named entity extraction
  - Synonymy expansion
  - Parsing?
Software Resources

- General: Machine learning tools
- Passage/Document retrieval:
  - Information retrieval engine:
    - Lucene, Indri/lemur, MG
    - Sentence breaking, etc..
- Query processing:
  - Named entity extraction
  - Synonymy expansion
  - Parsing?
- Answer extraction:
  - NER, IE (patterns)
Evaluation

- Candidate criteria:
  - Relevance
  - Correctness
  - Conciseness:
    - No extra information
  - Completeness:
    - Penalize partial answers
  - Coherence:
    - Easily readable
  - Justification

- Tension among criteria
Evaluation

- Consistency/repeatability:
  - Are answers scored reliability
Evaluation

- Consistency/repeatability:
  - Are answers scored reliably?

- Automation:
  - Can answers be scored automatically?
  - Required for machine learning tune/test
Evaluation

- Consistency/repeatability:
  - Are answers scored reliability?

- Automation:
  - Can answers be scored automatically?
  - Required for machine learning tune/test
    - Short answer answer keys
      - Litkowski’s patterns
Evaluation

- Classical:
  - Return ranked list of answer candidates
Evaluation

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score
  - Measure: Mean Reciprocal Rank (MRR)
Evaluation

- Classical:
  - Return ranked list of answer candidates
  - Idea: Correct answer higher in list => higher score

- Measure: Mean Reciprocal Rank (MRR)
  - For each question,
    - Get reciprocal of rank of first correct answer
      - E.g. correct answer is 4 => \( \frac{1}{4} \)
      - None correct => 0
    - Average over all questions

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

- Applications
Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
  - Factoid -> List, relation, etc
- Answer types
Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
  - Factoid -> List, relation, etc
- Answer types
  - Predominantly extractive, short answer in context
- Evaluation:
Dimensions of TREC QA

- Applications
  - Open-domain free text search
  - Fixed collections
  - News, blogs
- Users
  - Novice
- Question types
  - Factoid -> List, relation, etc
- Answer types
  - Predominantly extractive, short answer in context
- Evaluation:
  - Official: human; proxy: patterns
- Presentation: One interactive track