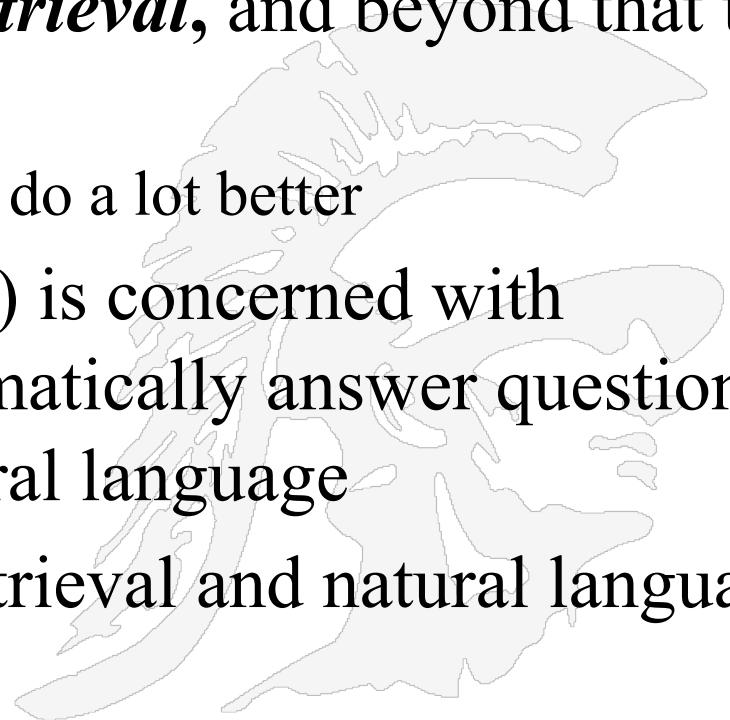


Search Engine Question Answering



Information Retrieval v. Question Answering

- The name “**information retrieval**” is standard, but as traditionally practiced, it’s not really right
- All you get is *document retrieval*, and beyond that the job is up to you
 - Modern search engines now do a lot better
- **Question Answering (QA)** is concerned with building systems that automatically answer questions posed by humans in a natural language
- It combines information retrieval and natural language processing



People *want* to ask questions...

Examples from Ask.com query log

how much should I weigh

what does my name mean

how to get pregnant

where can I find pictures of hairstyles

who is the richest man in the world

what is the meaning of life

why is the sky blue

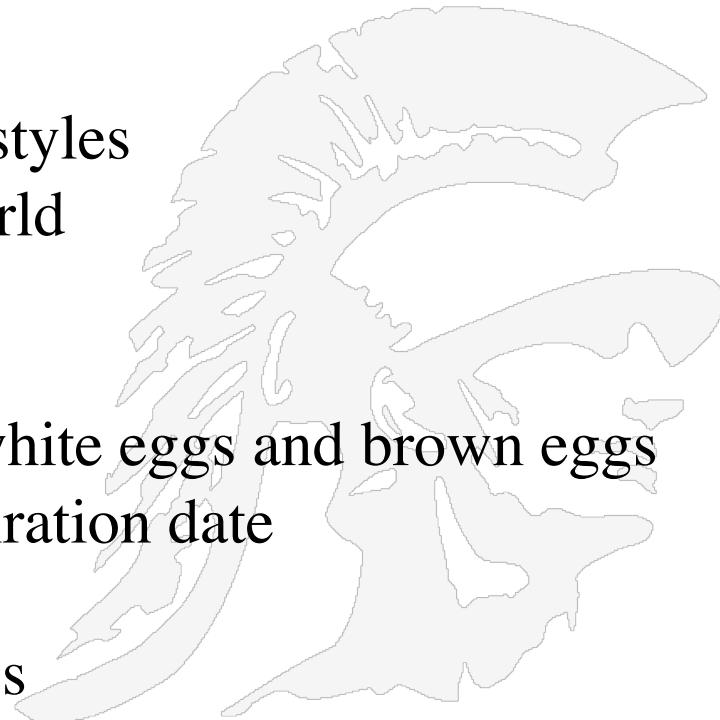
what is the difference between white eggs and brown eggs

can you drink milk after the expiration date

what is true love

what is the jonas brothers address

Around 10-20% of query logs are questions such as these



Question: *Who was the prime minister of Australia during the Great Depression?*
 Answer: *James Scullin (Labor) 1929–31*

Google Search: Who was the prime minister of Australia during the Great Depression? - Microsoft Internet Explorer

Address: +of+Australia+during+the+Great+Depression%3F&btnG=Google+Search

Google Search Site News

Advanced Search Preferences Language Tools Search Tips

Who was the prime minister of Australia Google Search

The following words are very common and were not included in your search: Who was the of the. [details]

Web Images Groups Directory News

Searched the web for **Who was the prime minister of Australia during the Great Depression?**. Results 1 - 1

Asking a question? Try out [Google Answers](#).

How Google used to respond to questions

From Poor Boy to Prime Minister

... how did he come to lead **Australia during** World War ... April 1939 Menz takes over as **Prime Minister** after the death of Lyons; Sept 3 1939 **Australia** declares war ... john.curtin.edu.au/manofpeace/boytopm.html - 23k - Mar 1, 2003 - [Cached](#) - [Similar pages](#)

Activity: Banning of the Communist Party in World War II

... The **Great Depression** had brought enormous suffering to workers ... by the 'Prime Minister' and His ... the Communist Party in **Australia during** john.curtin.edu.au/letters/activities/communism.html - 8k - [Cached](#) - [Similar pages](#) [More results from john.curtin.edu.au]

Prime Ministers of Australia - Chifley

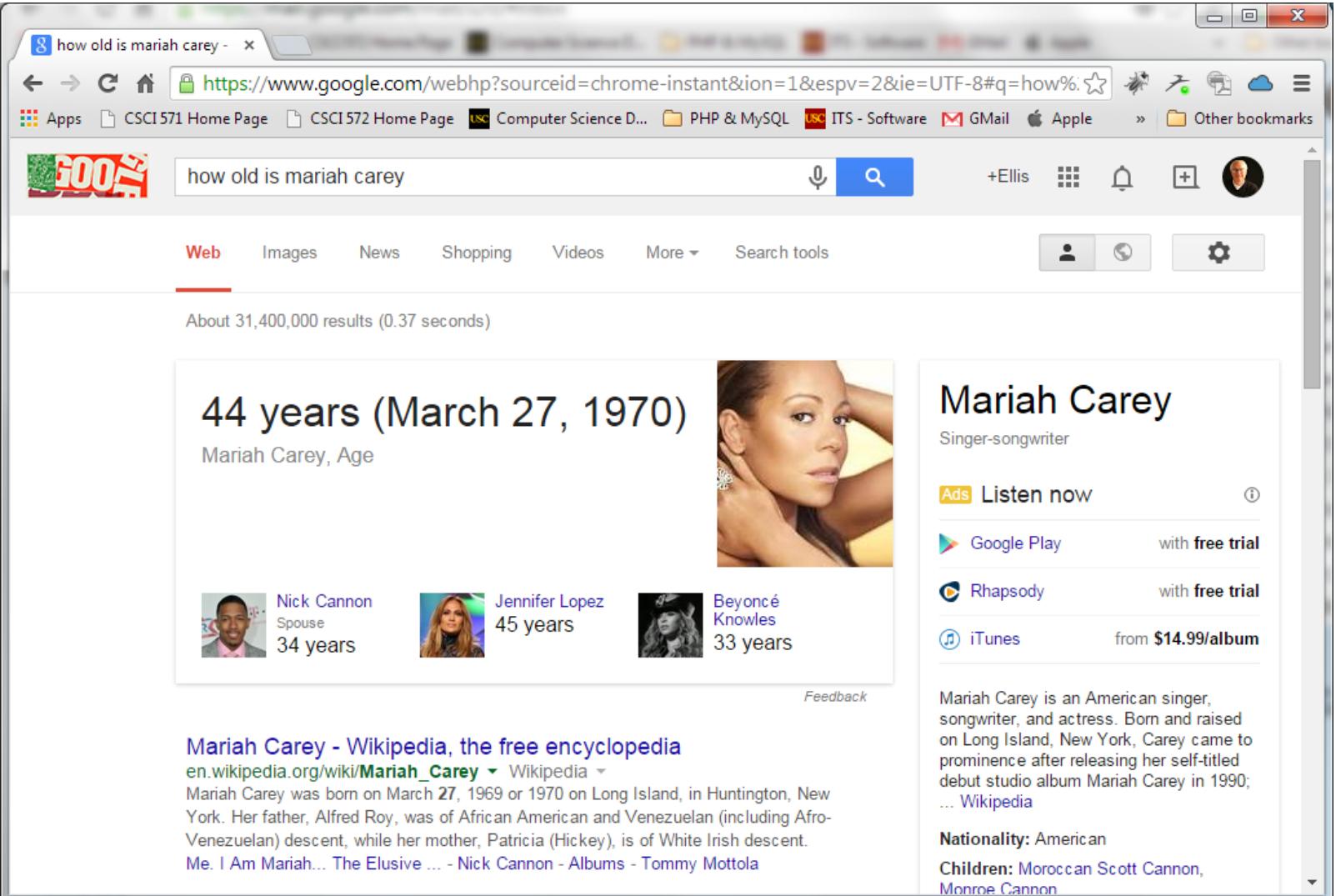
... defying the federal United **Australia** Party government ... Second World War until led du the 1930s. ... He became **Prime Minister** following Curtin's death, succeeding www.nma.gov.au/primeministers/3.htm - 30k - [Cached](#) - [Similar pages](#)

Page about Curtin (WW II Labor Prime Minister)
 (Can deduce answer)

Page about Curtin (WW II Labor Prime Minister)
 (Lacks answer)

Page about Chifley (Labor Prime Minister)
 (Can deduce answer)

Google has Improved Its Ability to Answer Some Questions



how old is mariah carey

About 31,400,000 results (0.37 seconds)

44 years (March 27, 1970)

Mariah Carey, Age

Nick Cannon
Spouse
34 years

Jennifer Lopez
45 years

Beyoncé Knowles
33 years

Mariah Carey

Singer-songwriter

Ads Listen now

Google Play with free trial

Rhapsody with free trial

iTunes from \$14.99/album

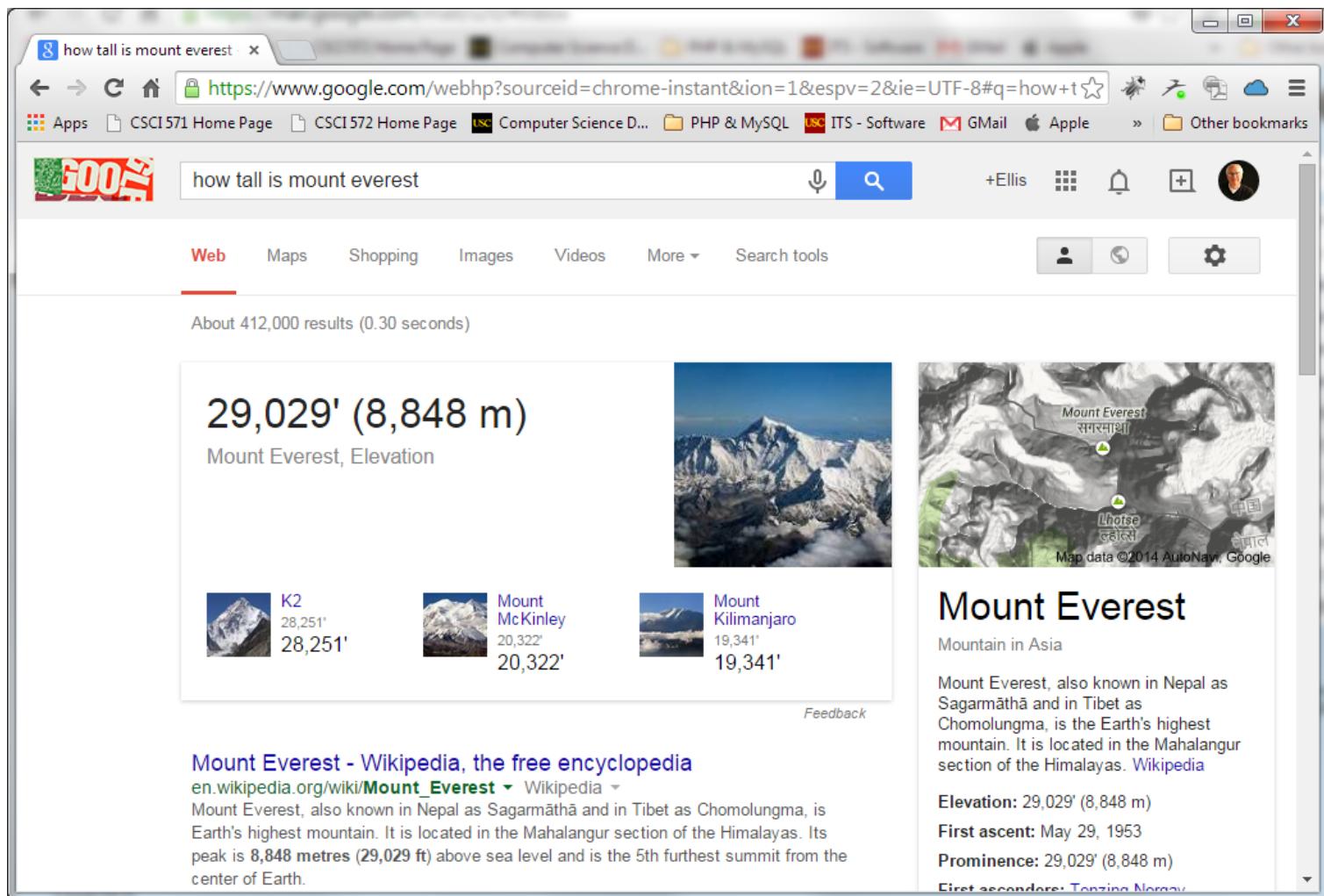
Mariah Carey is an American singer, songwriter, and actress. Born and raised on Long Island, New York, Carey came to prominence after releasing her self-titled debut studio album *Mariah Carey* in 1990; ... Wikipedia

Nationality: American

Children: Moroccan Scott Cannon, Monroe Cannon

Copyright EINN HOROWITZ 2011-2018

Some Questions are Easily Answered



Google search results for "how tall is mount everest".

Search query: how tall is mount everest

Results:

- 29,029' (8,848 m)**
Mount Everest, Elevation
-  **K2**
28,251'
28,251'
-  **Mount McKinley**
20,322'
20,322'
-  **Mount Kilimanjaro**
19,341'
19,341'

Feedback

Mount Everest - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Mount_Everest ▾ Wikipedia ▾

Mount Everest, also known in Nepal as Sagarmāthā and in Tibet as Chomolungma, is Earth's highest mountain. It is located in the Mahalangur section of the Himalayas. Its peak is 8,848 metres (29,029 ft) above sea level and is the 5th furthest summit from the center of Earth.


Map data ©2014 AutoNavi, Google

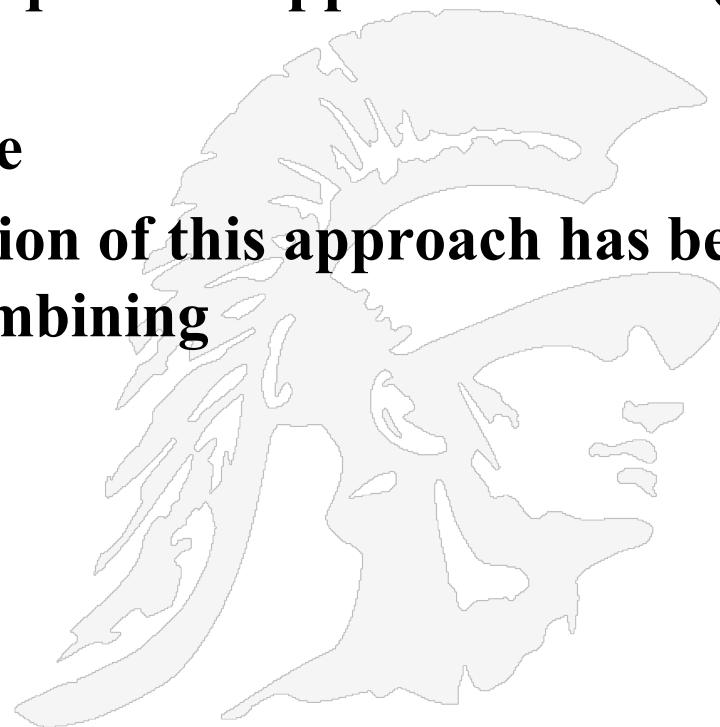
Mount Everest
Mountain in Asia

Mount Everest, also known in Nepal as Sagarmāthā and in Tibet as Chomolungma, is the Earth's highest mountain. It is located in the Mahalangur section of the Himalayas. [Wikipedia](#)

Elevation: 29,029' (8,848 m)
First ascent: May 29, 1953
Prominence: 29,029' (8,848 m)
First ascender: Tenzing Norgay

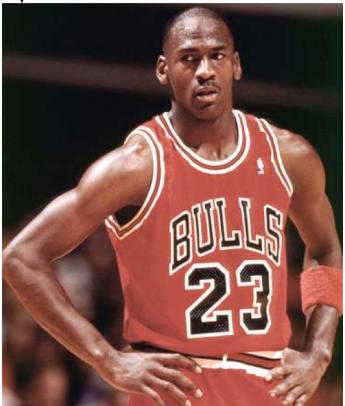
The Original Google Approach

- 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 poorly most of the time
- But a more sophisticated version of this approach has been introduced in recent years combining
 - Knowledge graph
 - N-grams



Many Questions Pose Semantic Difficulties

- Who is Michael Jordan?
 - Michael Jordan the basketball player or the Machine Learning guy?
- Key requirement is that entities get identified and disambiguated



who is michael jordan - Google Search

<https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF8>

who is michael jordan

Web News Images Videos Shopping More Search tools

About 324,000,000 results (0.39 seconds)

Michael Jordan - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Michael_Jordan Wikipedia
 Michael Jeffrey Jordan (born February 17, 1963), also known by his initials, MJ, is an American former professional basketball player, entrepreneur, and principal owner and chairman of the Charlotte Hornets.
 1984 NBA draft - Jeffrey Jordan - Air Jordan - Marcus Jordan

Michael Jordan - Biography - Basketball Player - Biography ...
www.biography.com/people/michael-jordan-9358066 FYI
 Follow the career of former basketball star Michael Jordan, from his college career to being the Chicago Bulls' MVP, to his multiple retirements, ...

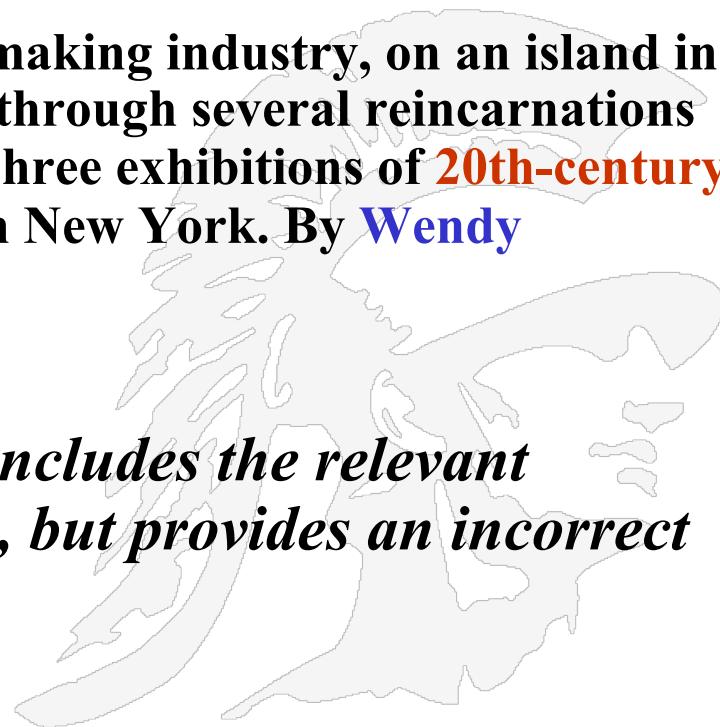
In the news

Is Leon Hall the Michael Jordan of the NFL?
[Cincinnati.com](https://cincinnati.com/article/2013/02/01/is-leon-hall-the-michael-jordan-of-the-nfl) - 1 day ago
 You've heard the perhaps-exaggerated story that Michael Jordan was cut from his high ...

Aries Spears -- I Hate Kobe Bryant ... He's No Michael Jordan

Why Natural Language Processing is Required

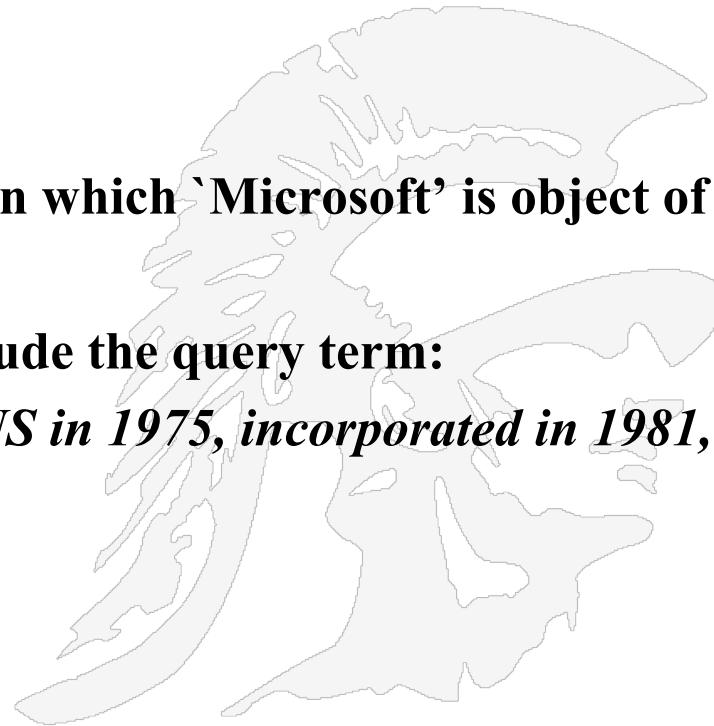
- **Question:** “When was Wendy’s founded?”
- **Passage candidate:**
 - “The renowned Murano glassmaking industry, on an island in the Venetian lagoon, has gone through several reincarnations since it was **founded** in 1291. Three exhibitions of **20th-century** Murano glass are coming up in New York. By **Wendy Moonan**.”
- **Answer:** **20th Century**
- *the candidate passage below includes the relevant keywords (Wendy's, founded), but provides an incorrect answer*



More NLP Challenges

Predicate-Argument Structure

- Q336: *When was Microsoft established?*
- Difficult because Microsoft tends to establish lots of things...
Microsoft plans to establish manufacturing partnerships in Brazil and Mexico in May.
- Need to be able to detect sentences in which 'Microsoft' is object of 'establish' or close synonym.
- A correct result might *not* even include the query term:
Microsoft Corp was founded in the US in 1975, incorporated in 1981, and established in the UK in 1982.

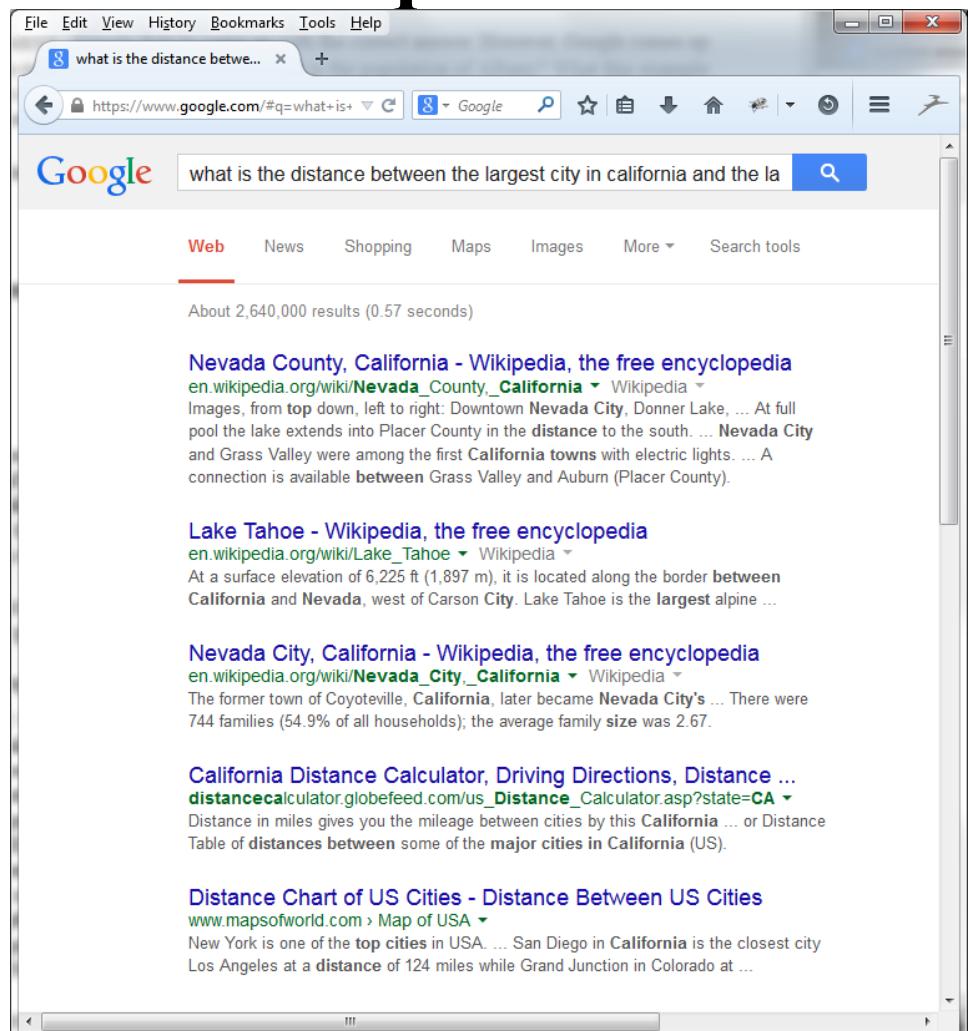


What is the distance between the largest city in California and the largest city in Nevada?

Google does poorly on this query, misinterpreting Nevada as Nevada County, California

ps: Google may have improved its results

Some Questions Require Inferences



File Edit View History Bookmarks Tools Help

what is the distance betwe... https://www.google.com/#q=what+is+ ↻ Google

what is the distance between the largest city in california and the la

Google Web News Shopping Maps Images More Search tools

About 2,640,000 results (0.57 seconds)

Nevada County, California - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Nevada_County,_California Wikipedia
Images, from top down, left to right: Downtown Nevada City, Donner Lake, ... At full pool the lake extends into Placer County in the distance to the south. ... Nevada City and Grass Valley were among the first California towns with electric lights. ... A connection is available between Grass Valley and Auburn (Placer County).

Lake Tahoe - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Lake_Tahoe Wikipedia
At a surface elevation of 6,225 ft (1,897 m), it is located along the border between California and Nevada, west of Carson City. Lake Tahoe is the largest alpine ...

Nevada City, California - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Nevada_City,_California Wikipedia
The former town of Coyoteville, California, later became Nevada City's ... There were 744 families (54.9% of all households); the average family size was 2.67.

California Distance Calculator, Driving Directions, Distance ...
distancecalculator.globefeed.com/us_Distance_Calculator.asp?state=CA
Distance in miles gives you the mileage between cities by this California ... or Distance Table of distances between some of the major cities in California (US).

Distance Chart of US Cities - Distance Between US Cities
[www.mapsofworld.com > Map of USA](https://www.mapsofworld.com/map-of-usa)
New York is one of the top cities in USA. ... San Diego in California is the closest city Los Angeles at a distance of 124 miles while Grand Junction in Colorado at ...

how many Ph.D. degrees in mathematics were granted by European universities in 1986?

All results are irrelevant



a more recent result; still no relevant links



Google search results for "how many Ph.D. degrees in mathematics were granted by European universities in 1986?"

- [Fuzzy Logic and the Semantic Web - Page 172 - Google Book...](https://books.google.com/books?id=0080460488)
- [Doctor of Philosophy - Wikipedia, the free encyclopedia](https://en.wikipedia.org/wiki/Doctor_of_Philosophy)
- [Doctor of Science - Wikipedia, the free encyclopedia](https://en.wikipedia.org/wiki/Doctor_of_Science)

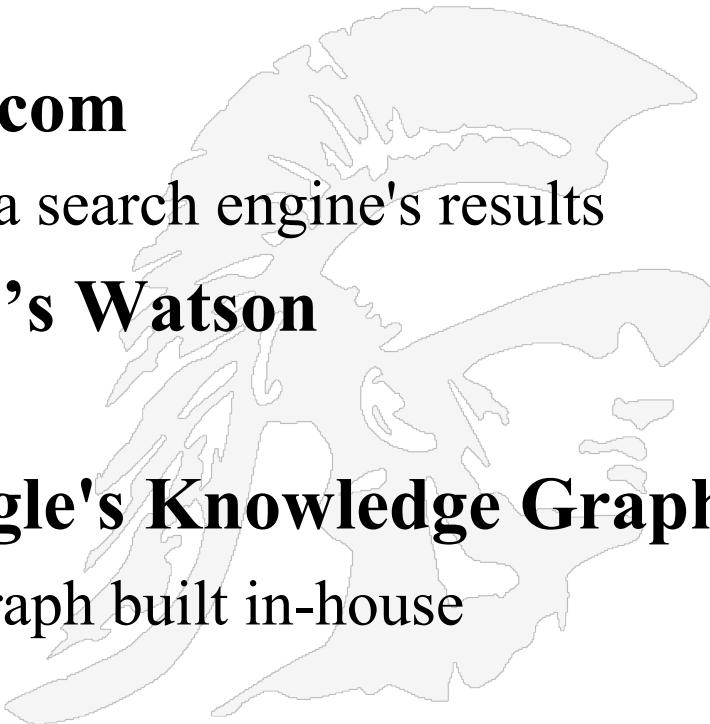
In Some Cases the Data May Not Exist

Bing search results for "how many Ph.D. degrees in mathematics were granted by European un..."

- [Doctorate Degrees: Are They Worth It? - Online Phd Programs](http://www.online-phd-programs.org/doctorate-degrees)
- [Doctor of Philosophy - Wikipedia, the free encyclopedia](https://en.wikipedia.org/wiki/PhD)
- [More women, minorities seek, earn Ph.D.s, according to ...](http://chronicle.uchicago.edu/00427/phds.shtml)
- [Doctorate - Wikipedia, the free encyclopedia](https://en.wikipedia.org/wiki/Doctorate)
- [Chapter 2: Higher Education in Science and Engineering ...](http://www.nsf.gov/statistics/seind02/c2/c2s3.htm)

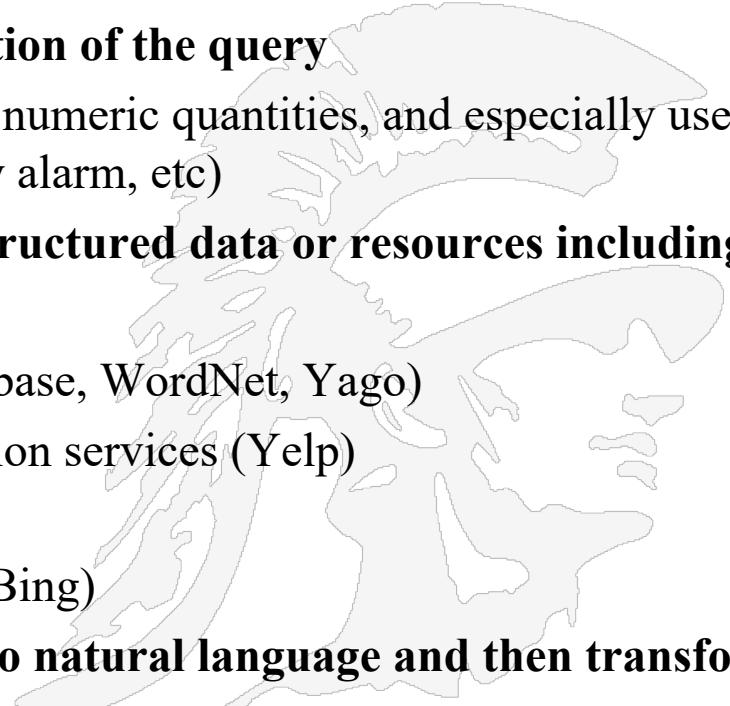
Some Popular Products Designed for Question/Answering

- **Approach 1: used by Siri**
 - map to known entities and use existing databases over the internet
- **Approach 2: used by Ask.com**
 - detect question type and use a search engine's results
- **Approach 3: used by IBM's Watson**
 - combine approaches 1 and 2
- **Approach 4: used by Google's Knowledge Graph**
 - use an entity - relationship graph built in-house



Approach used by Siri: Knowledge-Based Approach

- Siri was begun as a DARPA project called CALO/PAL (Personalized Assistant that Learns)
 1. First your voice query is put through a recognizer and a language model and Siri comes up with an interpretation of what was said
 2. Second Siri builds a semantic representation of the query
 - Extract times, dates, locations, entities, numeric quantities, and especially user actions (e.g. schedule a meeting, set my alarm, etc)
 3. Siri maps from this semantics to query structured data or resources including:
 - Geospatial databases
 - Ontologies (Wikipedia infoboxes, Freebase, WordNet, Yago)
 - Restaurant review sources and reservation services (Yelp)
 - Scientific databases (Wolfram Alpha)
 - Conventional search engines (Google, Bing)
 4. Siri then transforms the output above into natural language and then transforms the text back to speech



AskJeeves (now Ask.com)

- Earlier AskJeeves.com was well-known as a search engine specializing in Questions/Answers
- Though it still exists, it performs far weaker than sites such as Google

Ask how old is mariah carey

Q How old is Mariah Carey

[Mariah Carey Caution Tour - Tickets On Sale for All Dates](#)
Buy 2019 Mariah Carey Tickets at Vivid Seats! Order Online Today.
100% Buyer Guarantee - Unmatched Service - Instant Download - Email Delivery
Types: Instant Download Tickets, e-Mail Delivery Tickets, Interactive Seat Maps, Last Minute Tickets
★★★★★ rating for vivids seats.com

[Mariah Carey - 11 Things You Didn't Know](#)
We all remember Mariah Carey's lip sync fail, but did you know these 11 things?
Shows: Real Housewives, Dancing With the Stars, Teen Mom, Gilmore Girls, Grey's Anatomy
11 Bey And Jay Scandals
10 Facts About Shakira

[Mariah carey biography - Amazon Official Site](#)
Browse & Discover Thousands of Book Titles, for Less.
Read Ratings & Reviews - Explore Amazon Devices - Try Prime for Free - Shop Our Huge Selection
Gift cards Shop Kindle E-Readers Give the gift of Prime - Amazon Exclusives - Textbooks

Video results for how old is mariah carey

 Mariah Carey Transformation | From 1 to 47 Years Old!

 Happy Birthday, Mariah Carey! Watch the 46-Year-Old's Most Divalicious Moments

 Mariah Carey Vs. Ariana Grande SAME AGE Vocal Battle! (Live)

[More videos >](#)

[Mariah Carey - Wikipedia](#)
Mariah Carey (born March 27, 1969 or 1970) is an American singer and songwriter. Referred to By the age of four, Carey recalled that she had begun to sneak the radio under her covers at night, and just sing and try to find peace within the ...

How old is Mariah Carey
Snapshot taken 11/2018

Copyright Ellis Horowitz 2011-2018

Ask how tall is mount everest

Q How tall is Mount Everest

[Journey to Everest Base Camp - Hike mountain trails in Nepal](#)
Our exclusive eco-campsites offer added comfort & stunning views of the Himalaya
Trekking, hiking, cycling - Himalaya, Polar, Africa - The paths less travelled - Travel expert since 1975
Brands: Quality Small Group Trips, Exclusive Eco Campsites, Empowering Local People
Trekking & Hiking Trips - Expert local guide & crew

[Mount Everest - Wikipedia](#)
Mount Everest, known in Nepal as Sagarmatha (सगरमाथा) and in Tibetan as Chomolungma ... In 2010, an agreement was reached by both sides that the height of Everest is 8,848 m, and Nepal recognizes China's claim that the rock height ...

[How Tall Is Mount Everest? For Nepal, It's a Touchy Question. - The ...](#)
www.nytimes.com/2018/02/03/world/asia/mount-everest-how-tall-nepal.html
Feb 3, 2018 ... KATHMANDU, Nepal — Mount Everest is the tallest mountain in the world, but precisely how tall is it? It's not such a simple question. In the past ...

[3 Mountains Taller Than Everest | Mental Floss](#)
Jan 29, 2017 ... by James Hunt. At almost 30,000 feet, Mount Everest may seem like it's as tall and as high a mountain as has ever existed. But the title of Tallest ...

Q How Tall is Mount Everest in Meters?

[How Tall is Mount Everest? - National Geographic Education Blog](#)
Feb 6, 2018 ... Mountains that are part of high-elevation ranges such as the ... Most surveyors put Mount Everest's elevation at 8,850 meters (29,029) feet.

How tall is mount Everest
Snapshot taken 11/2018

3 Main Phases for Question/Answering

1. QUESTION PROCESSING

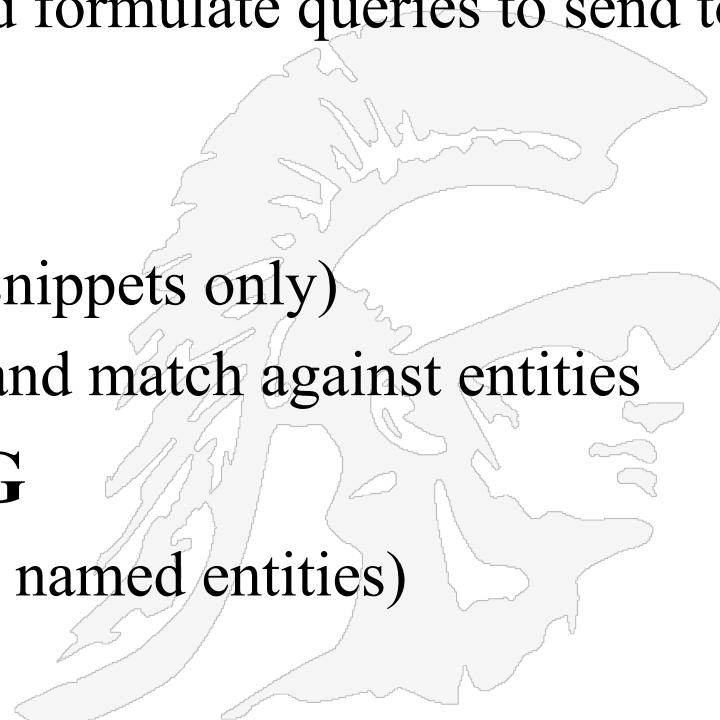
- Detect question type (who, what, when, where, etc)
- Identify important entities and formulate queries to send to a search engine

2. PASSAGE RETRIEVAL

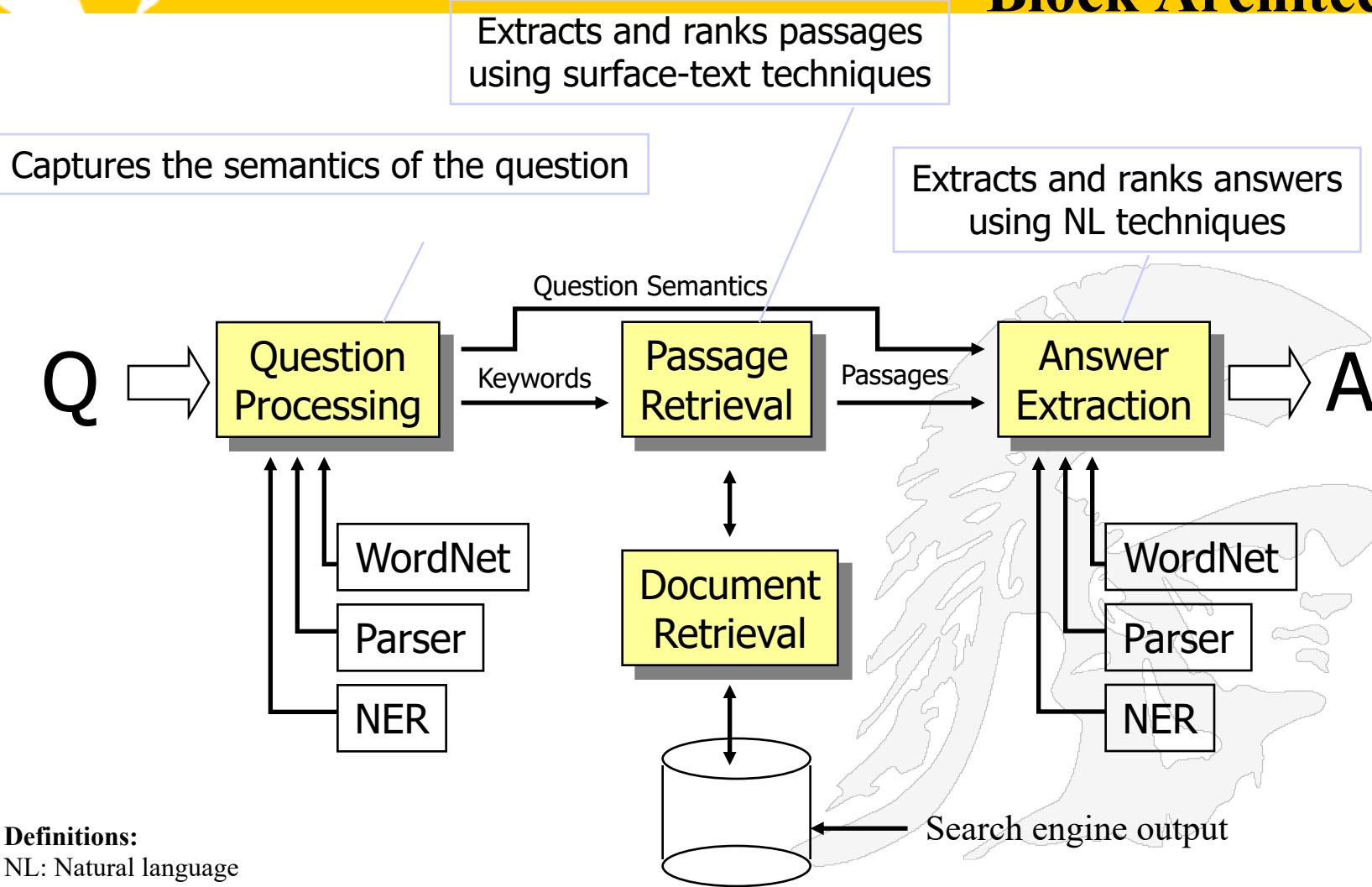
- Retrieve ranked documents (snippets only)
- Break into suitable passages and match against entities

3. ANSWER PROCESSING

- Extract candidate answers (as named entities)
- Rank candidates
 - using evidence from relations in the text and external sources



QA 3 Phase Block Architecture



Question Types: Many Questions Fall into Distinct Categories

Who	Person, Organization
When	Date, Year
Where	Location
In What	Location
How many	Number

However

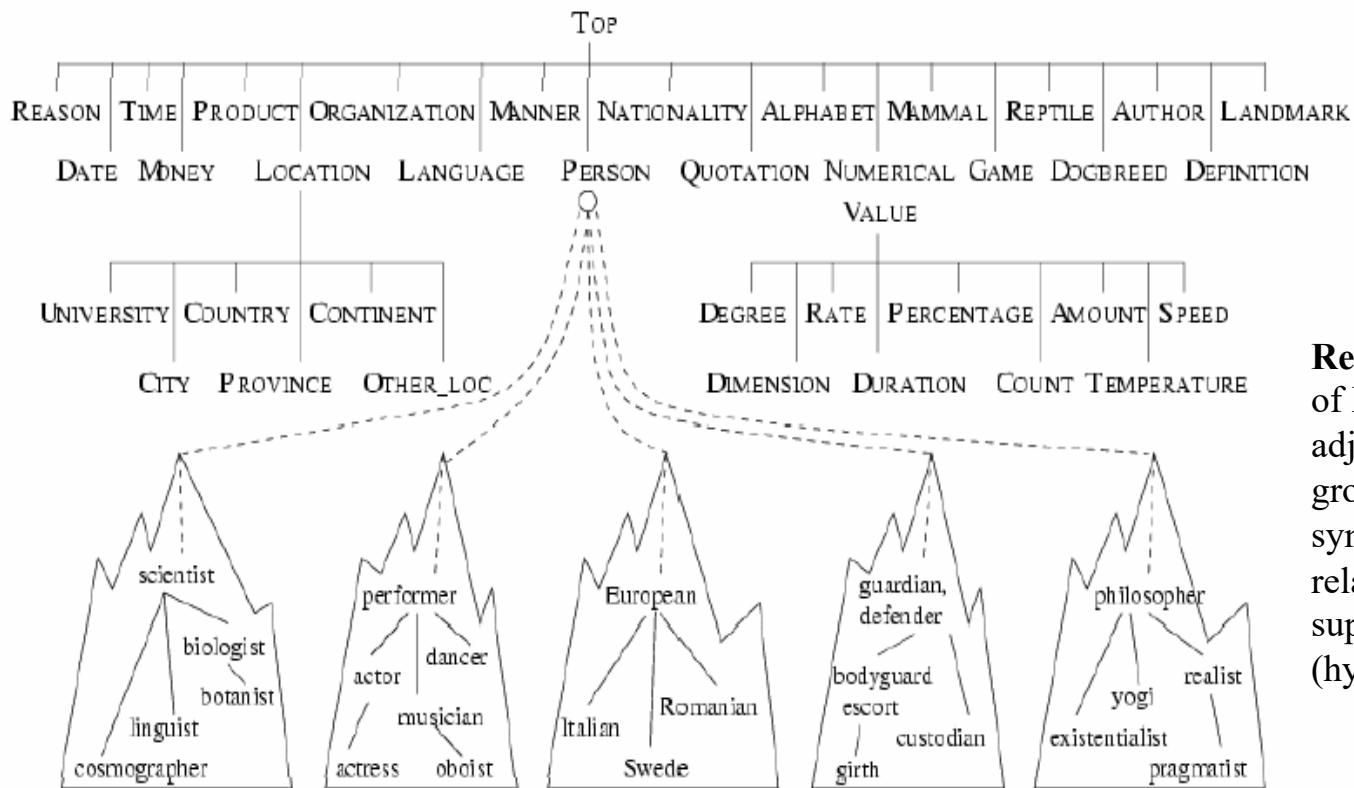
Question Types Can be Very Large

- Factoid Questions**
 - Who, where, when, how many
 - The answers fall into a limited and somewhat predictable set of categories
- This set is from a paper by Li and Roth, 2005, *Learning Question Classifiers*
- Their major categories include
 - Abbreviation
 - Description
 - Entity
 - Human
 - Location
 - Numeric
- Questions can be labeled at one or two levels, e.g. either as Numeric or as Numeric:date or as Location or Location:mountain
- Conclusion:** using the template approach it is difficult to determine which template applies

Tag	Example
ABBREVIATION	
abb	What's the abbreviation for limited partnership?
exp	What does the "c" stand for in the equation E=mc2?
DESCRIPTION	
definition	What are tannins?
description	What are the words to the Canadian National anthem?
manner	How can you get rust stains out of clothing?
reason	What caused the Titanic to sink?
ENTITY	
animal	What are the names of Odin's ravens?
body	What part of your body contains the corpus callosum?
color	What colors make up a rainbow ?
creative	In what book can I find the story of Aladdin?
currency	What currency is used in China?
disease/medicine	What does Salk vaccine prevent?
event	What war involved the battle of Chapultepec?
food	What kind of nuts are used in marzipan?
instrument	What instrument does Max Roach play?
lang	What's the official language of Algeria?
letter	What letter appears on the cold-water tap in Spain?
other	What is the name of King Arthur's sword?
plant	What are some fragrant white climbing roses?
product	What is the fastest computer?
religion	What religion has the most members?
sport	What was the name of the ball game played by the Mayans?
substance	What fuel do airplanes use?
symbol	What is the chemical symbol for nitrogen?
technique	What is the best way to remove wallpaper?
term	How do you say " Grandma " in Irish?
vehicle	What was the name of Captain Bligh's ship?
word	What's the singular of dice?
HUMAN	
description	Who was Confucius?
group	What are the major companies that are part of Dow Jones?
ind	Who was the first Russian astronaut to do a spacewalk?
title	What was Queen Victoria's title regarding India?
LOCATION	
city	What's the oldest capital city in the Americas?
country	What country borders the most others?
mountain	What is the highest peak in Africa?
other	What river runs through Liverpool?
state	What states do not have state income tax?
NUMERIC	
code	What is the telephone number for the University of Colorado?
count	About how many soldiers died in World War II?
date	What is the date of Boxing Day?
distance	How long was Mao's 1930s Long March?
money	How much did a McDonald's hamburger cost in 1963?
order	Where does Shanghai rank among world cities in population?
other	What is the population of Mexico?
period	What was the average life expectancy during the Stone Age?
percent	What fraction of a beaver's life is spent swimming?
temp	How hot should the oven be when making Peachy Oat Muffins?
speed	How fast must a spacecraft travel to escape Earth's gravity?
size	What is the size of Argentina?
weight	How many pounds are there in a stone?

Answer Type Taxonomy

- Contains ~9000 concepts reflecting expected answer types
- Merges named entities with the WordNet hierarchy

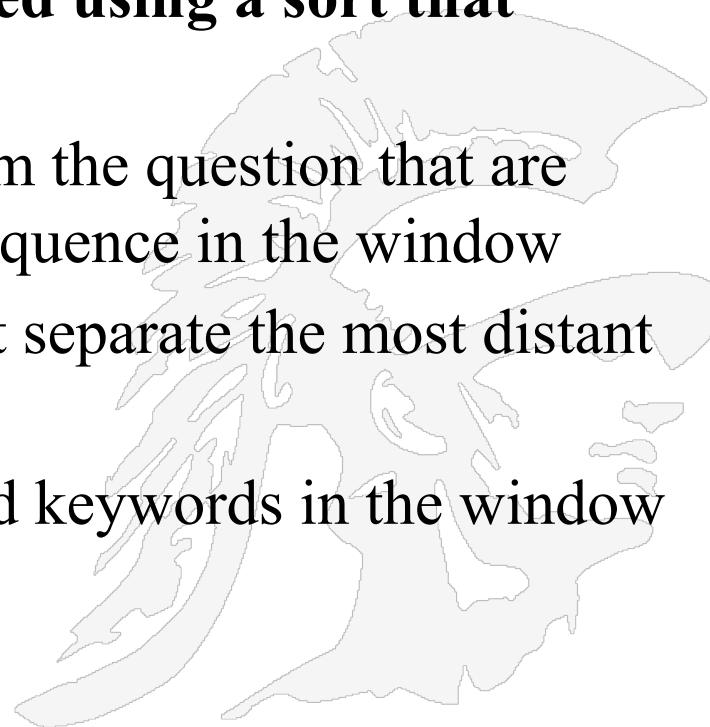


Recall: WordNet is a database of English nouns, verbs, adjectives and adverbs grouped into sets of synonyms (synsets) and relations showing super/subordinate relations (hyperonymy/hyponymy)

If you know the answer should be a person
 WordNet helps determine what sort of person

Passage Scoring

- Once snippets are returned, the answers must be extracted; sorting is involved
- Passage ordering is performed using a sort that involves three scores:
 1. The number of words from the question that are recognized in the same sequence in the window
 2. The number of words that separate the most distant keywords in the window
 3. The number of unmatched keywords in the window



Ranking Candidate Answers

Q066: Name the first private citizen to fly in space.

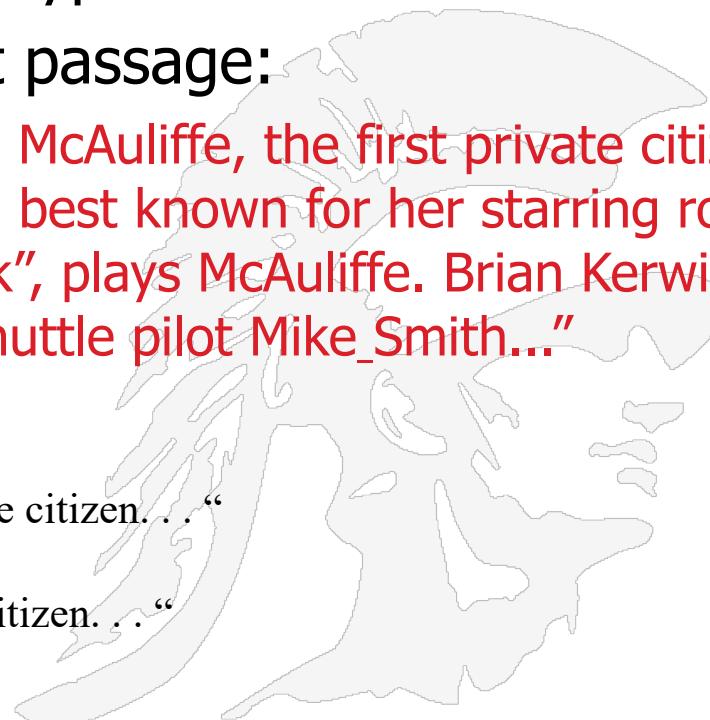
- Answer type: Person
- Text passage:

“Among them was Christa McAuliffe, the first private citizen to fly in space. Karen Allen, best known for her starring role in “Raiders of the Lost Ark”, plays McAuliffe. Brian Kerwin is featured as shuttle pilot Mike_Smith...”

There are five words from the questions “the first private citizen. . . ”

The answer is adjacent to “the first private citizen. . . ”

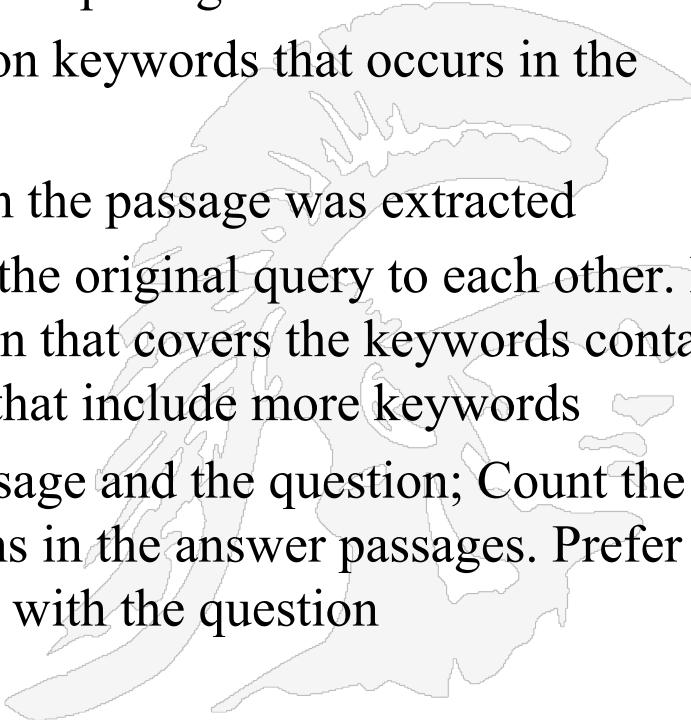
There are no unmatched keywords in “the first private citizen. . . ”



Phase 3

Ranking Retrieved Passages

- **Using supervised machine learning the candidate passages are ranked according to six criteria**
 1. The number of named entities of the right type in the passage
 2. The number of question keywords in the passage
 3. The longest exact sequence of question keywords that occurs in the passage
 4. The rank of the document from which the passage was extracted
 5. The proximity of the keywords from the original query to each other. For each passage identify the shortest span that covers the keywords contained in that passage. Prefer smaller spans that include more keywords
 6. The N-gram overlap between the passage and the question; Count the N-grams in the question and the N-grams in the answer passages. Prefer the passages with higher N-gram overlap with the question



Ranking Candidate Answers

Q066: Name the first private citizen to fly in space.

- Answer type: Person
- Text passage:

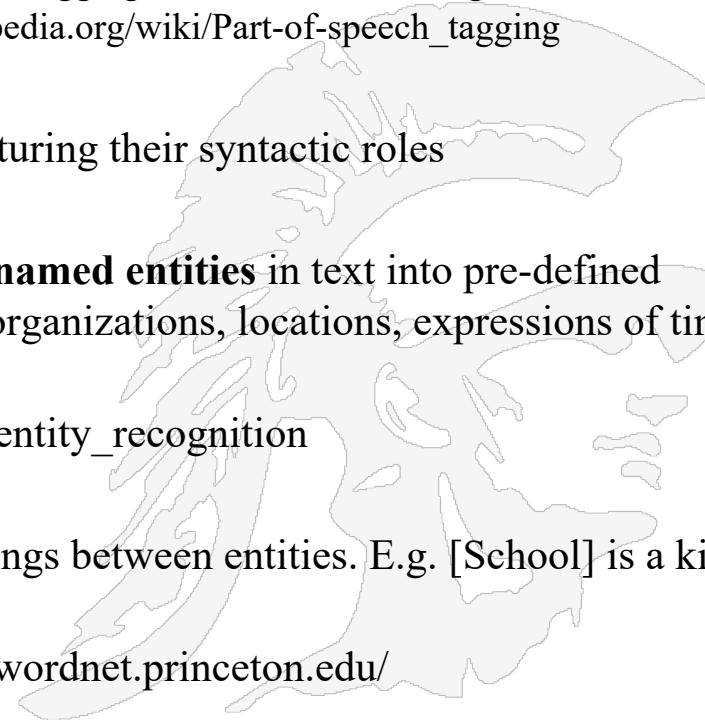
“Among them was Christa McAuliffe, the first private citizen to fly in space. Karen Allen, best known for her starring role in “Raiders of the Lost Ark”, plays McAuliffe. Brian Kerwin is featured as shuttle pilot Mike_Smith...”

- Best candidate answer: Christa McAuliffe

Karen Allen is rejected as an answer as it does not occur in the sentence “the first private citizen. . . ”
Brian Kerwin is rejected as the name is far away from “the first private citizen. . . ”

Some General Approaches

- **Part-of-Speech Tagging**
 - a piece of software that reads text in some language and assigns **parts of speech** to each word, such as noun, verb, adjective, etc.
 - Markov Models are now the standard method for part-of-speech assignment
 - Some current major algorithms for part-of-speech tagging include the Viterbi algorithm, Brill tagger, and Baum-Welch algorithm, see https://en.wikipedia.org/wiki/Part-of-speech_tagging
- **Parsing**
 - analyzing a sentence into its parts and capturing their syntactic roles
- **Named Entity Extraction**
 - Software that seeks to locate and classify **named entities** in text into pre-defined categories such as the **names** of persons, organizations, locations, expressions of times, quantities ...
 - See https://en.wikipedia.org/wiki/Named-entity_recognition
- **Determining Semantic Relations**
 - **semantic relations** are concepts or meanings between entities. E.g. [School] is a kind of [educational institution]
 - Opportunity to use WordNet, <https://wordnet.princeton.edu/>
- **Dictionaries**



Example of a Part-of-Speech Question parse

WP:Wh-pronoun

VBD:Verb, past tense

DT:Determiner

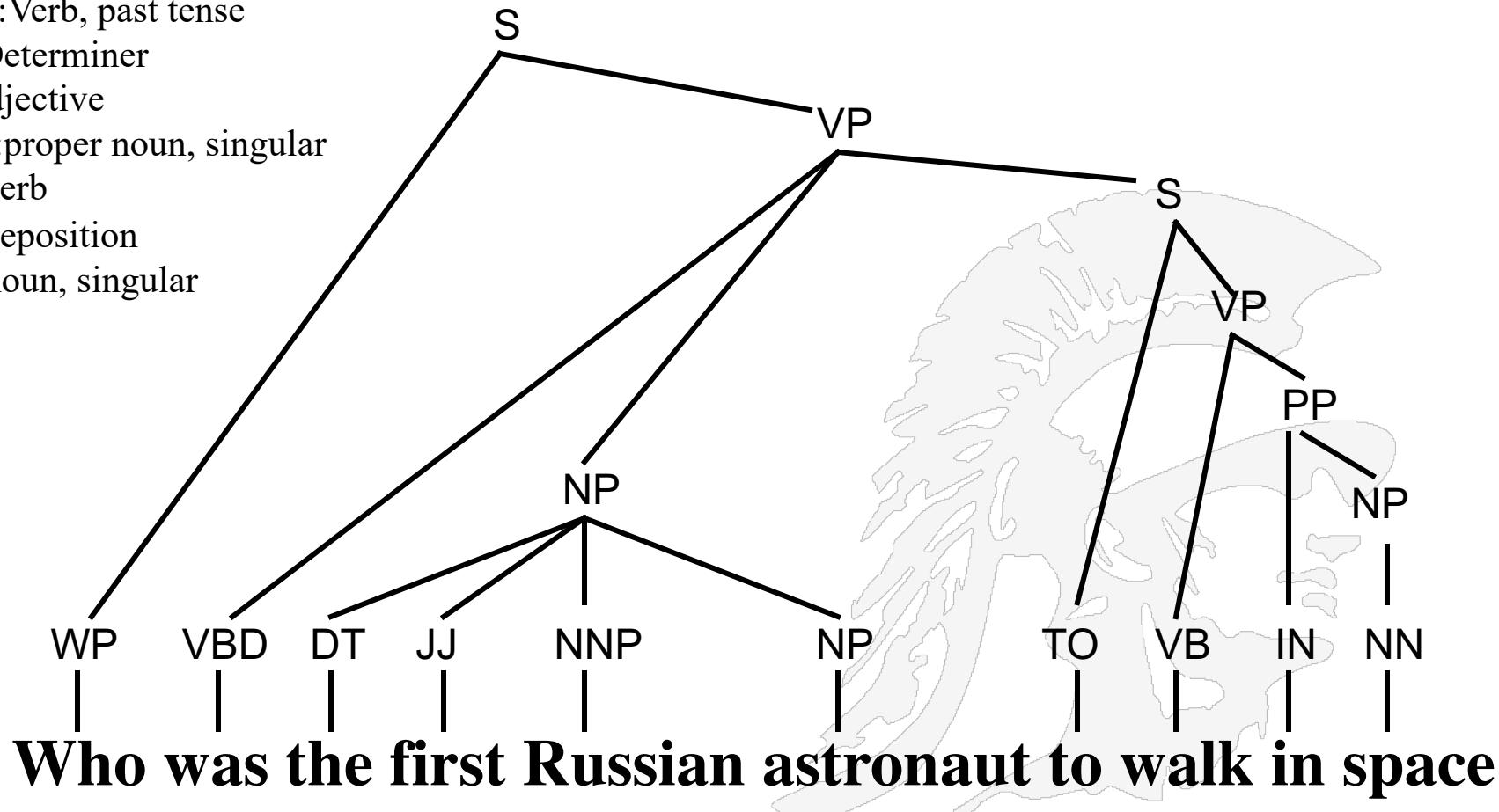
JJ:Adjective

NNP:proper noun, singular

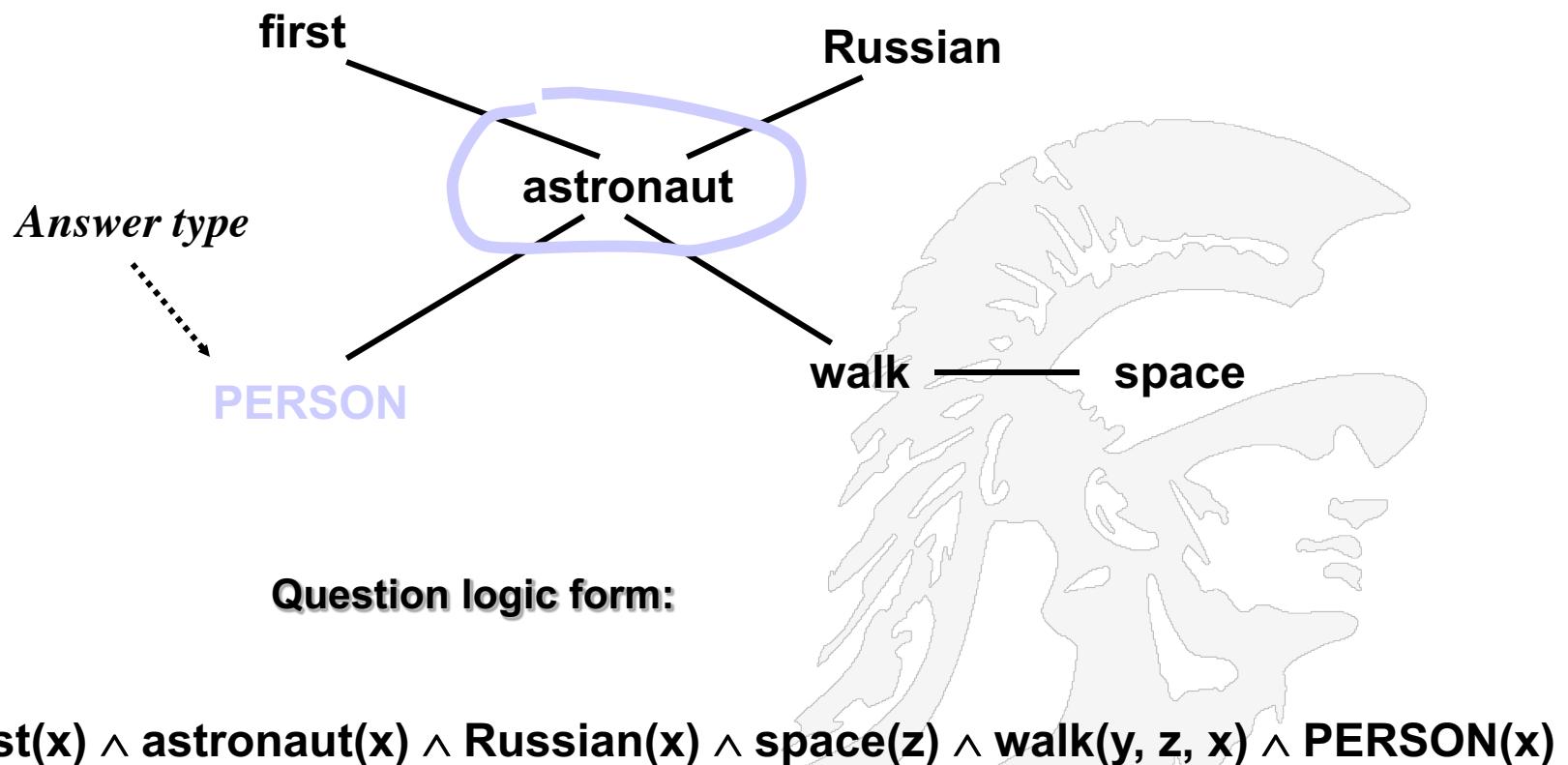
VB:verb

IN:preposition

NN:noun, singular

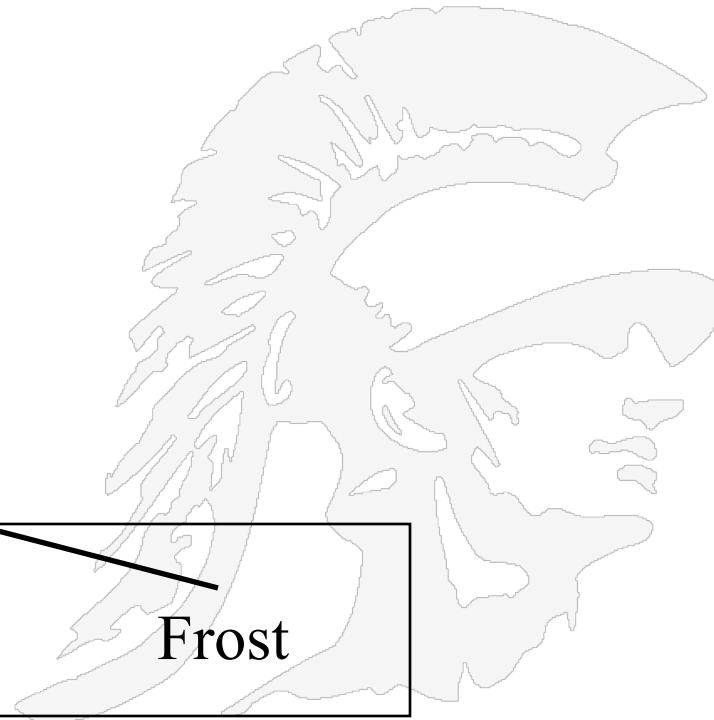
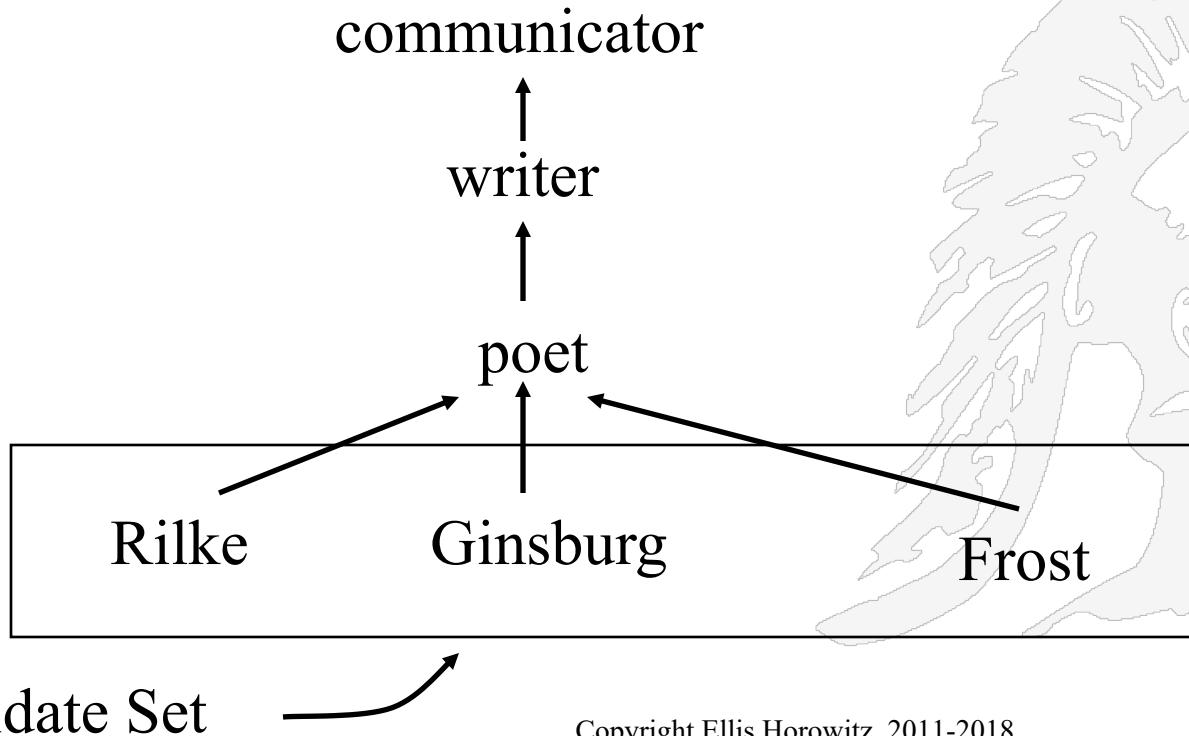


A Question's Semantic Form



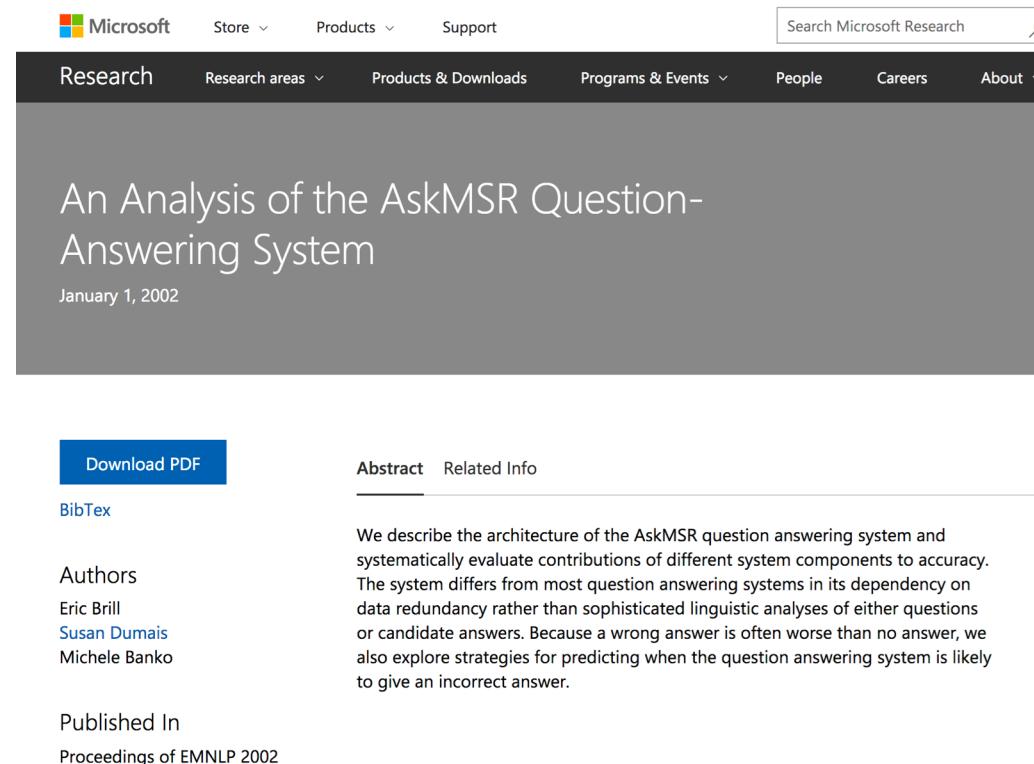
“What 20th century poet wrote Howl?”

WordNet permits refinement
of poet to specific instances



- **AskMSR is a question answering system developed at Microsoft**
- **Rather than doing sophisticated linguistic analyses it relies upon information scattered around the web**
- **AskMSR system stressed how much could be achieved by very simple methods**

http://research.microsoft.com/en-us/um/people/sdumais/EMNLP_Final.pdf



The screenshot shows a Microsoft Research page for a paper titled "An Analysis of the AskMSR Question-Answering System" by Eric Brill, Susan Dumais, and Michele Banko, published in January 1, 2002. The page includes links for Download PDF, BibTeX, Authors, Abstract, Related Info, and Published In.

An Analysis of the AskMSR Question-Answering System

January 1, 2002

Download PDF

BibTeX

Authors

Eric Brill
Susan Dumais
Michele Banko

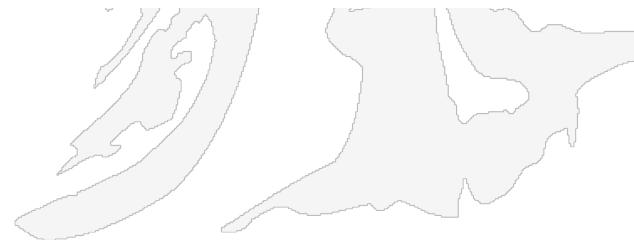
Abstract

We describe the architecture of the AskMSR question answering system and systematically evaluate contributions of different system components to accuracy. The system differs from most question answering systems in its dependency on data redundancy rather than sophisticated linguistic analyses of either questions or candidate answers. Because a wrong answer is often worse than no answer, we also explore strategies for predicting when the question answering system is likely to give an incorrect answer.

Related Info

Published In

Proceedings of EMNLP 2002



AskMSR: Shallow Approach

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

Abraham Lincoln, 1809-1865

*LINCOLN, ABRAHAM was born near Hodgenville, Kentucky, on February 12, 1809. In 1816, the Lincoln family moved to Pigeon Creek in Perry (now Spencer) County. Two years later, Abraham Lincoln's mother died and his father married a woman known as his "angel" mother. Lincoln attended a formal school for only a few months but acquired knowledge through the reading of books. He moved to Illinois, in 1830 where he obtained a job as a store clerk and the local postmaster. He served without distinction in the Black Hawk War. Lincoln lost his attempt at the state legislature, but two years later he tried again, was successful, and Lincoln was admitted to the bar and became noteworthy as a witty, honest, competent circuit court judge. He served one year term in the U.S. House in 1846, at which time he opposed the war with Mexico. By 1859, Lincoln had gained national attention for his series of debates with Stephen A. Douglas. Although he lost the election he became a significant figure in his party.



Sixteenth President
1861-1865
Married to Mary Todd Lincoln

of his inauguration on March 4, seven southern states had seceded from the Union, and the Civil War began. The South had 23 states and the North had 15. The South had superior artillery. Lincoln called for 75,000 volunteers (approximately 100,000 men) to serve for three years. The South had seceded, for a total of 11. Lincoln immediately took action to keep the Union together. He issued the Emancipation Proclamation which expanded the purpose of the war. On November 19, 1863, Lincoln gave the Gettysburg Address during the dedication of a national cemetery in Gettysburg, Pennsylvania. The address is considered one of the greatest speeches in history. It begins with the words, "Four score and seven years ago our fathers brought forth on this continent a new nation, conceived in liberty and dedicated to the proposition that all men are created equal."

War emblem
General T
ience at Fo

Abraham Lincoln

16th President of the United States (March 4, 1861 to April 15, 1865)

Born: February 12, 1809, in Hardin County, Kentucky

Died: April 15, 1865, at Petersen's Boarding House in Washington, D.C.

"I was born February 12, 1809, in Hardin County, Kentucky. My parents were both born in Virginia, of undistinguished families, perhaps I should say. My mother, who died in my tenth year, was of a family of the name of Hanks.

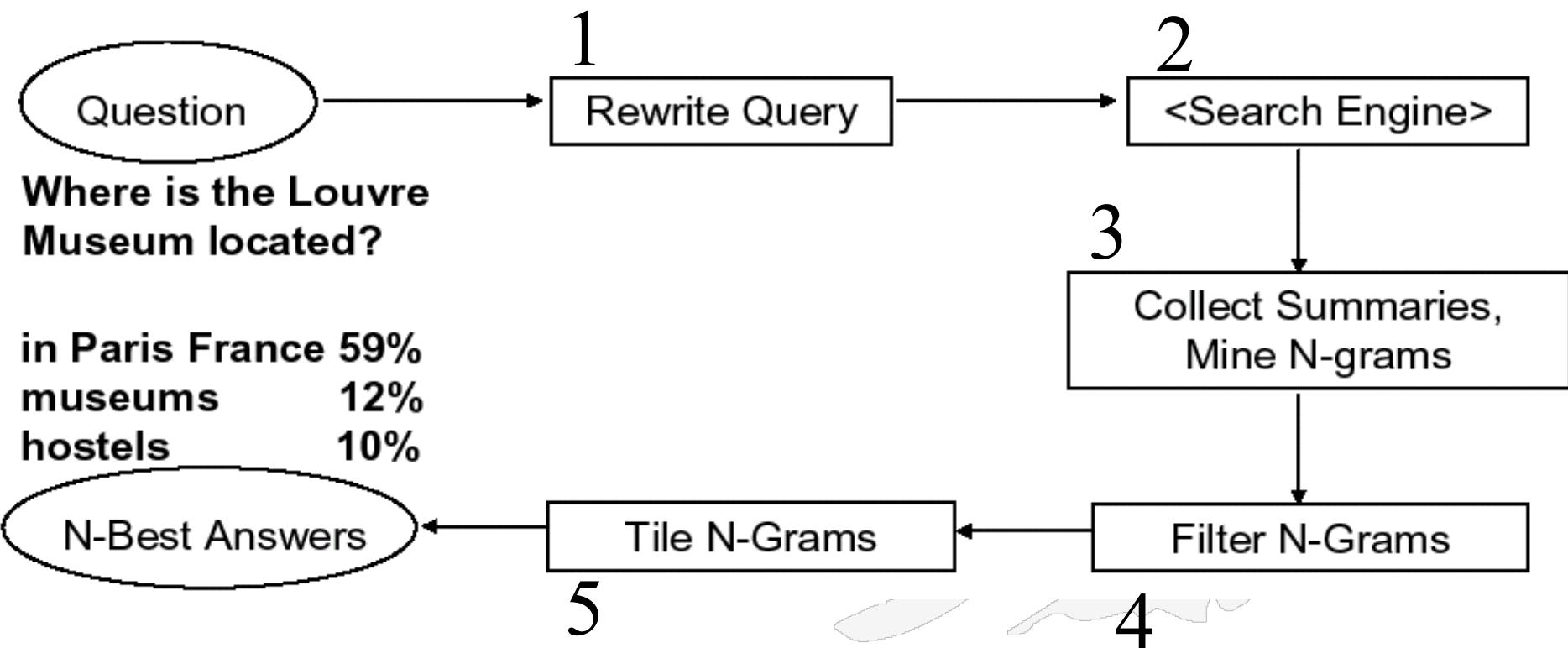


ABRAHAM LINCOLN

Sixteenth President of the United States

Born in 1809 - Died in 1865

AskMSR: Details



Step 1:Query Rewriting

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

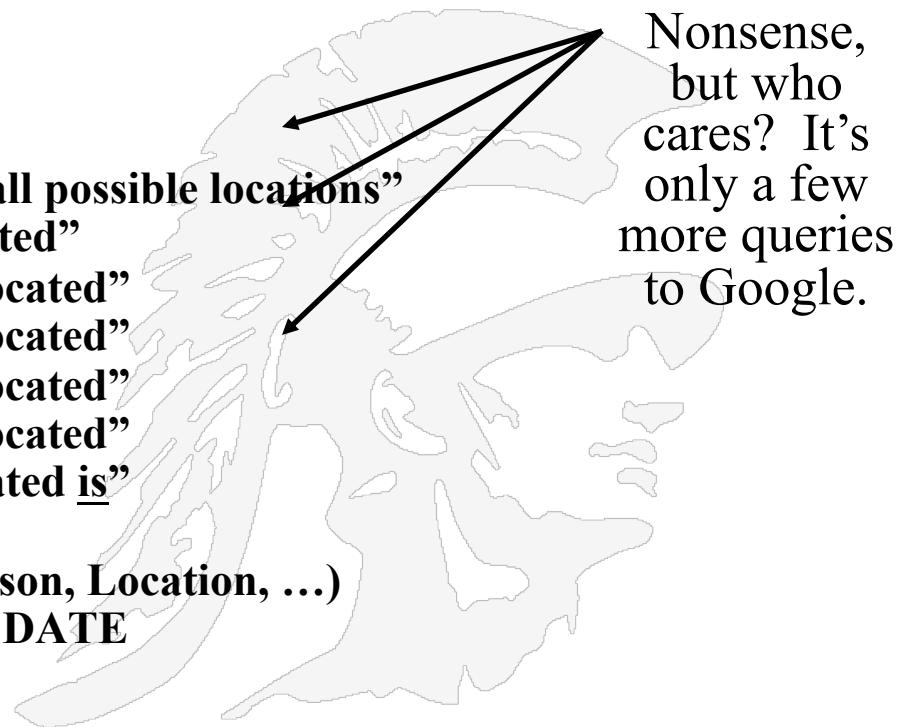
a. Category-specific transformation rules

eg “For Where questions, move ‘is’ to all possible locations”
“Where is the Louvre Museum located”

- “is the Louvre Museum located”
- “the is Louvre Museum located”
- “the Louvre is Museum located”
- “the Louvre Museum is located”
- “the Louvre Museum located is”

b. Expected answer “Datatype” (eg, Date, Person, Location, ...)

When was the French Revolution? → DATE



Query Rewriting - Weights

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

Where is the Louvre Museum located?

Weight 1

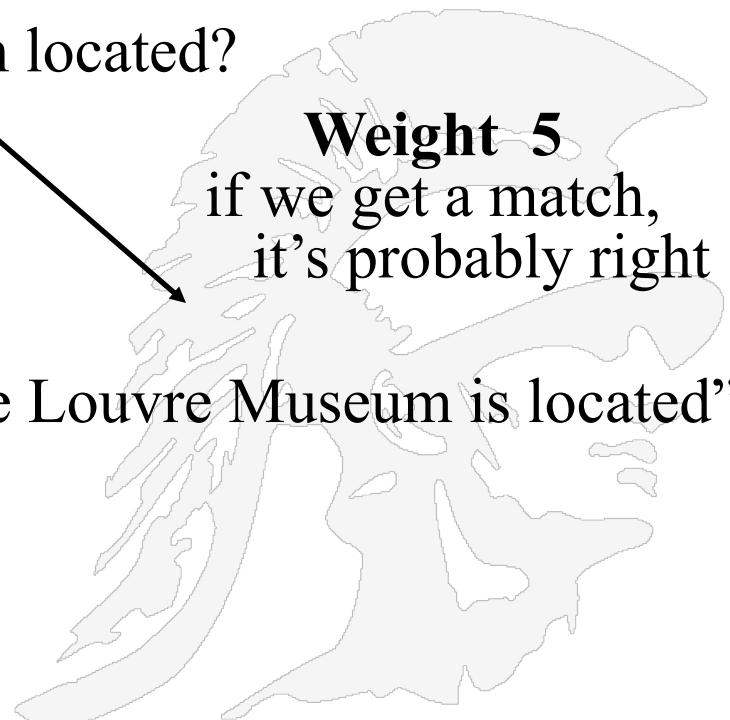
Lots of non-answers
could come back too

+Louvre +Museum +located

Weight 5

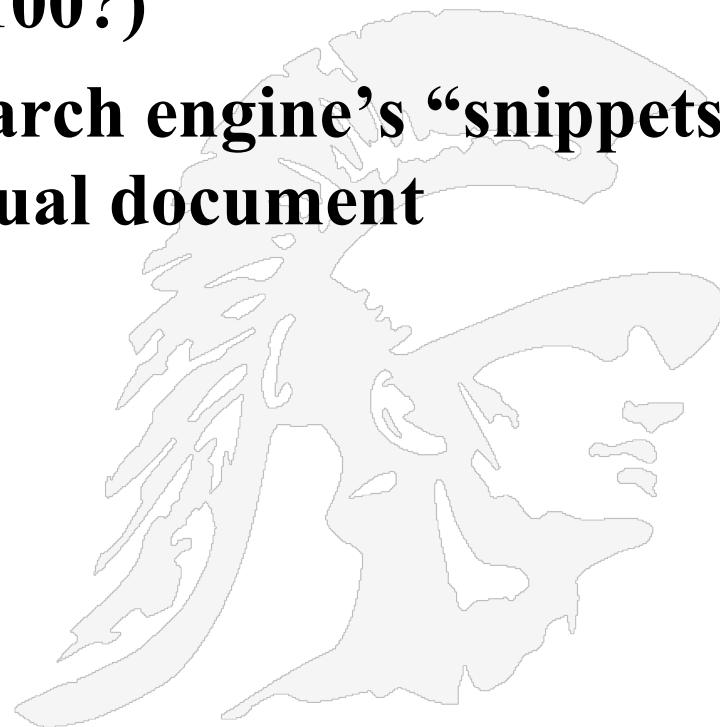
if we get a match,
it's probably right

+“the Louvre Museum is located”



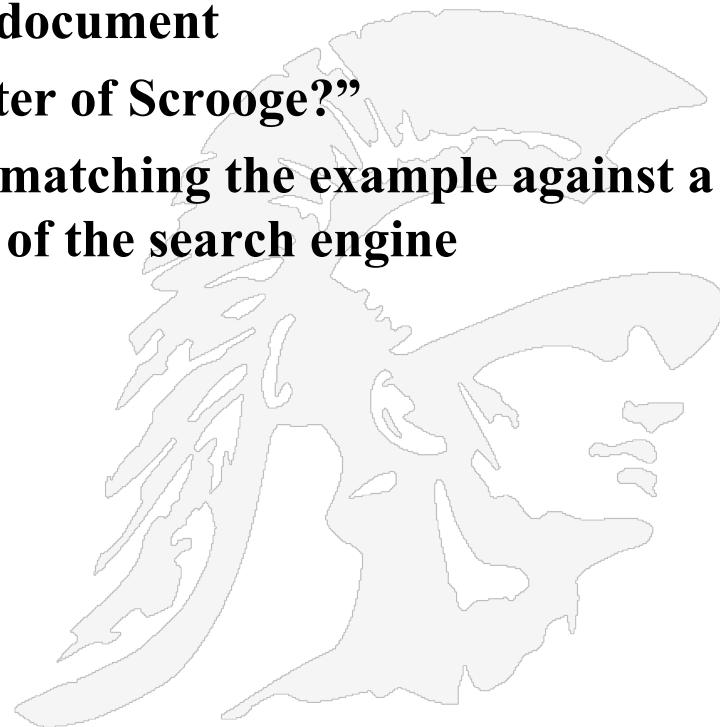
Step 2: Query Search Engine

- 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



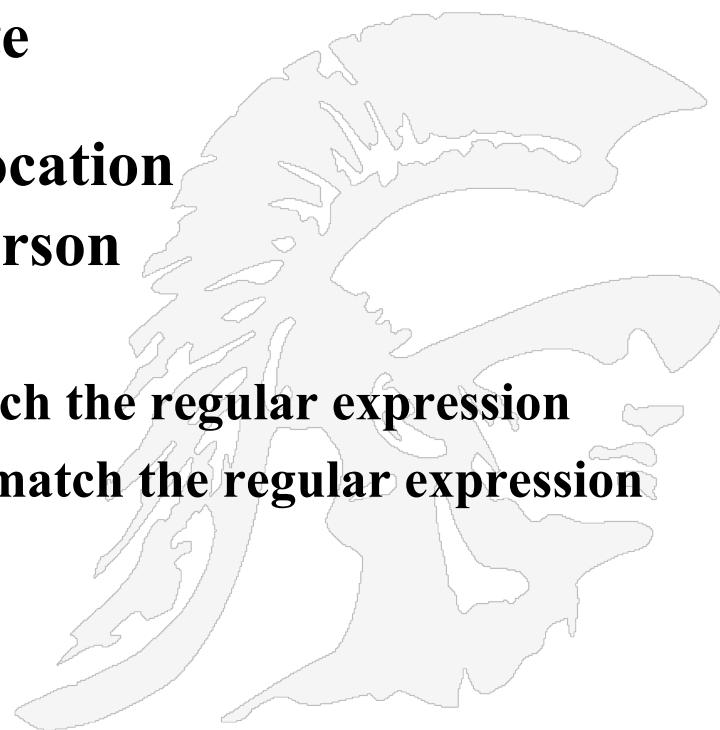
Step 3: Mining N-Grams

- **Simple:** Enumerate all N-grams ($N=1,2,3$ say) in all retrieved snippets
 - Use hash table and other data structures 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?”
- **Below are the weights produced by matching the example against a set of N-grams in the N-gram database of the search engine**
 - Dickens - 117
 - Christmas Carol - 78
 - Charles Dickens - 75
 - Disney - 72
 - Carl Banks - 54
 - A Christmas - 41
 - Christmas Carol - 45
 - Uncle - 31



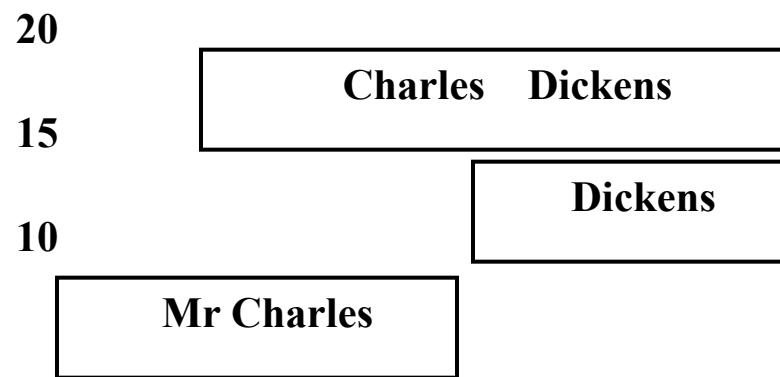
Step 4: Filtering N-Grams

- Each question type is associated with one or more “data-type filters” = regular expression
- When... → Date
- Where... → Location
- What ... → Person
- Who ... → Date
- Boost score of n-grams that do match the regular expression
- Lower score of n-grams that don’t match the regular expression



Step 5: Tiling the Answers

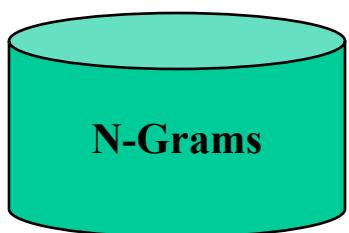
Scores



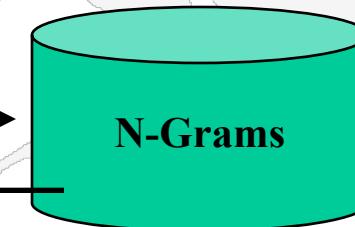
merged, discard
old n-grams

Score 45

Mr Charles Dickens



tile highest-scoring n-gram



Repeat, until no more overlap

Judging the Quality of Answers: Mean Reciprocal Rank (MRR) Scoring

- The mean reciprocal rank is a statistical measure for evaluating any process that produces a list of possible responses to a sample of queries, ordered by probability of correctness.
The **reciprocal rank** of a query response is the multiplicative inverse of the **rank** of the first correct answer
- The mean reciprocal rank is the average of the reciprocal ranks of results for a sample of queries
- For example, suppose we have the following three sample queries for a system that tries to translate English words to their plurals. In each case, the system makes three guesses, with the first one being the one it thinks is most likely correct:

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}.$$

Query	Results	Correct response	Rank	Reciprocal rank
cat	catten, cati, cats	cats	3	1/3
tori	torii, tori , toruses	tori	2	1/2
virus	viruses , virii, viri	viruses	1	1

the mean reciprocal rank as $(1/3 + 1/2 + 1)/3 = 11/18$ or about 0.61.