Question Answering

CS486 / 686

University of Waterloo

Lecture 23: April 1st, 2014

Question Answering

- Extension to search engines

Watson at Jeopardy

Siri Interface
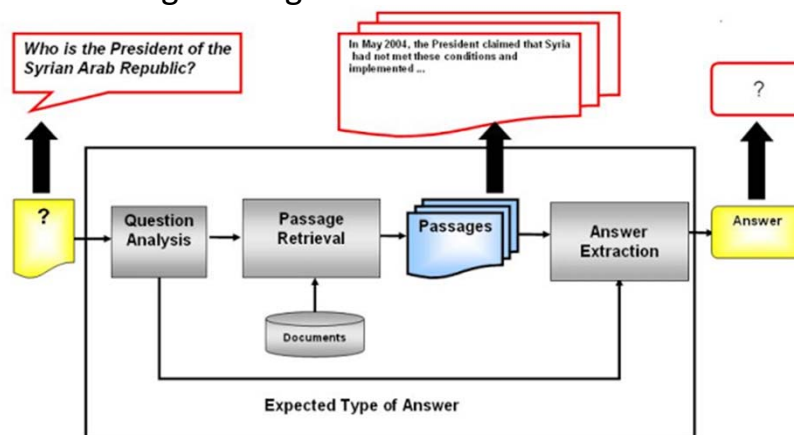
Question Answering

- Question: Natural language (text or speech)
  - E.g., Where was Gauss born?

- Answer retrieved from
  - A database
    - E.g. Triples: entity-relation-entity
  - Corpus of natural language text
    - E.g., the web

Common Architecture

- Lots of engineering

Three Stages

- Question analysis
  - Question classification
  - Query formulation

- Passage Retrieval
  - Document retrieval
  - Paragraph/sentence retrieval

- Answer extraction
  - Answer formulation

Question Classification

- Infer the type of answer that is expected
- Examples:
  - Location: Where was Gauss born?
  - Date/time: When did the US become independent?
  - Person: Who was the first man to walk on the moon?
  - Sport: What is the national sport of Canada?
  - Yes/No: Is it raining outside?
  - Money: How much does this cost?

Question Typology

- Li & Roth 2005
- TREC-2002
  - 1000 questions
  - # of questions/class

Class	#	Class	#
ABBREVIATION	18	term	19
abbreviation	2	vehicle	7
expression	16	word	0
DESCRIPTION	153	**HUMAN**	171
definition	126	group	24
description	13	individual	140
manner	7	title	4
reason	7	description	3
ENTITY	174	**LOCATION**	195
animal	27	city	44
body	5	country	21
color	12	mountain	5
creative	14	other	114
currency	8	state	11
disease/medicine	3	**NUMERIC**	289
event	6	code	1
food	7	count	22
instrument	1	date	146
lang	3	distance	38
letter	0	money	9
other	19	order	0
plant	7	other	24
product	9	period	18
religion	1	percent	7
sport	3	speed	9
substance	20	temp	7
symbol	2	vol.size	4
technique	1	weight	4

Classification Techniques

- Rule-based systems
  - E.g., Hovy et al. 2002: 276 hand-written rules associated with approximately 180 answer types

- Machine learning
  - E.g., Li & Roth 2005: supervised learning
    - Mapping from question features to answer types

Supervised learning

- Dataset: labeled questions from TREC competitions
- Features:
  - Words in the question
  - Syntactic features: part-of-speech tags, syntactic chunks, first head verb and head noun in the question,
  - Semantic features: word net synsets, types of named entities, presence of specific relations
- Algorithms:
  - Naïve Bayes, decision tree, support vector machine, logistic regression, etc.

CS486/686 Slides (c) 2014 P. Poupart   11

Common Architecture

- Lots of engineering
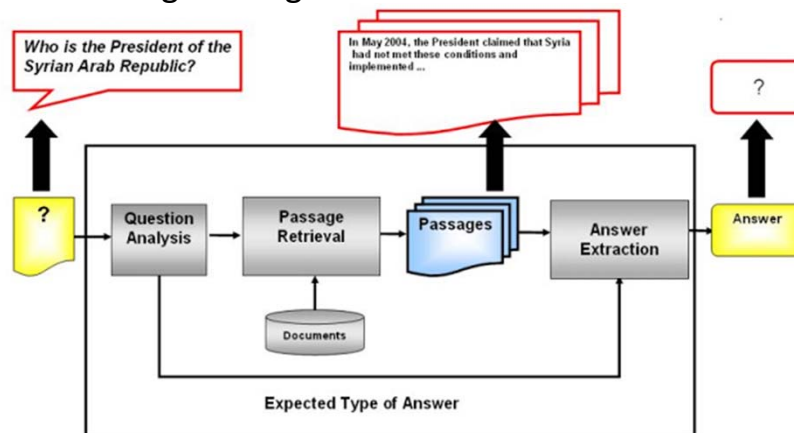
CS486/686 Slides (c) 2014 P. Poupart   12

Query Formulation

- Transform question into a query that a search engine can use to retrieve relevant passages

- Possible transformations:
  - No transformation (keep question as is)
  - Correct typos with a spell checker
  - Keyword extraction (i.e, remove stop words)
  - Keyword expansion (i.e., add synonyms of keywords)
  - Find closest question in database of common questions
  - Entity/relation/concept extraction
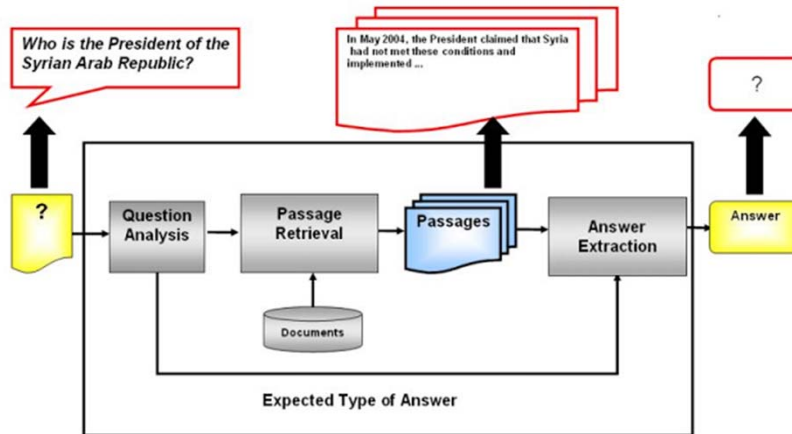
Passage Retrieval

- Task: given some relevant documents, retrieve candidate sentences, paragraphs or sections

- Rule-based systems or machine learning techniques

- Rules or features based on the presence/absence of
  - The desired answer type
  - Question keywords or n-grams
  - Question entities and relations

Common Architecture

- Lots of engineering

Answer Type Detection

- Factoid Questions
  - Location, person, organization, date/time, etc.
  - Answer often consists of one or a few words
  - Answer types often correspond to entity types

- Non-Factoid Questions
  - Definitions, lists, descriptions, reasons, etc.
  - Answer often consists of a phrase (multiple words)
  - Answer types may be detected by patterns, e.g., regular expressions

Non-Factoid Questions

- Examples
  - What is autism?
    - "… development disorders such as autism…"
  - What is caldera?
    - "… the Long Valley caldera, a volcanic crater 19 km long …"

- Pattern-based detection
  - <AP> such as <QP>
  - <QP>, a <AP>

  AP means answer phrase and QP means question phrase

Answer Extraction

- Task: extract one or several answers consisting of sequences of words ordered by relevance
  - Keep surrounding text to help users verify/understand the answer

- Example: How tall is mount Everest?
  1. The official height of mount Everest is 29035 feet
  2. Mount Everest is the highest peak in the Himalayas at more than 8km high
  3. Mount Everest is taller than 25000 feet

State of the Art

- Conferences for the evaluation of QA systems
  - TREC (Text REtrieval Conference) (USA)
    - Question Answering Track
  - NTCIR (National institute of informatics Test Collection for Information Retrieval) (Japan)
    - Multi-lingual question answering track
    - E.g., Questions asked in English, but answers returned in Japanese
  - CLEF (Conferences & Labs of the Evaluation Forum) (EU)
    - QA4MRE track (Question Answering for Machine Reading Evaluation)
    - Extract or verify answer in a single document

Watson at Jeopardy

Jeopardy

- Host reads a clue in the form of an answer
  - But it is really a question
- Contestants respond with a question
  - But it is really an answer

- Clue: When hit by electrons, a phosphor gives off electromagnetic energy in this form.
  - What form of electromagnetic energy does a phosphor give when hit by electrons?
- Response: What is a photon?
  - Photon

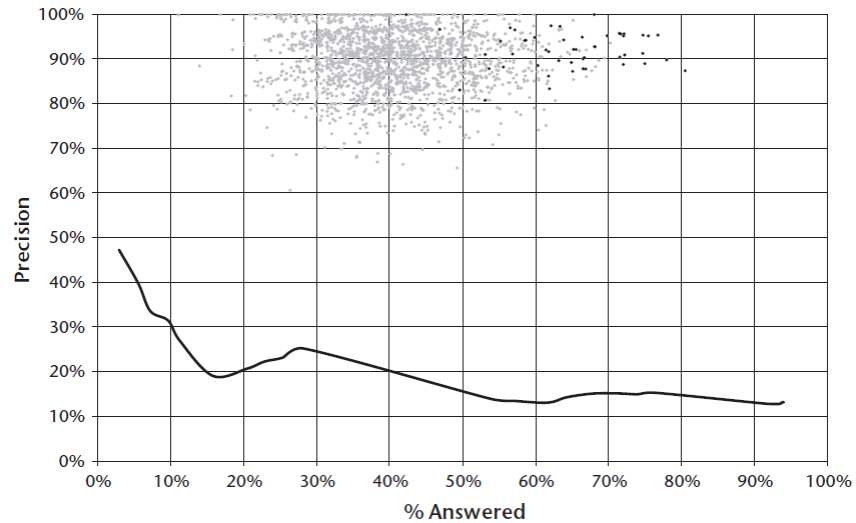CS486/686 Slides (c) 2014 P. Poupart   21

QA Systems in 2007

- Designed for TREC (not Jeopardy)

- Two state of the art QA systems
  - IBM: PIQUANT (Practical Intelligent QUestion ANswering Technology)
  - CMU: OpenEphyra (Open source QA framework)

CS486/686 Slides (c) 2014 P. Poupart   22

PIQUANT vs Jeopardy Champions

Jeopardy vs TREC

Jeopardy

- No specific corpus
- No internet access
- 1-6 seconds per question
- Complex questions
- Confidence is critical
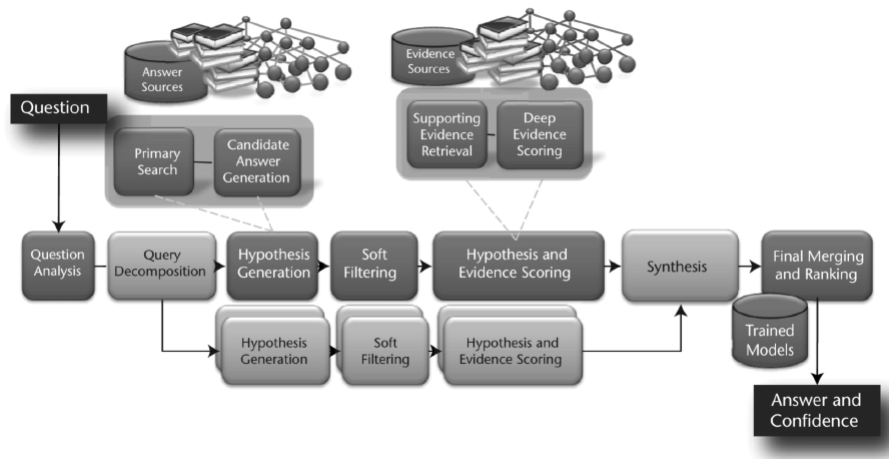
TREC

- Corpus: 1 million docs
- Internet access
- 1 week: answer 500 quest.
- Simple questions
- Confidence not measured

DeepQA Architecture

Key Aspects

1. Ensemble framework
   – Multiple techniques for each component
   – Combine/rank hypotheses produced by each technique

2. Pervasive confidence measures
   – All algorithms produce a hypothesis and a score

Content Acquisition

- No set corpus and no internet access

- Acquisition of relevant content
  - Manual and automated steps
  - encyclopedias, dictionaries, thesauri, newswire articles, literary works
  - Freebase, WordNet, DBPedia, etc.
  - Passages of some web pages

Question Analysis

- Compute shallow parses, deep parses, logical forms, semantic role labels, coreference, relations, named entities

- Question Classification:
  - puzzle question, math question, definition question, named entity, lexical answer type detection

Lexical Answer Type

- When hit by electrons, a phosphor gives off electromagnetic energy in this form.
  - Answer type: definition

- This title character was the crusty and tough city editor of the *Los Angeles Tribune*
  - Answer type: named entity

Hypothesis Generation

- Generate candidate hypotheses from content sources

- text search engines with different approaches
- document search as well as passage search
- knowledge base search
- named entity recognition

- Focus on recall: generate lots of possible hypotheses

Hypothesis scoring
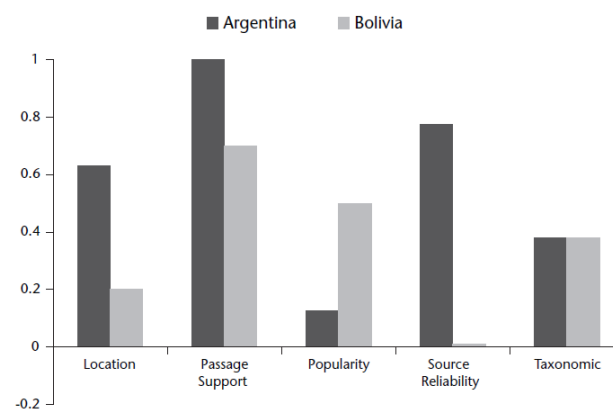
- Focus on precision: filter and rank hypotheses

- Many scoring techniques to verify different dimensions
  - Taxonomic, Geospatial (location), Temporal, Source Reliability, Gender, Name Consistency, Relational, Passage Support, Theory Consistency

CS486/686 Slides (c) 2014 P. Poupart 31

Example

- Chile shares its longest land border with this country.
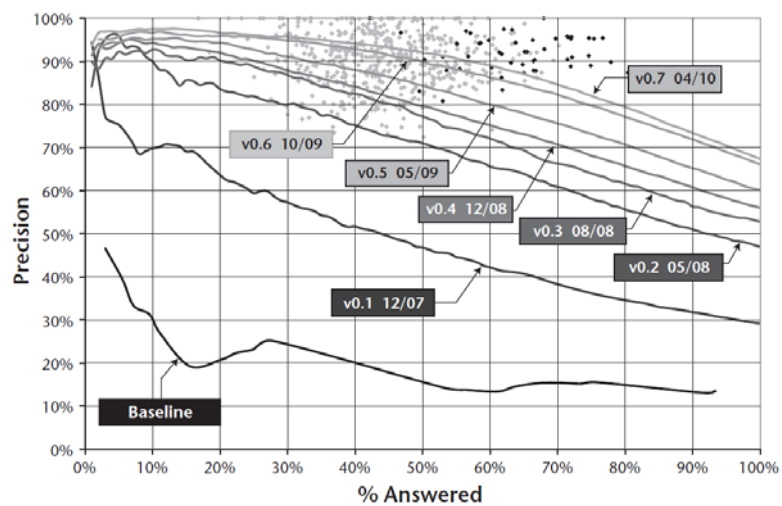

CS486/686 Slides (c) 2014 P. Poupart 32

Ranking

- Combine scores to rank hypotheses
  - Supervised learning
  - Ensemble and hierarchical models

Improvements