

The Question-Answering and Chatbot challenges

Konstantinos Skianis, 2nd year Ph.D. student



January 27, 2017

Outline

- 1 Introduction
 - Definition
 - Architecture
 - History

- 2 Question Answering
 - Types of QA
 - Components of QA
 - Question Processing
 - Passage Retrieval and Answer Extraction

- 3 Conversational Agents
 - Introduction
 - Siri
 - Summary
 - Amazon Alexa

1 Introduction

- Definition
- Architecture
- History

2 Question Answering

- Types of QA
- Components of QA
- Question Processing
- Passage Retrieval and Answer Extraction

3 Conversational Agents

- Introduction
- Siri
- Summary
- Amazon Alexa

Question Answering

Definition

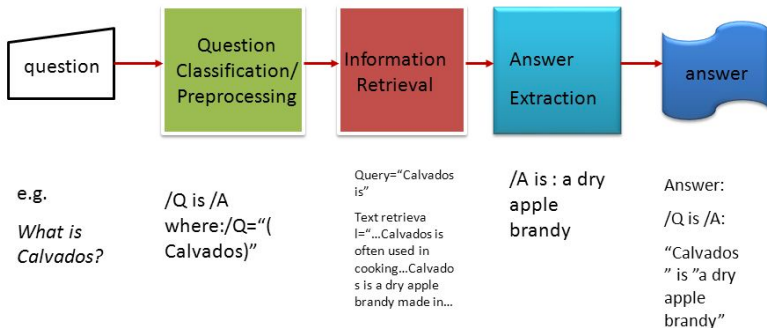
Question Answering (QA) lies within the fields of IR and NLP, which is concerned with building systems that automatically answer questions posed by humans in a natural language.

Corpora

A QA implementation constructs its answers by querying a structured database of knowledge or information (knowledge base), or pull answers from an unstructured collection of natural language documents.

- internal organization documents and web pages
- a set of Wikipedia pages
- a subset of World Wide Web pages

Architecture



Used in Apple's Siri, Wolfram Alpha, IBM Watson.

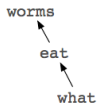
Some history

Old NLP task

Simmons, Klein, McConlogue.1964.Indexing and Dependency Logic for Answering English Questions. American Documentation 15:30, 196-204

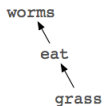
Question:

What do worms eat?

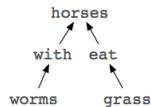


Potential Answers

Worms eat grass



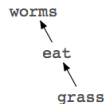
Horses with worms eat grass



Birds eat worms



Grass is eaten by worms



More history

Question Answering: IBM's Watson

Won Jeopardy on February 16, 2011! 1 million dollars prize!

**WILLIAM WILKINSON'S
"AN ACCOUNT OF THE PRINCIPALITIES OF
WALLACHIA AND MOLDOVA"
INSPIRED THIS AUTHOR'S
MOST FAMOUS NOVEL**



Bram Stoker

- 1 Introduction
 - Definition
 - Architecture
 - History

- 2 Question Answering
 - Types of QA
 - Components of QA
 - Question Processing
 - Passage Retrieval and Answer Extraction

- 3 Conversational Agents
 - Introduction
 - Siri
 - Summary
 - Amazon Alexa

Types of questions

Factoid Questions (simple)

- Where is Ecole Polytechnique?
- Who is the current President of the United States of America?

Narrative Questions (hard)

- In children with an acute febrile illness, what is the efficacy of acetaminophen in reducing fever?
- What do scholars think about Jefferson's position on dealing with pirates?

Commercial systems

What do we ask?

Where is the Louvre Museum located?	In Paris, France
What's the abbreviation for limited partnership?	L.P.
What are the names of Odin's ravens?	Huginn and Muninn
What currency is used in China?	The yuan
What kind of nuts are used in marzipan?	almonds
What instrument does Max Roach play?	drums
What is the telephone number for Stanford University?	650-723-2300

what is the capital of greece  


All Images Maps News Shopping More Settings Tools

About 109,000,000 results (1.20 seconds)

Greece / Capital



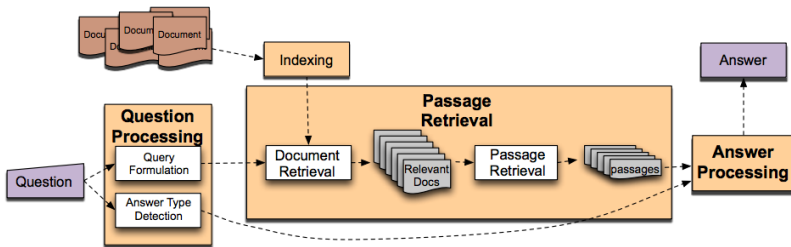
Athens

 Plan a trip and points of interest

Feedback

- People also ask
- What country is Greece in?
 - What Greece is known for?
 - Is Greek in Asia?
 - Is Greece a state?

IR-based Factoid QA



(1) Question Processing

- Detect question type, answer type, focus, relations
- Formulate queries to send to a search engine

(2) Passage Retrieval

- Retrieve ranked documents
- Break into suitable passages and rerank

(3) Answer processing

- Extract candidate answers
- Rank candidates using evidence from text and external sources

Question Processing

Things to extract from the question

- Answer Type Detection
 - Decide the named entity type (person, place) of the answer
- Query Formulation
 - Choose query keywords for the IR system
- Question Type classification
 - Is this a definition question, a math question, a list question?
- Focus Detection
 - Find the question words that are replaced by the answer
- Relation Extraction
 - Find relations between entities in the question

Question Processing

Example

They're the two states you could be reentering if you're crossing Florida's northern border

- Answer Type Detection: US state
- Query: two states, border, Florida, north
- Focus: the two states
- Relations: borders(Florida, ?x, north)

Answer Type Detection: Named Entities

- Who founded Virgin Airlines? **PERSON**
- What Canadian city has the largest population? **CITY**

Answer Type Detection

- Most often, we treat the problem as machine learning classification
 - **Define** a taxonomy of question types
 - **Annotate** training data for each question type
 - **Train** classifiers for each question class using a rich set of features

Features

- Question words and phrases
- Part-of-speech tags
- Parse features (headwords)
- Named Entities
- Semantically related words

Passage Retrieval

- (1) IR engine retrieves documents using query terms
- (2) Segment the documents into shorter units
 - something like paragraphs
- (3) Passage ranking
 - Use answer type to help rerank passages

Features for Passage Ranking

- Number of Named Entities of the right type in passage
- Number of query words in passage
- Number of question N-grams also in passage
- Proximity of query keywords to each other in passage
- Longest sequence of question words
- Rank of the document containing passage

Answer Extraction

- Run an answer-type named-entity tagger on the passages
 - Each answer type requires a named-entity tagger that detects it
 - If answer-type is CITY, tagger has to tag CITY
 - Can be full NER, simple regular expressions, or hybrid
- Return the string with the right type:
 - Who is the prime minister of India? (PERSON)
Manmohan Singh, Prime Minister of India, had told left leaders that the deal would not be renegotiated.
 - How tall is Mt. Everest? (LENGTH)
The official height of Mount Everest is 29035 feet

Ranking Candidate Answers

- But what if there are multiple candidate answers!
Q: Who was Queen Victoria's second son?
- Answer Type: Person
- Passage:
The Marie biscuit is named after Marie Alexandrovna, the daughter of Czar Alexander II of Russia and wife of Alfred, the second son of Queen Victoria and Prince Albert

Machine Learning: features for Ranking Candidate Answers

Answer type match: Candidate contains a phrase with the correct answer type.

Pattern match: Regular expression pattern matches the candidate.

Question keywords: # of question keywords in the candidate.

Keyword distance: Distance in words between the candidate and query keywords

Novelty factor: A word in the candidate is not in the query.

Apposition features: The candidate is an appositive to question terms.

Punctuation location: The candidate is immediately followed by a comma, period, quotation marks, semicolon, or exclamation mark.

Sequences of question terms: The length of the longest sequence of question terms that occurs in the candidate answer.

Common Evaluation Metrics

- (1) **Accuracy** (does answer match gold+labeled answer?)
- (2) **Mean ReciprocalRank**
 - For each query return a ranked list of M candidate answers.
 - Query score is 1/Rank of the first correct answer
 - If first answer is correct:1
 - else if second answer is correct: $\frac{1}{2}$
 - else if third answer is correct: $\frac{1}{3}$, etc.
 - Score is 0 if none of the M answers are correct
 - Take the mean over all N queries

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

Relation Extraction

- Answers: Databases of Relations
 - born-in("Emma Goldman", "June 27 1869")
 - author-of("Cao Xue Qin", "Dream of the Red Chamber")
 - Draw from Wikipedia infoboxes, DBpedia, FreeBase, etc.
- Questions: Extracting Relations in Questions

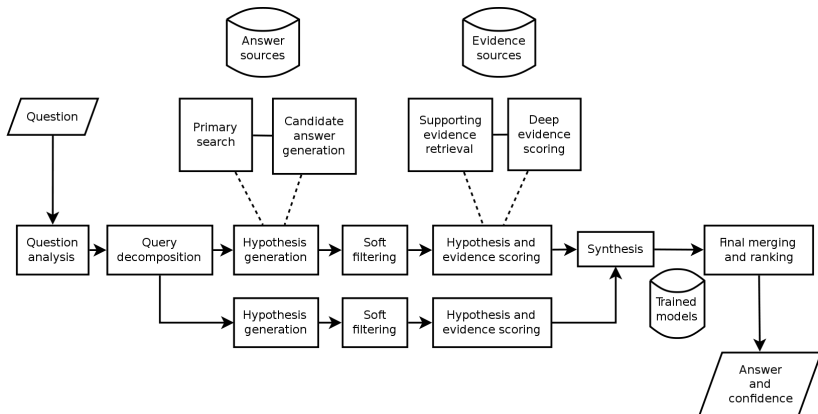
Whose granddaughter starred in E.T.?

(acted-in ?x "E.T.")

(granddaughter-of ?x ?y)

Deep Learning for QA

IBM Watson's DeepQA system



Squad dataset

Test Set Leaderboard

Since the release of our dataset ([and paper](#)), the community has made rapid progress! Here are the ExactMatch (EM) and F1 scores of the best models evaluated on the test and development sets of v1.1.

Rank	Model	Test EM	Test F1
1	r-net (ensemble) (Microsoft Research Asia)	74.8	82.2
2	BiDAF (ensemble) Allen Institute for AI & University of Washington (Seo et al. '16)	73.3	81.1
3	Dynamic Coattention Networks (ensemble) Salesforce Research (Xiong & Zhong et al. '16)	71.6	80.4
4	BiLSTM-Fusion German Research Center for Artificial Intelligence	70.8	78.9
5	r-net (single model) Microsoft Research Asia	70.1	78.8
6	jNet (single model)	68.7	77.4

- 1 Introduction
 - Definition
 - Architecture
 - History

- 2 Question Answering
 - Types of QA
 - Components of QA
 - Question Processing
 - Passage Retrieval and Answer Extraction

- 3 Conversational Agents
 - Introduction
 - Siri
 - Summary
 - Amazon Alexa

Conversational Agents

Conversational Agents AKA Dialog Agents

- Phone-based Personal Assistants: Apple SIRI, Windows Cortana, Google Now, Amazon Alexa
- Talking to your car
- Communicating with robots
- Transaction with banks
- Clinical uses for mental health
- Chatting for fun

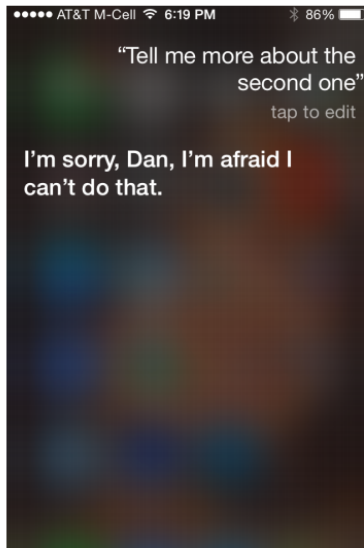
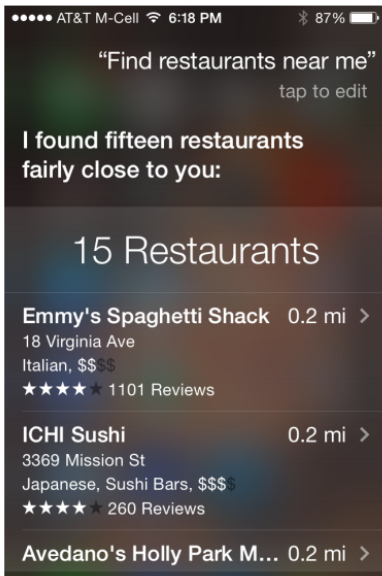
Two classes of systems

- (1) Goal-based dialog agents
 - SIRI, interfaces to cars, robots
- (2) Chatbots

Components

- (1) The user speaks, converts to plain text by the system's input recognizer/decoder with:
 - automatic speech recognizer (ASR), gesture recognizer, handwriting recognizer
- (2) The text is analyzed by a Natural language understanding unit (NLU), which may include:
 - Proper Name identification, POS, Syntactic/semantic parser
- (3) Semantic information is analyzed by the dialog manager, that keeps the history and state of the dialog and manages the general flow.
- (4) Usually, the dialog manager contacts one or more task managers, that have knowledge of the specific task domain.
- (5) The dialog manager produces output with:
 - natural language generator, gesture generator, layout engine
- (6) Finally, the output is rendered using:
 - text-to-speech engine (TTS), talking head, robot or avatar

Siri example



Siri architecture

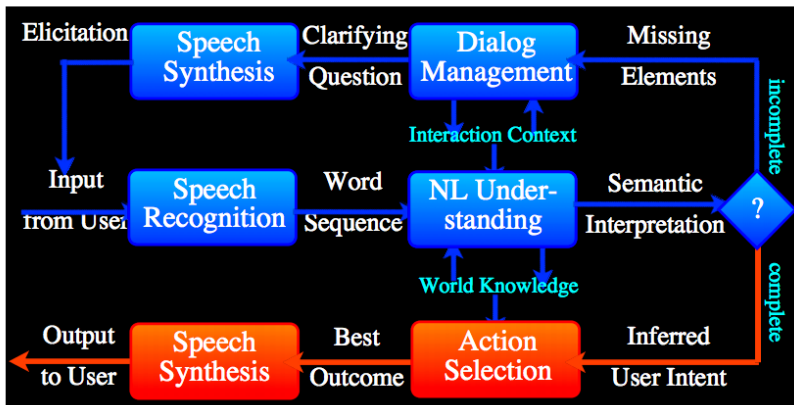


Figure from Jerome Bellegarda

Chatbot architectures

Rule-based

- Pattern-action rules (Eliza)
- A mental model (Parry)

Corpus-based (from large chat corpus)

- Information Retrieval
- Deep neural nets

Neural Net Chatbot

How are you doing?

I'm good, you?

I am off all week next week.

What time you going?

I know you love me and I love you.

I love you too, baby.

Ohh I've never seen that! How long does it take you guys to learn the drill?

Like 2 weeks ago!!

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2015. A Diversity-Promoting Objective Function for Neural Conversation Models.

Summary

Commercial Systems

- Goal-oriented: ontology + hand-written rules for slot fillers
- Non-goal-oriented:
 - Simple rule-based systems
 - IR-based: mine datasets of conversations.

Future

- More use of data
 - IR-based models
 - Deep learning models
- Problems:
 - Acquiring the data
 - Integrating goal-based and chatbot-based systems
 - More generalized models

Amazon Alexa

The Alexa Prize

\$2.5 Million to Advance Conversational Artificial Intelligence

September 2016 – November 2017



Amazon Selects Teams to Compete for Inaugural \$2.5 Million Alexa Prize

Amazon will sponsor 12 university teams to compete in the 2016-2017 Alexa Prize. This year's inaugural competition focuses on the grand challenge of building a socialbot that can converse coherently and engagingly with humans on popular topics for 20 minutes. The sponsored teams will receive a \$100,000 stipend, Alexa-enabled devices, free Amazon Web Services (AWS) services to support their development efforts, and support from the Alexa Skills Kit (ASK) team.

The end!

Thank you! Demo time!