CS544: Question Answering, Question Classification

February 12, 2013

Zornitsa Kozareva
USC/ISI
Marina del Rey, CA
kozareva@isi.edu
www.isi.edu/~kozareva
What do you do when you have a question?

Q> How do you get a perfect sun tan?

Search Engine

The Perfect Suntan Guide
www.maxtimag.com/loose-wires/.../the-perfect-suntan-guide.html
Apr 21, 2006 – Don’t get browned off on holiday - too much sun is more dangerous than crime and terrorists. Follow our handy tips for a perfect safe suntan.

How to Get a Good Tan Without Getting Sunburned: 10 steps ...
www.wikihow.com › ... › Skin Care › Sun and Sunless Tanning
Dec 17, 2011 – How to Get a Good Tan Without Getting Sunburned. ... Apply moisturizer or after sun and don’t scratch the sunburn or it will make it worse. 8 ...

How to Get the Perfect Tan | eHow.com
www.ehow.com › Beauty & Personal Care
Since we know more about the dangers of worshiping the sun, such as the risks of developing skin cancer and wrinkles, getting the perfect tan now includes ...

How to get a great suntan
www.fitnesstipsforlife.com/how-to-get-a-great-suntan.html
***** 3 votes
Exfoliating your skin is a very good idea. When you are getting a suntan the color that you get is a result of the skin reacting to the sun. If you have dead skin on ...
What do you do when you have a question?

Q> How do you get a perfect sun tan?  
A> Lie in the sun
Where do we stand today

• Modern search engines provide document retrieval facilities

• (+) Given a document collection and a list of key words, they return a ranked list of relevant documents (snippets)

• (-) However
  – search engines do not speak our language
  – a document is not an answer to a question
  – relevant documents might not have the answer
  – user spends too much time to find the answer
Question Answering (QA)

• Ask the computer a question in natural language and quickly receive the exact answer (and sufficient context to validate the answer)

Ex. **Question:** Where was Einstein born?

**Answer:** Ulm

**Answer found:** One day, from among his city views of Ulm, Otto chose a water color to send to Albert Einstein as a remembrance of Einstein’s birthplace.
Jeopardy Challenge which IBM’s Watson won on Feb 16th, 2011

WILLIAM WILKINSON’S “AN ACCOUNT OF THE PRINCIPALITIES OF WALLACHIA AND MOLDOVIA” INSPIRED THIS AUTHOR’S MOST FAMOUS NOVEL
Siri. 🎧 Beta
Your wish is its command.
Paradigms for QA

• IR-based approaches
  – TREC; IBM Watson; Google

• Knowledge-based and Hybrid approaches
  – IBM Watson; Apple Siri; Wolfram Alpha; True Knowledge Evi
IR-based Factoid QA

• QUESTION PROCESSING
  – Detect question type, answer type, focus, relations
  – Formulate queries to send to a search engine

• PASSAGE RETRIEVAL
  – Retrieve ranked documents
  – Break into suitable passages and rerank

• ANSWER PROCESSING
  – Extract candidate answers
  – Rank candidates
Simplified Architecture

Human

Question

WEB

Search Engine

Paragraph Sorter

Answer Processor
Question Processing
(Things to extract from the question)

• Answer Type Detection
  – decide the named entity (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

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

- Answer Type: US state
- Query: two states, border, Florida, north
- Focus: the two states
- Relations: borders(Florida, ?x, north)
Many questions can already be answered by web search

Huginn and Muninn - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Huginn_and_Muninn
The names of the ravens are sometimes modernly anglicized as Hugin and Munin. In the Poetic Edda, a disguised Odin expresses that he fears that they may ...
Attestations - Archaeological record - Theories - See also
IR-based Question Answering

Search

About 904,000 results (0.30 seconds)

Everything
Images
Maps
Videos
News

Best guess for Louvre Museum Location is Paris, France
Mentioned on at least 7 websites including wikipedia.org, answers.com and east-buc.k12.ia.us - Show sources - Feedback

Musée du Louvre - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Musée_du_Louvre
Musée du Louvre is located in Paris. Location within Paris. Established, 1793. Location, Palais Royal, Musée du Louvre, 75001 Paris, France. Type, Art museum ...
Louvre Palace - List of works in the Louvre - Category:Musée du Louvre

Example slide from Chris Manning
TREC IR-based QA Systems

• Where do lobsters like to live?  
  — on the table

• Where are zebras most likely found?  
  — in the dictionary

• What is an invertebrate?  
  — Dukakis

Michael Dukakis is a member of the Democratic Party, I have long suspected that elected officials from the Democratic Party are some previously unclassified form of invertebrate, a totally spineless creature capable of great noise but no real movement or action.
Paradigms for QA

• **IR-based approaches**
  – TREC; IBM Watson; Google

• **Knowledge-based and Hybrid approaches**
  – IBM Watson; Apple Siri; Wolfram Alpha; True Knowledge Evi
Knowledge-based approaches (Siri)

- Build a semantic representation of the query
  - Times, dates, locations, entities, numeric quantities

- Map from this semantics to query structured data or resources
  - Geospatial databases
  - Ontologies (Wikipedia infoboxes, DBPedia, WordNet, Yago)
  - Restaurant review sources and reservation services
  - Scientific databases
Hybrid approaches (IBM Watson)

- Build a shallow semantic representation of the query
- Generate answer candidates using IR methods
  - Augmented with ontologies and semi-structured data
- Score each candidate using richer knowledge sources
  - Geospatial databases
  - Temporal reasoning
  - Taxonomical classification
## Commercial Systems

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Where is the Louvre Museum located?</td>
<td>In Paris, France</td>
</tr>
<tr>
<td>What’s the abbreviation for limited partnership?</td>
<td>L.P.</td>
</tr>
<tr>
<td>What are the names of Odin’s ravens?</td>
<td>Huginn and Muninn</td>
</tr>
<tr>
<td>What currency is used in China?</td>
<td>yuan</td>
</tr>
<tr>
<td>What kind of nuts are used in marzipan?</td>
<td>almonds</td>
</tr>
<tr>
<td>What instrument does Max Roach play?</td>
<td>drums</td>
</tr>
<tr>
<td>What is the telephone number for Stanford University?</td>
<td>650-723-2300</td>
</tr>
</tbody>
</table>

* Handle mainly factoid questions
Question Classification

- The process by which a system analyzes a question and labels the question with a given answer type

Q: “Who was the first Prime Minister of Canada?”

Answer Type: Person

- It is an integral part of any QA system
  - limits the range of data/corpora to be searched
  - makes the QA problem tractable
Task Definition

• Given a set of questions, build an automated system that can learn the mapping between the questions and the answer types
Task Definition

• Given a set of questions, build an automated system that can learn the mapping between the questions and the answer types

• What’s in a Question?
  – Question Type
  – Answer Type
  – Question Focus
  – Question Topic
Question Type

• **Def:** An idiomatic categorization of questions for the purpose of distinguishing between different processing strategies and/or answer formats.

  - FACTOID: “How far is it from Earth to Mars?”
  - LIST: “List the names of chewing gums”
  - DEFINITION: “Who is Vlad the Impaler?”
  - YES-NO: “Is Saddam Hussein alive?”
  - OPINION: “What do most Americans think of gun control?”
  - ...
Answer Type

• **Def:** The class of object (type of sentence) sought by the question.

  - PERSON (from “Who ...”)
  - PLACE (from “Where ...”)
  - DATE (from “When ...”)
  - NUMBER (from “How many ...”)

  but also

  - EXPLANATION (from “Why ...”)
  - METHOD (from “How ...”)


Question Focus

• **Def:** The property or entity that is being sought by the question.

  – “In what state is the Grand Canyon?”
  – “What is the population of Bulgaria?”
  – “What color is a pomegranate?”
Question Focus

• **Def:** The property or entity that is being sought by the question.
  
  – “In what **state** is the Grand Canyon?”
  – “What is the **population** of Bulgaria?”
  – “What **color** is a pomegranate?”
Question Topic

- **Def:** The object (person, place, ...) or event that the question is about. Note that the question might well be about a property of the topic, which will be the question focus.

  - “What is the height of Mt. Everest?”
    - **height** is the focus
    - **Mt. Everest** is the topic
Data Creation

- Who creates the Answer Types?

- How and why are certain Answer Types selected?
  
  - Answer Types are usually tied to the classes recognized by output of Named Entity Recognition Systems
  
  - Answer Types depends on the final application
TREC QA Data Sets

• Each year a new set of questions is created

• The questions are not labeled with expected answer types

• The named entity tags are generated by BBN’s Identifinder named entity tagger

• The repository contains 31 Answer Types
# Examples of Answer Types

<table>
<thead>
<tr>
<th>Example question</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>What do you call a group of geese?</td>
<td>Animal</td>
</tr>
<tr>
<td>Who was Monet?</td>
<td>Biography</td>
</tr>
<tr>
<td>How many types of lemurs are there?</td>
<td>Cardinal</td>
</tr>
<tr>
<td>What is the effect of acid rain?</td>
<td>Cause / Effect / Influence</td>
</tr>
<tr>
<td>What is the street address of the White House?</td>
<td>Contact Info</td>
</tr>
<tr>
<td>Boxing Day is celebrated on what day?</td>
<td>Date</td>
</tr>
<tr>
<td>What is sake?</td>
<td>Definition</td>
</tr>
<tr>
<td>What is another name for nearsightedness?</td>
<td>Disease</td>
</tr>
<tr>
<td>What was the famous battle in 1836 between Texas and Mexico?</td>
<td>Event</td>
</tr>
<tr>
<td>What is the tallest building in Japan?</td>
<td>Facility</td>
</tr>
<tr>
<td>What type of bridge is the Golden Gate Bridge?</td>
<td>Facility Description</td>
</tr>
<tr>
<td>What is the most popular sport in Japan?</td>
<td>Game</td>
</tr>
<tr>
<td>What is the capital of Sri Lanka?</td>
<td>Geo-Political Entity</td>
</tr>
<tr>
<td>Name a Gaelic language.</td>
<td>Language</td>
</tr>
<tr>
<td>What is the world’s highest peak?</td>
<td>Location</td>
</tr>
<tr>
<td>How much money does the Sultan of Brunei have?</td>
<td>Money</td>
</tr>
</tbody>
</table>
Distribution of Answer Types

date, definition, GPE, person \( \sim 57\% \)

facility description, time are seen only 1
Answer Type Taxonomy

Xin Li, Dan Roth. 2002. Learning Question Classifiers. COLING'02

• 6 coarse classes
  – ABBREVIATION, ENTITY, DESCRIPTION, HUMAN, LOCATION, NUMERIC

• 50 finer classes
  – LOCATION: city, country, mountain...
  – HUMAN: group, individual, title, description
  – ENTITY: animal, body, color, currency...
Li & Roth’s Answer Type Taxonomy

Example slide from Chris Manning
Answer Types in Jeopardy


• 2500 answer types in 20,000 Jeopardy question sample
• The most frequent 200 answer types cover < 50% of data
• The 40 most frequent Jeopardy answer types

he, country, city, man, film, state, she, author, group, here, company, president, capital, star, novel, character, woman, river, island, king, song, part, series, sport, singer, actor, play, team, show, actress, animal, presidential, composer, musical, nation, book, title, leader, game
N-gram Features

• N-grams are ordered arrangement of $n$ words
  – $n=1$ unigram
  – $n=2$ bigram
  – ...

Ex. “How far is it from Earth to Mars?”
  unigrams: {how, far, is, it}
  bigrams: {how far, far is, is it ...}

*Is it good to use higher order n-grams? Why?*
Question Word Feature

• Use the question word itself as a separate feature

Ex. “How far is it from Earth to Mars?”
  unigrams: {how, far, is, it}
  question word: {how}

But isn't this very repetitive? Would it actually help or make a difference?

Ex. “What color does litmus paper turn when it comes into contact with a strong acid?”
  avoid misclassifying the question as time
Limitations of the Discussed Features

• They will incorrectly classify a question that has multiple types

Ex. *What* question

– What is the proper name for a female *walrus*?
– What is Nicholas Cage’s *profession*?
– What is the *population* of Seattle?
– What *caused* the Lynmouth floods?
– What is the *telephone number* for the University of Kentucky?
– What is another name for *nearsightedness*?
Semantic Features

• Let’s incorporate some semantics
• Where does semantic information come from?

  – What is the proper name for a female *walrus*?
  
  – What is the *population* of Seattle?

  – What is the tallest *mountain*?

  - What is the tallest *building* in Japan?
WordNet Features

**Headword Extraction Procedure**

For each question:
- run POS-tagger
- find the first noun phrase and assume it is the main one
- extract the rightmost word tagged as noun
- use WordNet to determine all hypernyms of the headword

What is the **tallest building** in Japan?

---

works ~90%
Web-based Features

• For each question and its headword, submit to the Web lexico-syntactic patterns of the form:

  “* such as headword”

• Retrieve and rerank all semantic classes learned from the Web

What is the proper name for a female walrus?
Other Features

• Part-of-speech tags
• Noun phrases
• Question length
• Named Entities (helps generalize the question)

Who is Duke Ellington?  <- Who is <person_name>?
Who is Charles Lindbergh?  <- Who is <person_name>?
Machine Learning Classifier

• The most robust performance is achieved with Support Vector Machines (SVM)

• Training:
  – For each question word (what, how, when, ...) build different SVM classifiers

• Testing:
  – For each unseen question, extract the question word
  – Build a feature vector using the same features which were used in training
  – Use the SVM classifier corresponding to the question word
What is another good measure for the evaluation of question classification?

<table>
<thead>
<tr>
<th>example</th>
<th>Cap.</th>
<th>inDicPer</th>
<th>inDicOrg</th>
<th>inDicLoc</th>
<th>NP</th>
<th>Predicted Answer</th>
<th>True Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>LOCATION</td>
<td>LOCATION</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>LOCATION</td>
<td>OTHER</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>PERSON</td>
<td>PERSON</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>ORGANIZATION</td>
<td>ORGANIZATION</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>OTHER</td>
<td>ORGANIZATION</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>OTHER</td>
<td>OTHER</td>
</tr>
</tbody>
</table>
Evaluation Measures

• Mean Reciprocal Rank (MRR) is used to evaluate a list of possible responses to a query, ordered by probability of correctness.

\[ MRR = \frac{1}{Q} \sum_{i=1}^{Q} \frac{1}{R_{AT}} \]

- \( Q \) – total number of questions to evaluate
- \( R_{AT} \) – rank at which answer type is found

<table>
<thead>
<tr>
<th>Question</th>
<th>Result of Predicted Answer Type</th>
<th>Correct Answer Type</th>
<th>Rank</th>
<th>Reciprocal Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is Nicholas Cage’s profession?</td>
<td>person, bio, contact info, definition</td>
<td>bio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>What is the tallest building in Japan?</td>
<td>entity, location, gpe, facility</td>
<td>facility</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Evaluation Measures

- Mean Reciprocal Rank (MRR) is used to evaluate a list of possible responses to a query, ordered by probability of correctness.

\[ MRR = \frac{1}{Q} \sum_{i=1}^{Q} \frac{1}{R_{AT}} \]

- \( Q \) – total number of questions to evaluate
- \( R_{AT} \) – rank at which answer type is found

<table>
<thead>
<tr>
<th>Question</th>
<th>Result of Predicted Answer Type</th>
<th>Correct Answer Type</th>
<th>Rank</th>
<th>Reciprocal Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is Nicholas Cage’s profession?</td>
<td>person, <strong>bio</strong>, contact info, definition</td>
<td><strong>bio</strong></td>
<td>2</td>
<td>1/2</td>
</tr>
<tr>
<td>What is the tallest <strong>building</strong> in Japan?</td>
<td><strong>entity</strong>, location, gpe, <strong>facility</strong></td>
<td><strong>facility</strong></td>
<td>4</td>
<td>1/4</td>
</tr>
</tbody>
</table>

MRR=0.375
Question Classification Performance

• Baseline model
  – bag-of-word features (i.e. the individual words that make up the question)

• Obtained results with described features

<table>
<thead>
<tr>
<th>Rank R</th>
<th>Baseline</th>
<th>Syntax (bigram)</th>
<th>Semantic</th>
<th>Syntax+Semantic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>77.59</td>
<td>78.81</td>
<td>80.39</td>
<td>81.25</td>
</tr>
<tr>
<td>2</td>
<td>86.42</td>
<td>86.85</td>
<td>90.45</td>
<td>88.58</td>
</tr>
<tr>
<td>3</td>
<td>89.51</td>
<td>90.30</td>
<td>93.10</td>
<td>93.32</td>
</tr>
<tr>
<td>4</td>
<td>91.16</td>
<td>91.74</td>
<td>95.55</td>
<td>94.40</td>
</tr>
<tr>
<td>5</td>
<td>92.60</td>
<td>92.82</td>
<td>97.34</td>
<td>95.33</td>
</tr>
<tr>
<td>10</td>
<td>95.33</td>
<td>95.40</td>
<td>95.40</td>
<td>97.20</td>
</tr>
<tr>
<td>MRR</td>
<td>.8437</td>
<td>.8517</td>
<td>.8727</td>
<td>.8737 (+3)</td>
</tr>
</tbody>
</table>
Error Analysis

• Inconsistency in the labeled data
  
  Who is *Duke Ellington*?
  
  Who is *Charles Lindbergh*?

  What does *CNN* stand for?
  
  What does *USPS* stand for?
Error Analysis

• Ambiguous answer types

   Where was Lincoln assassinated?

   Where was the actress, Marion Davies, buried?
Error Analysis

- Inherently difficult questions

  What is the name of the Lion King’s son in the movie “The Lion King”?

  Who developed potlatch?
Error Analysis

• POS-tagging misclassifications

What US Government agency registers trademarks?

What/WP US/NNP Government/NN agency/NN registers/NNS trademarks/NNS ?/.
Error Analysis

• WordNet coverage and ambiguity

What cereal goes “snap, crackle, pop”?

WordNet Sense 1

<table>
<thead>
<tr>
<th>cereal, cereal grass</th>
<th>grass</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>gramineous plant, graminaceous plant</td>
</tr>
<tr>
<td></td>
<td>herb, herbaceous plant</td>
</tr>
<tr>
<td></td>
<td>vascular plant</td>
</tr>
</tbody>
</table>

Type

product
Can you build your own QC system?

In what city was Peter born?
What is Peter’s birthplace?
What is the birthplace of Peter?
Name Peter’s birthplace.

What is the tallest mountain?
What is the highest mountain in the world?
Name the highest mountain.
What is the name of the tallest mountain?