

# Question Answering

## Lecture 21: November 20, 2013

CS886-2 Natural Language Understanding  
University of Waterloo

# Question Answering

- Extension to search engines

Where was Gauss born?

**Web** Images Maps More ▾ Search tools

About 1,040,000 results (0.19 seconds)

**Braunschweig, Germany**

Carl Friedrich Gauss, Place of birth

[Feedback/More info](#)

## Watson at Jeopardy



CS886 Lecture Slides (c) 2013 P. Poupart

3

## Siri Interface



CS886 Lecture Slides (c) 2013 P. Poupart

4

## 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

CS886 Lecture Slides (c) 2013 P. Poupart

5

## In Theory

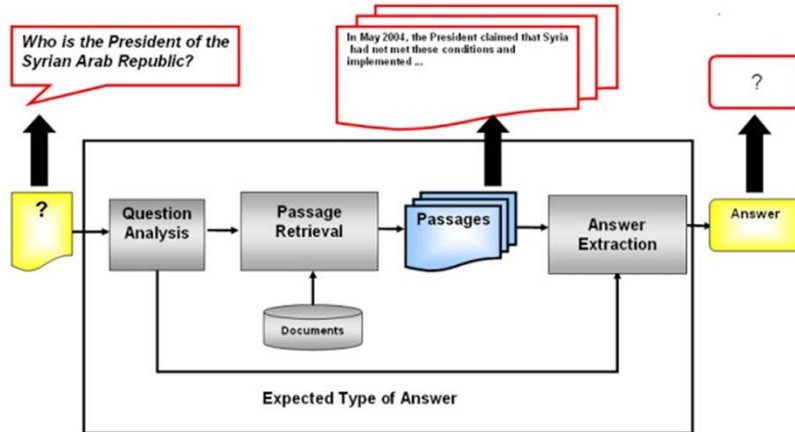
- Question answering is a form of co-reference resolution
- Example: Where was Gauss born?
- Text: Carl Friedrich Gauss is a famous mathematician. He grew up in Braunschweig, Germany.

CS886 Lecture Slides (c) 2013 P. Poupart

6

## In Practice

- Lots of engineering



CS886 Lecture Slides (c) 2013 P. Poupart

7

## Three Stages

- **Question analysis**
  - Question classification
  - Query formulation
- **Passage Retrieval**
  - Document retrieval
  - Paragraph/sentence retrieval
- **Answer extraction**
  - Answer formulation

CS886 Lecture Slides (c) 2013 P. Poupart

8

## 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?

CS886 Lecture Slides (c) 2013 P. Poupart

9

## Question Typology

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

Class	#	Class	#
<b>ABBREVIATION</b>	18	term	19
abbreviation	2	vehicle	7
expression	16	word	0
<b>DESCRIPTION</b>	153	<b>HUMAN</b>	171
definition	126	group	24
description	13	individual	140
manner	7	title	4
reason	7	description	3
<b>ENTITY</b>	174	<b>LOCATION</b>	195
animal	27	city	44
body	5	country	21
color	12	mountain	5
creative	14	other	114
currency	8	state	11
disease/medicine	3	<b>NUMERIC</b>	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	spood	9
substance	20	temp	7
symbol	2	vol.size	4
technique	1	waight	4

CS886 Lecture Slides (c) 2013 P. Poupart

10

## 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

CS886 Lecture Slides (c) 2013 P. Poupart

11

## 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.

CS886 Lecture Slides (c) 2013 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