Question-Answering: Overview

Ling573
Systems & Applications
April 4, 2013

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 capabilties
  - 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 (60s-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

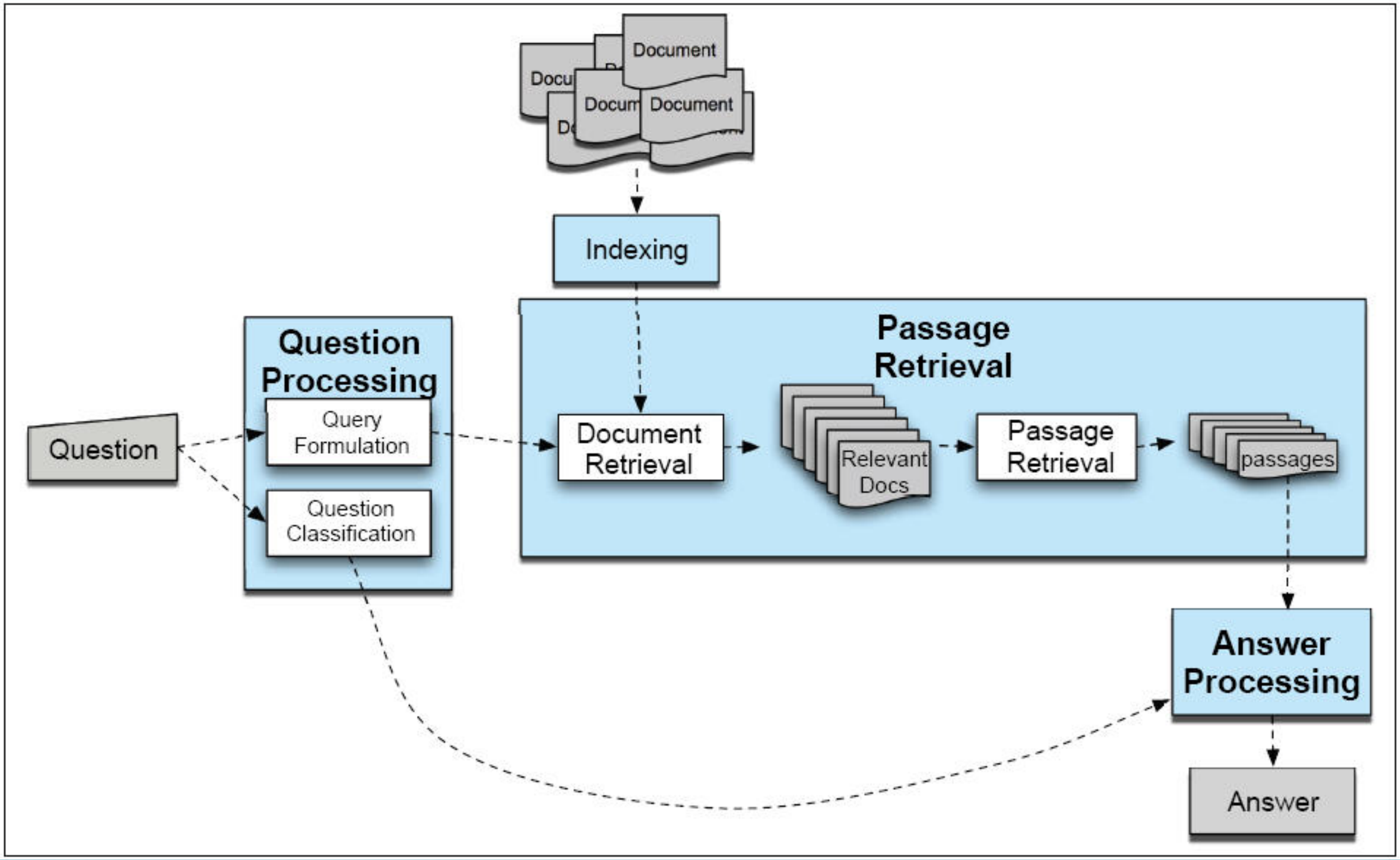
(Very) Brief History

- 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 (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)
- 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

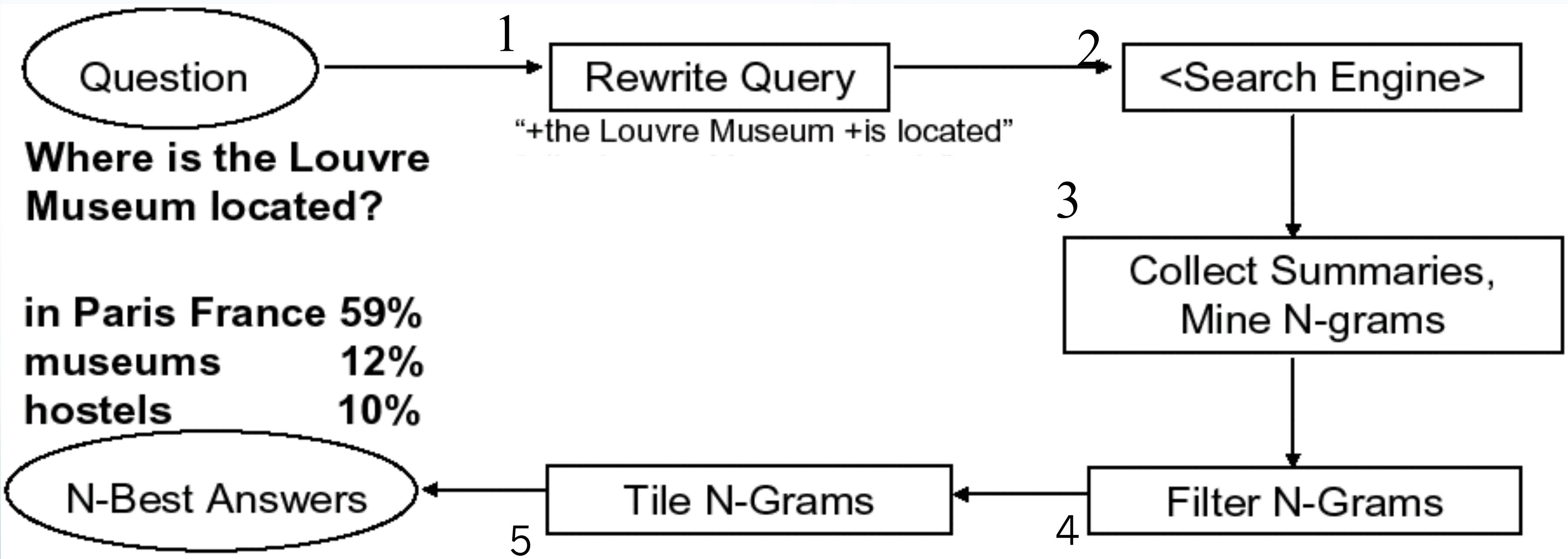
- Given a document collection and a query:
- Execute the following steps:
  - Question processing
  - Document collection processing
  - Passage retrieval
  - Answer processing and presentation
  - Evaluation

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

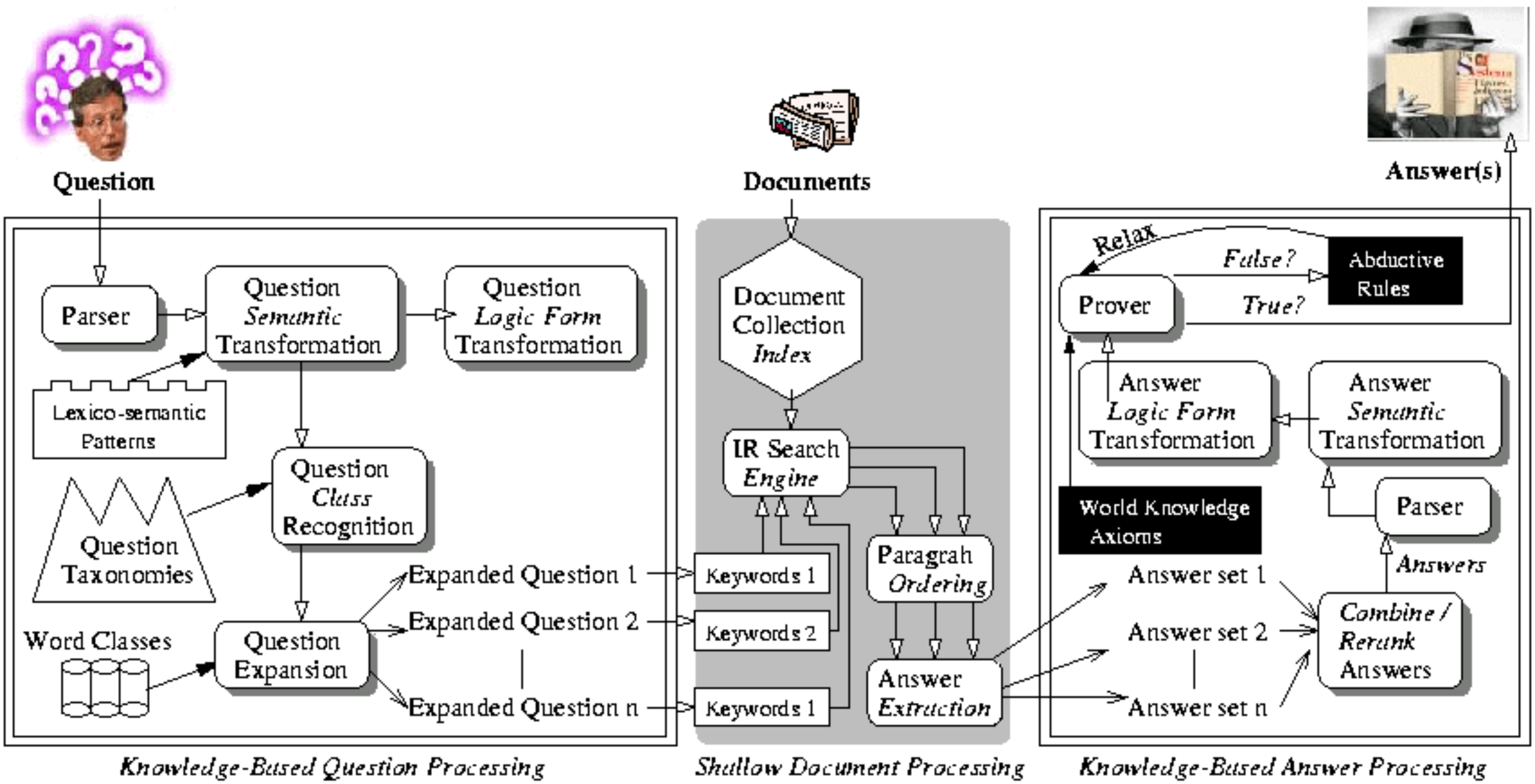
AskMSR

- Shallow Processing for QA



Deep Processing Technique for QA

- LCC (Moldovan, Harabagiu, et al)



Query Formulation

- Convert question to 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

Tag	Example
ABBREVIATION	
abb	What's the abbreviation for limited partnership?
exp	What does the "c" stand for in the equation E=mc2?
DESCRIPTION	
definition	What are tannins?
description	What are the words to the Canadian National anthem?
manner	How can you get rust stains out of clothing?
reason	What caused the Titanic to sink ?
ENTITY	
animal	What are the names of Odin's ravens?
body	What part of your body contains the corpus callosum?
color	What colors make up a rainbow ?
creative	In what book can I find the story of Aladdin?
currency	What currency is used in China?
disease/medicine	What does Salk vaccine prevent?
event	What war involved the battle of Chapultepec?
food	What kind of nuts are used in marzipan?
instrument	What instrument does Max Roach play?
lang	What's the official language of Algeria?
letter	What letter appears on the cold-water tap in Spain?
other	What is the name of King Arthur's sword?
plant	What are some fragrant white climbing roses?
product	What is the fastest computer?
religion	What religion has the most members?
sport	What was the name of the ball game played by the Mayans?
substance	What fuel do airplanes use?
symbol	What is the chemical symbol for nitrogen?
technique	What is the best way to remove wallpaper?
term	How do you say " Grandma " in Irish?
vehicle	What was the name of Captain Bligh's ship?
word	What's the singular of dice?

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

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

Pattern	**Question**	**Answer**
<AP> such as <QP>	What is autism?	", developmental disorders such as autism"
<QP>, a <AP>	What is a caldera?	"the Long Valley caldera, a volcanic crater 19 miles long"

  - 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

Pattern	Question	Answer
<AP> such as <QP>	What is autism?	", developmental disorders such as autism"
<QP>, a <AP>	What is a caldera?	"the Long Valley caldera, a volcanic crater 19 miles long"

  - 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 reliability?

- 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 => ¼
        - None correct => 0
    - Average over all questions

$$MRR = \frac{\sum_{i=1}^{N} \frac{1}{rank_i}}{N}$$

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

Watson & Jeopardy!™ vs QA

- QA vs Jeopardy!
- TREC QA systems on Jeopardy! task
- Design strategies
- Watson components
- DeepQA on TREC

TREC QA vs Jeopardy!

- Both:

TREC QA vs Jeopardy!

- Both:
  - Open domain ‘questions’; factoids
- TREC QA:
  - ‘Small’ fixed doc set evidence, can access Web
  - No timing, no penalty for guessing wrong, no betting

TREC QA vs Jeopardy!

- Both:
  - Open domain ‘questions’; factoids
- TREC QA:
  - ‘Small’ fixed doc set evidence, can access Web
  - No timing, no penalty for guessing wrong, no betting
- Jeopardy!:
  - Timing, confidence key; betting
  - Board; Known question categories; Clues & puzzles
  - No live Web access, no fixed doc set

TREC QA Systems for Jeopardy!

- TREC QA somewhat similar to Jeopardy!

TREC QA Systems for Jeopardy!

- TREC QA somewhat similar to Jeopardy!
- Possible approach: extend existing QA systems
  - IBM's PIQUANT:
    - Closed document set QA, in top 3 at TREC: 30+%
  - CMU's OpenEphyra:
    - Web evidence-based system: 45% on TREC2002

TREC QA Systems for Jeopardy!

- TREC QA somewhat similar to Jeopardy!
- Possible approach: extend existing QA systems
  - IBM's PIQUANT:
    - Closed document set QA, in top 3 at TREC: 30+%
  - CMU's OpenEphyra:
    - Web evidence-based system: 45% on TREC2002
- Applied to 500 random Jeopardy questions
  - Both systems under 15% overall
    - PIQUANT ~45% when 'highly confident'

DeepQA Design Strategies

- Massive parallelism
  - Consider multiple paths and hypotheses

DeepQA Design Strategies

- Massive parallelism
  - Consider multiple paths and hypotheses
- Combine experts
  - Integrate diverse analysis components

DeepQA Design Strategies

- Massive parallelism
  - Consider multiple paths and hypotheses
- Combine experts
  - Integrate diverse analysis components
- Confidence estimation:
  - All components estimate confidence; learn to combine

DeepQA Design Strategies

- Massive parallelism
  - Consider multiple paths and hypotheses
- Combine experts
  - Integrate diverse analysis components
- Confidence estimation:
  - All components estimate confidence; learn to combine
- Integrate shallow/deep processing approaches

Watson Components: Content

- Content acquisition:
  - Corpora: encyclopedias, news articles, thesauri, etc
  - Automatic corpus expansion via web search
  - Knowledge bases: DBs, dbPedia, Yago, WordNet, etc

Watson Components: Question Analysis

- Uses
  - "Shallow & deep parsing, logical forms, semantic role labels, coreference, relations, named entities, etc"

Watson Components: Question Analysis

- Uses
  - “Shallow & deep parsing, logical forms, semantic role labels, coreference, relations, named entities, etc”
- Question analysis: question types, components
- Focus & LAT detection:
  - Finds lexical answer type and part of clue to replace with answer

Watson Components: Question Analysis

- Uses
  - "Shallow & deep parsing, logical forms, semantic role labels, coreference, relations, named entities, etc"
- Question analysis: question types, components
- Focus & LAT detection:
  - Finds lexical answer type and part of clue to replace with answer
- Relation detection: Syntactic or semantic rel's in Q
- Decomposition: Breaks up complex Qs to solve

Watson Components: Hypothesis Generation

- Applies question analysis results to support search in resources and selection of answer candidates

Watson Components: Hypothesis Generation

- Applies question analysis results to support search in resources and selection of answer candidates
- ‘Primary search’:
  - Recall-oriented search returning 250 candidates
  - Document- & passage-retrieval as well as KB search

Watson Components: Hypothesis Generation

- Applies question analysis results to support search in resources and selection of answer candidates
- 'Primary search':
  - Recall-oriented search returning 250 candidates
  - Document- & passage-retrieval as well as KB search
- Candidate answer generation:
  - Recall-oriented extracted of specific answer strings
    - E.g. NER-based extraction from passages

Watson Components: Filtering & Scoring

- Previous stages generated 100s of candidates
  - Need to filter and rank

Watson Components: Filtering & Scoring

- Previous stages generated 100s of candidates
  - Need to filter and rank
- Soft filtering:
  - Lower resource techniques reduce candidates to ~100

Watson Components: Filtering & Scoring

- Previous stages generated 100s of candidates
  - Need to filter and rank
- Soft filtering:
  - Lower resource techniques reduce candidates to ~100
- Hypothesis & Evidence scoring:
  - Find more evidence to support candidate
    - E.g. by passage retrieval augmenting query with candidate
  - Many scoring fns and features, including IDF-weighted overlap, sequence matching, logical form alignment, temporal and spatial reasoning, etc, etc..

Watson Components: Answer Merging and Ranking

- Merging:
  - Uses matching, normalization, and coreference to integrate different forms of same concept
    - e.g., ‘President Lincoln’ with ‘Honest Abe’

Watson Components: Answer Merging and Ranking

- Merging:
  - Uses matching, normalization, and coreference to integrate different forms of same concept
    - e.g., ‘President Lincoln’ with ‘Honest Abe’
- Ranking and Confidence estimation:
  - Trained on large sets of questions and answers
  - Metalearner built over intermediate domain learners
    - Models built for different question classes

Watson Components: Answer Merging and Ranking

- Merging:
  - Uses matching, normalization, and coreference to integrate different forms of same concept
    - e.g., ‘President Lincoln’ with ‘Honest Abe’
- Ranking and Confidence estimation:
  - Trained on large sets of questions and answers
  - Metalearner built over intermediate domain learners
    - Models built for different question classes
- Also tuned for speed, trained for strategy, betting

Retuning to TREC QA

- DeepQA system augmented with TREC-specific:
  - Question analysis and classification
  - Answer extraction

  - Used PIQUANT and OpenEphyra answer typing

Retuning to TREC QA

- DeepQA system augmented with TREC-specific:
  - Question analysis and classification
  - Answer extraction

  - Used PIQUANT and OpenEphyra answer typing

  - 2008: Unadapted: 35% -> Adapted: 60%
  - 2010: Unadapted: 51% -> Adapted: 67%

Summary

- Many components, analyses similar to TREC QA
  - Question analysis →Passage Retrieval → Answer extr.
    - May differ in detail, e.g. complex puzzle questions
- Some additional:
  - Intensive confidence scoring, strategizing, betting
- Some interesting assets:
  - Lots of QA training data, sparring matches
- Interesting approaches:
  - Parallel mixtures of experts; breadth, depth of NLP