Information Retrieval and Web Search

Question Answering

Instructor: Rada Mihalcea
QA on Google

Google search for "what is the capital of Italy" shows Rome as the capital of Italy.
QA on Google

Google search for "what is the population of rome"

Rome Population

- Madrid: 3.273 million
- Rome: 2.753 million
- Milan: 1.316 million

Rome
Capital of Italy

Rome is a city and special comune in Italy. Rome is the capital of Italy and region of Lazio. With 2.9 million residents in 1,285 km², it is also the country's largest and most populated comune and ... Wikipedia

Population elsewhere

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Paris</td>
<td>2.211 million (2008)</td>
</tr>
<tr>
<td>Italy</td>
<td>59.28 million (2010)</td>
</tr>
<tr>
<td>Florence</td>
<td>370,092 (2010)</td>
</tr>
</tbody>
</table>

Sources include: UNdata, World Bank
According to legend, Rome was founded in 753 BC by twin sons Romulus and Remus who were raised by a she-wolf. During its twelve-century history, the Roman civilization shifted from a monarchy to an oligarchic republic to an immense empire.
QA on Google

how is life in rome

About 205,000,000 results (0.36 seconds)

Life in Roman Times - PBS
www.pbs.org/empires/romans/empire/life.html
As with many cultures, a person's quality of life depended in many ways on their rank within the social structure. Two Romans living at the same time in the ...

50 Signs That Rome Is Really Home - Revealed Rome
www.revealedrome.com/2014/01/living-in-rome-italy.html
Jan 13, 2014 - Rome expat life. The "real" Monti. 19: You've learned that you're not just small (piccola), but really tiny (piccolissima). You're not just beautiful ...

BEING THERE: ROME | More Intelligent Life
moreintelligentlife.com/content/john.../being-there-rome
It's firmly on the tourist circuit, yet visiting is very different from living there. John Hooper has come to see it as a city of many layers, where nothing can be taken ...

Cost of Living in Rome, Italy. Prices in Rome. Updated Mar ...
www.numbeo.com > Numbeo > Cost of Living > Italy
Various interesting economical indices calculated for Rome in Italy. ... Quality Of Life - Quality of Life ... you live in Rome? We need your help: Add data for Rome!

Roman Society, Roman Life - The Roman Empire
www.roman-empire.net/society/society.html
The entire concept of Roman life seemed to center around the city, be this the city of Rome itself or any other town. The countryside was a nice place to retire to ...

Living in Rome | InterNations.org
www.internations.org/rome-expats/guide/living-in-rome-15732
Are you planning a new life in Rome as an expatmate? Explore the InterNations guide to living in Rome for info on housing, health, transportation, and schools!
Why is QA hard?
Why is QA hard?

• **Understanding** that a query means a question.

• Translating the question into a sequence of tokens useful for **retrieval**.

• Retrieve appropriate **units of text** (vanilla IR may not work!)

• Filter, rank, and/or **combine** the units into an answer.

• **Summarize** the answer.

• **Visualize** it.

• Different **types** of question need different treatments.
Factoid Questions
Factoid Questions

- Who founded Virgin Airlines?
- What is the average age of the onset of autism?
- Where is Apple Computer based?

- Courtesy: *Speech and Language Processing* by Dan Jurafsky and James Martin, Chapter 23.
Non-factoid Questions

• Who is Celia Cruz?
• What is a Hajj?
• In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?

• Courtesy: *Speech and Language Processing* by Dan Jurafsky and James Martin, Chapter 23.
Factoid vs Non-factoid QA

- Factoid QA needs information extraction, followed by ranking.
- Non-factoid QA needs filtering and summarization on top of that.
  - Query-based summarization
  - Focused summarization
- PERSON, LOCATION, ORGANIZATION and TIME questions are generally easier to handle than REASON and DESCRIPTION questions.
What role does IR play in QA?
What role does IR play in QA?

• We need to **retrieve** and **rank** units of text in response to a query.
• The units should best be **indexed** beforehand.
• **Snippets** are very helpful.
What role does IR play in QA?

- **Snippets** are very helpful.
What role does IR play in QA?

- **Snippets** are very helpful.
Question Answering from Text

• An idea originating from the IR community
• With massive collections of full-text documents, simply finding *relevant documents* is of limited use: we want *answers* from textbases
• QA: give the user a (short) answer to their question, perhaps supported by evidence.
People *Want* to Ask Questions...

Examples from AltaVista query log
who invented surf music?
how to make stink bombs
where are the snowdens of yesteryear?
which english translation of the bible is used in official catholic liturgies?
how to do clayart
how to copy psx
how tall is the sears tower?

Examples from Excite query log (12/1999)
how can i find someone in texas
where can i find information on puritan religion?
what are the 7 wonders of the world
how can i eliminate stress
What vacuum cleaner does Consumers Guide recommend

Around 12–15% of query logs
Applications

• General search for information
• Education
  – Tutoring systems
• Healthcare
• Customer service
• Travel reservations
General Factoid QA Framework

Courtesy: Prof Dan Jurafsky
The Google Answer #1

- Include question words etc. in your stop-list
- Do standard IR

- Sometimes this (sort of) works:

- Question: *Who was the prime minister of Australia during the Great Depression?*
- Answer: *James Scullin (Labor) 1929–31.*
Prime Minister during the Great Depression. Labor Prime Minister James Scullin was Australia's first Catholic Prime Minister and the first to be of Irish descent. He came to power on 22 October 1929, just a week before the stock market crash in Wall Street.

Great Depression in Australia - Wikipedia, the free encyclopedia

Australia suffered badly during the period of the Great Depression of the 1930s. Joseph Lyons, popular United Australia Party Prime Minister from 1932-1939. 1920s: The calm before the storm - 1929: The storm erupts

Great Depression in Canada - Wikipedia, the free encyclopedia

Thus, the British market played a vital role in helping Canada and Australia stabilize their economies. When the Depression began Mackenzie King was Prime Minister. During the Great Depression in Canada the demand for radical action peaked...

Great Depression | australia.gov.au

Oct 1, 2009 - Prime Ministers during the Great Depression. Labor Prime Minister James Scullin was Australia's first Catholic Prime Minister and the first to be of Irish descent. He came to power on 22 October 1929, just a week before the stock market crash in Wall Street.

The Great Depression - Ancient Australian History

The Great Depression was an economic catastrophe that caused a decline in... time... Australia had a lot of foreign debt due to infrastructure projects during the... This has been attributed to Prime Minister Lyons’ conservative and cautious...

National Museum of Australia - James Scullin

Jim Scullin was the first Catholic to become Australian Prime Minister, the first Prime Minister to have come from an Irish background, and the first Labor Prime Minister born in Australia. He became Prime Minister in 1929 - the start of the Great Depression - a devastating time for Australia and the rest of the world.
But Often It Doesn’t...

- Question: *How much money did IBM spend on advertising in 2006?*
- Answer: *I don’t know, but I’d like to ... 😞*
Lot of ads on Google these days!

No relevant info (Marketing firm page)

No relevant info (Mag page on ad exec)

No relevant info (Mag page on MS-IBM)
The Google Answer #2

• Take the question and try to find it as a string on the web
• Return the next sentence on that web page as the answer
• Works brilliantly if this exact question appears as a FAQ question, etc.
• Works lously most of the time
• But a slightly more sophisticated version of this approach has been revived in recent years with considerable success...
A Brief History

• In some sense question answering is not a new research area

• Question answering systems can be found in many areas of NLP research, including:
  • Natural language database systems
    - A lot of early NLP work on these
  • Spoken dialog systems

• The focus on open-domain QA is (relatively) new
  – MURAX (Kupiec 1993): Encyclopedia answers
  – Hirschman: Reading comprehension tests
Ask (former AskJeeves)

- **Ask.com** is probably one the most popular examples of “Question answering”
- It largely does pattern matching to match your question to their own knowledge base of questions
- If that works, you get the human-created answers to that known question
- If that fails, it falls back to regular web search
- A potentially interesting middle ground, but a fairly weak shadow of real QA
Other Online QA Examples

- EasyAsk
- MIT Start
- Apple Siri
- IBM Watson
- Wolfram Alpha
- Yahoo! Answers
- Quora
Question Answering at TREC

- Question answering competition at TREC consists of answering a set of 500 fact-based questions, e.g., “When was Mozart born?".

- For the first three years systems were allowed to return 5 ranked answer snippets (50/250 bytes) to each question.
  - IR think
  - Mean Reciprocal Rank (MRR) scoring:
    - 1, 0.5, 0.33, 0.25, 0.2, 0 for 1, 2, 3, 4, 5, 6+ doc
  - Mainly Named Entity answers (person, place, date, ...)

- More complex tasks in later years:
  - E.g., starting with 2002 the systems were only allowed to return a single exact answer and the notion of confidence has been introduced.
  - Incremental answers
  - “All the answers” in the collection
Sample TREC Questions

1. Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?
2. What was the monetary value of the Nobel Peace Prize in 1989?
3. What does the Peugeot company manufacture?
4. How much did Mercury spend on advertising in 1993?
5. What is the name of the managing director of Apricot Computer?
6. Why did David Koresh ask the FBI for a word processor?
7. What debts did Qintex group leave?
8. What is the name of the rare neurological disease with symptoms such as: involuntary movements (tics), swearing, and incoherent vocalizations (grunts, shouts, etc.)?
The TREC Document Collection

- The current collection uses news articles from the following sources:
  - AP newswire,
  - New York Times newswire,
  - Xinhua News Agency newswire,

- In total there are 1,033,461 documents in the collection. 3GB of text

- Clearly this is too much text to process entirely using advanced NLP techniques so the systems usually consist of an initial information retrieval phase followed by more advanced processing.

- Many supplement this text with use of the web, and other knowledge bases
Top Performing Systems

• The best performing systems at TREC can answer approximately 70% of the questions

• Approaches and successes have varied a fair deal
  – Knowledge-rich approaches, using a vast array of Natural Language Processing techniques got the best results in 2000, 2001
  – AskMSR system stressed how much could be achieved by very simple methods with enough text (and now various copycats)
  – Middle ground is to use large collection of surface matching patterns (ISI)
Web Question Answering: Is More Always Better?
- Dumais, Banko, Brill, Lin, Ng (Microsoft, MIT, Berkeley)

Q: “Where is the Louvre located?”

Want “Paris” or “France” or “75058 Paris Cedex 01” or a map

Don’t just want URLs
AskMSR: Shallow Approach

- In what year did Abraham Lincoln die?
- Ignore hard documents and find easy ones
AskMSR: Details

Where is the Louvre Museum located?

in Paris France 59%
museums 12%
hostels 10%

N-Best Answers

1. Rewrite Query
2. <Search Engine>
3. Collect Summaries, Mine N-grams
4. Filter N-Grams
5. Tile N-Grams
Step 1: Rewrite Queries

- Intuition: The user’s question is often syntactically quite close to sentences that contain the answer
  - Where is the Louvre Museum located?
  - The Louvre Museum is located in Paris
  - Who created the character of Scrooge?
  - Charles Dickens created the character of Scrooge.
Query Rewriting

- Classify question into seven categories
  - **Who** is/was/are/were...?
  - **When** is/did/will/are/were ...?
  - **Where** is/are/were ...?

a. Category-specific transformation rules
   eg “For Where questions, move ‘is’ to all possible locations”
   “Where is the Louvre Museum located”
   \[ \rightarrow \text{“is the Louvre Museum located”} \]
   \[ \rightarrow \text{“the is Louvre Museum located”} \]
   \[ \rightarrow \text{“the Louvre is Museum located”} \]
   \[ \rightarrow \text{“the Louvre Museum is located”} \]
   \[ \rightarrow \text{“the Louvre Museum located is”} \]

b. Expected answer “Datatype” (eg, Date, Person, Location, ...)
   **When** was the French Revolution? \[ \rightarrow \text{DATE} \]

- Hand-crafted classification/rewrite/datatype rules
  (Could they be automatically learned?)

Nonsense, but who cares? It’s only a few more queries to Google.
Query Rewriting - Weights

- One wrinkle: Some query rewrites are more reliable than others

Where is the Louvre Museum located?

**Weight 1**
Lots of non-answers could come back too

**Weight 5**
if we get a match, it’s probably right

+“the Louvre Museum is located”

+Louvre +Museum +located
Step 2: Query Search Engine

• Send all rewrites to a Web search engine
• Retrieve top N answers (100?)
• For speed, rely just on search engine’s snippets, not the full text of the actual document
Step 3: Mining N-Grams

- Unigram, bigram, trigram, ... N-gram: list of N adjacent terms in a sequence

- Eg, “Web Question Answering: Is More Always Better”
  - Unigrams: Web, Question, Answering, Is, More, Always, Better
  - Bigrams: Web Question, Question Answering, Answering Is, Is More, More Always, Always Better
Mining N-Grams

- Simple: Enumerate all N-grams (N=1,2,3 say) in all retrieved snippets
  - Use hash table to make this efficient

- Weight of an n-gram: occurrence count, each weighted by “reliability” (weight) of rewrite that fetched the document

- Example: “Who created the character of Scrooge?”
  - Dickens - 117
  - Christmas Carol - 78
  - Charles Dickens - 75
  - Disney - 72
  - Carl Banks - 54
  - A Christmas - 41
  - Christmas Carol - 45
  - Uncle - 31
Step 4: Filtering N-Grams

- Each question type is associated with one or more “data-type filters” = regular expression
  - When...
  - Where...
  - What ...
  - Who ...

- Boost score of n-grams that do match regexp
- Lower score of n-grams that don’t match regexp
Step 5: Tiling the Answers

Scores

20  Charles Dickens
15  Dickens
10  Mr Charles

N-Grams

merged, discard old n-grams

Score 45  Mr Charles Dickens

N-Grams

Repeat, until no more overlap

tile highest-scoring n-gram
Results

• Standard TREC contest test-bed:
  ~1M documents; 900 questions

• Technique doesn’t do too well (though would have placed in top 9 of ~30 participants!)
  – MRR = 0.262 (ie, right answered ranked about #4-#5)

• Using the Web as a whole, not just TREC’s 1M documents...
  MRR = 0.42 (ie, on average, right answer is ranked about #2-#3)
  – Why? Because it relies on the enormity of the Web!
ISI: Surface Patterns Approach

• ISI’s approach

• Use of Characteristic Phrases

• "When was <person> born”
  - Typical answers
    • "Mozart was born in 1756."
    • "Gandhi (1869-1948)…”
  - Suggests phrases like
    • "<NAME> was born in <BIRTHDATE>”
    • "<NAME> ( <BIRTHDATE>-”
  - as Regular Expressions can help locate correct answer
Use Pattern Learning

• Example:
  • “The great composer Mozart (1756-1791) achieved fame at a young age”
  • “Mozart (1756-1791) was a genius”
  • “The whole world would always be indebted to the great music of Mozart (1756-1791)”
- Longest matching substring for all 3 sentences is "Mozart (1756-1791)"
- Suffix tree would extract "Mozart (1756-1791)" as an output, with score of 3
Pattern Learning (cont.)

- Repeat with different examples of same question type
  - “Gandhi 1869”, “Newton 1642”, etc.

- Some patterns learned for BIRTHDATE
  - a. born in <ANSWER>, <NAME>
  - b. <NAME> was born on <ANSWER>,
  - c. <NAME> ( <ANSWER> -
  - d. <NAME> ( <ANSWER> - )
Experiments

• 6 different Q types
  – from Webclopedia QA Typology (Hovy et al., 2002a)
    • BIRTHDATE
    • LOCATION
    • INVENTOR
    • DISCOVERER
    • DEFINITION
    • WHY-FAMOUS
Experiments: Pattern Precision

• BIRTHDATE table:
  • 1.0  <NAME> ( <ANSWER> - )
  • 0.85 <NAME> was born on <ANSWER>,
  • 0.6  <NAME> was born in <ANSWER>
  • 0.59 <NAME> was born <ANSWER>
  • 0.53 <ANSWER> <NAME> was born
  • 0.50  - <NAME> ( <ANSWER>
  • 0.36  <NAME> ( <ANSWER> -

• INVENTOR
  • 1.0  <ANSWER> invents <NAME>
  • 1.0  the <NAME> was invented by <ANSWER>
  • 1.0  <ANSWER> invented the <NAME> in
Experiments

• DISCOVERER
  • 1.0 when <ANSWER> discovered <NAME>
  • 1.0 <ANSWER>'s discovery of <NAME>
  • 0.9 <NAME> was discovered by <ANSWER> in

• DEFINITION
  • 1.0 <NAME> and related <ANSWER>
  • 1.0 form of <ANSWER>, <NAME>
  • 0.94 as <NAME>, <ANSWER> and
Experiments

• WHY-FAMOUS
  • 1.0 <ANSWER> <NAME> called
  • 1.0 laureate <ANSWER> <NAME>
  • 0.71 <NAME> is the <ANSWER> of

• LOCATION
  • 1.0 <ANSWER>'s <NAME>
  • 1.0 regional : <ANSWER> : <NAME>
  • 0.92 near <NAME> in <ANSWER>

• Depending on question type, get high MRR (0.6–0.9), with higher results from use of Web than TREC QA collection
Shortcomings & Extensions

• Need for POS &/or semantic types
  • "Where are the Rocky Mountains?"
  • "Denver's new airport, topped with white fiberglass cones in imitation of the Rocky Mountains in the background, continues to lie empty"
  • <NAME> in <ANSWER>

• Named Entity Tagger &/or ontology could enable system to determine "background" is not a location
Shortcomings...

• Long distance dependencies
  • "Where is London?"
  • "London, which has one of the most busiest airports in the world, lies on the banks of the river Thames”
  • would require pattern like:
    <QUESTION>, (<any_word>)*, lies on <ANSWER>
  – Abundance & variety of Web data helps system to find an instance of patterns w/o losing answers to long distance dependencies
Shortcomings...

• System currently has only one anchor word
  – Doesn't work for Q types requiring multiple words from question to be in answer
    • "In which county does the city of Long Beach lie?"
    • "Long Beach is situated in Los Angeles County"
  • required pattern:
    <Q_TERM_1> is situated in <ANSWER> <Q_TERM_2>
References

• AskMSR: Question Answering Using the Worldwide Web
  – Michele Banko, Eric Brill, Susan Dumais, Jimmy Lin

• Web Question Answering: Is More Always Better?
  – Susan Dumais, Michele Banko, Eric Brill, Jimmy Lin, Andrew Ng