

←

1/37

→

\*\*\*

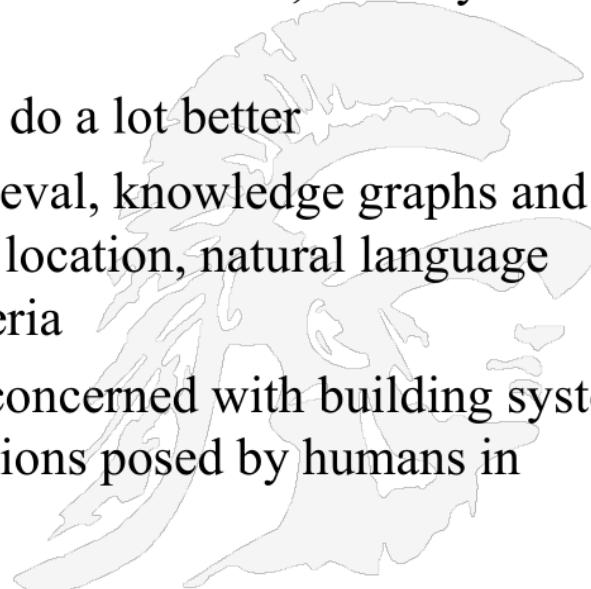
3:14:11

# Question answering

..

## Information Retrieval v. Question Answering

- The name “**information retrieval**” is standard, but as traditionally practiced, it’s not really right
- In the past all we got was ***document retrieval***, and beyond that the job is up to us
  - Modern search engines now do a lot better
- They combine information retrieval, knowledge graphs and inferencing, past query history, location, natural language processing and many other criteria
- **Question Answering (QA)** is concerned with building systems that automatically answer questions posed by humans in a natural language

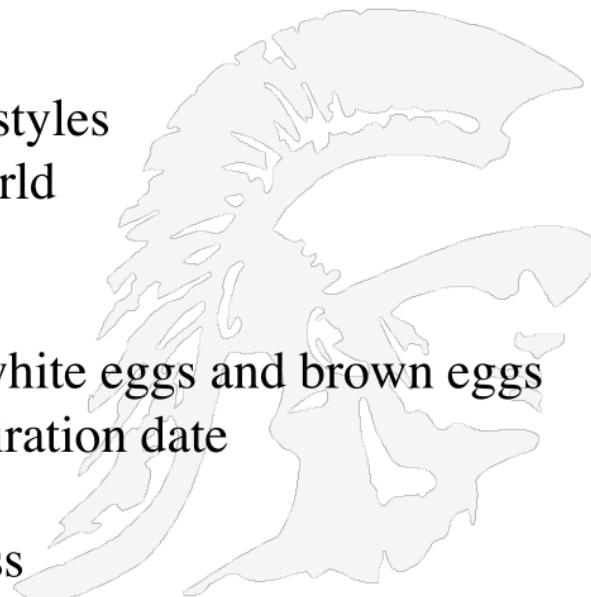


..

## 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 - 10

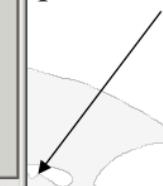
Asking a question? Try out Google Answers.

**From Poor Boy to Prime Minister**  
... how did he come to lead Australia during World War ... April 1939 Menzies 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, 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 under led by the 1930s. ... He became Prime Minister following Curtin's death, succeeding ... www.nma.gov.au/primeministers/3.htm - 30k - Cached - Similar pages

### How Google used to respond to questions



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)

•

University of Southern California  USC

**USC Viterbi**  
School of Engineering

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

Google's result today

→

who was the prime minister of australia during the great depression

All Images News Videos Shopping More Settings Tools

About 9,390,000 results (0.72 seconds)

James Scullin became the new prime minister and Bruce lost his own seat of Flinders, the first sitting Australian prime minister to do so. However, on 24 October 1929, one week after Labor took power, the US stock market crashed.

www.nma.gov.au › defining-moments › resources › great-depression › Great Depression | National Museum of Australia

About Featured Snippets Feedback

en.wikipedia.org › wiki › Great\_Depression\_in\_Australia › Great Depression in Australia - Wikipedia

Australia suffered badly during the period of the Great Depression of the 1930s. ... The conservative Prime Minister of Australia, Stanley Bruce, wished to ...  
1929–1935: Scullin and ... · Varying experiences of ... · Legacy of the Great ...

People also ask

Who was the prime minister of Australia during ww2?  
What areas of Australia were most affected by the Great Depression?  
Did the Great Depression affect Australia?  
Who was hit the hardest during the Great Depression?

Feedback

•

University of Southern California  USC

# Google has Improved Its Ability to Answer Many Questions

how old is mariah carey - Google

google.com/search?q=how+old+is+mariah+carey&lz=1C5CHFA\_enUS728US728&oq=how+ol...

CSCI 572 Home P... Piazza Spring2022 CSCI572\_Spring2... DEN D2L Page USC Schedule of...

Google how old is mariah carey

All Images News Books Videos More Tools

About 127,000,000 results (0.73 seconds)

Mariah Carey / Age

**53 years**

March 27, 1969



People also search for

 Nick Cannon 41 years  Jennifer Lopez 52 years  Beyoncé 40 years

Feedback

People also ask :

What is Mariah Carey's net worth 2021? ▼

Why does Mariah Carey touch her ear? ▼

Is Mariah Carey richer than Nick Cannon? ▼

Who is Mariah Carey husband? ▼

Feedback

**Mariah Carey**  
American singer-songwriter

Available on

 YouTube  
 Spotify  
 Apple Music

More music services

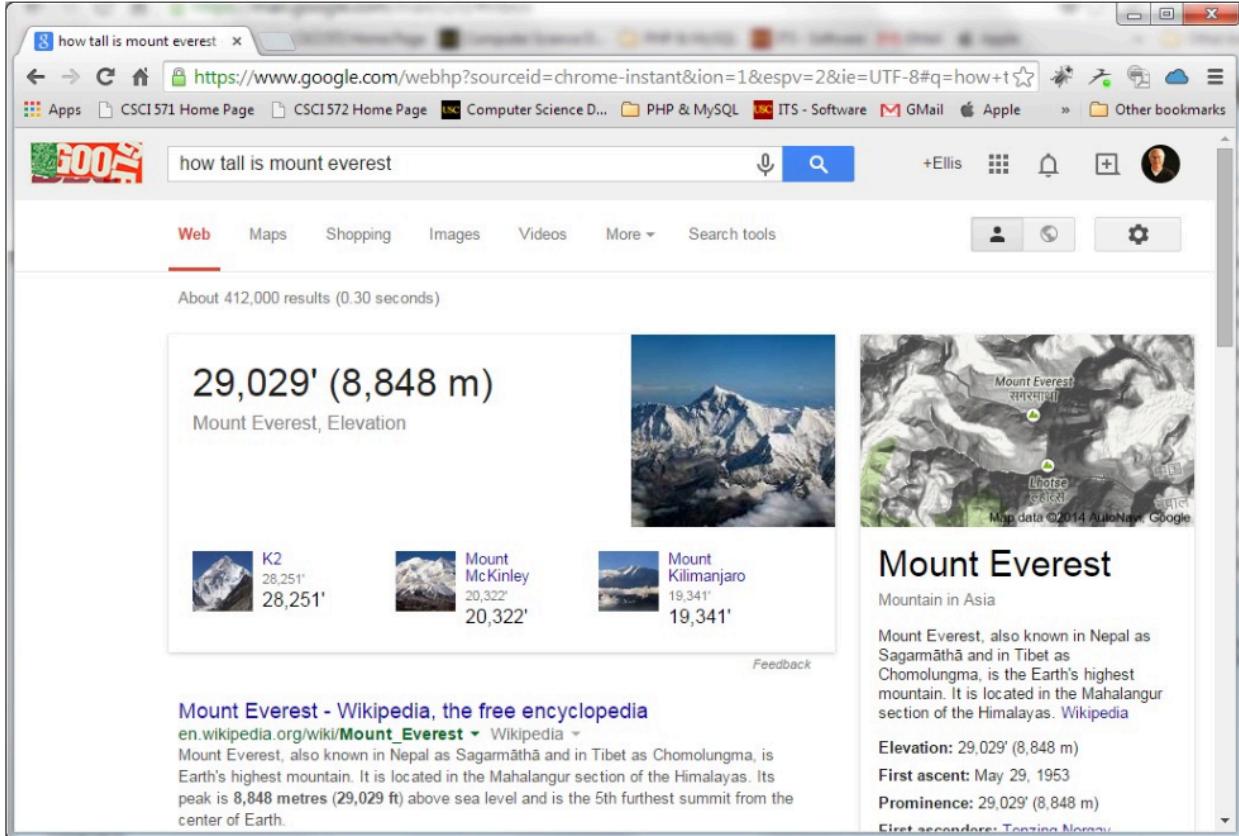
Mariah Carey is an American singer, actress, and record producer. Known for her powerful voice, "Songbird Supreme" and the "Queen of Pop", she is noted for her five-octave vocal range, melismatic singing style, and signature whistle register. Carey rose to fame with her eponymous debut album. [Wiki](#)

**Born:** March 27, 1969 (age 53 years)  
**Children:** Moroccan Scott Cannon  
**Spouse:** Nick Cannon (m. 2008–present); Mottola (m. 1993–1998)

•

University of Southern California  USC

**Some Questions are Easily Answered**



The screenshot shows a Google search results page for the query "how tall is mount everest". The top result displays the elevation "29,029' (8,848 m)" and a link to the Mount Everest Wikipedia page. To the right, there is a map showing the location of Mount Everest and its peak. Below the main result, there are cards for other mountains: K2 (28,251'), Mount McKinley (20,322'), and Mount Kilimanjaro (19,341'). The page also includes standard Google search navigation and a sidebar with various links.

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](https://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.

**Mount Everest**

Mountain in Asia

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. [Wikipedia](#)

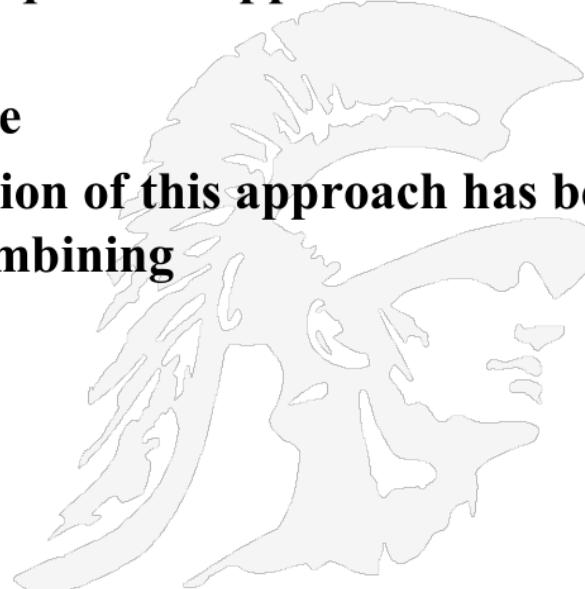
Elevation: 29,029' (8,848 m)  
First ascent: May 29, 1953  
Prominence: 29,029' (8,848 m)  
First ascender: Tenzing Norgay

Copyright Ellis Horowitz 2011-2022

..

## 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
  - WordNet
  - NLP techniques



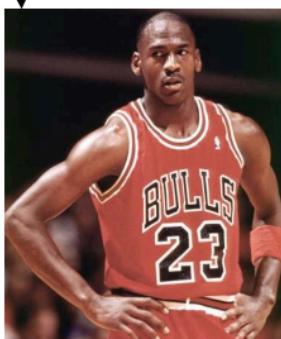
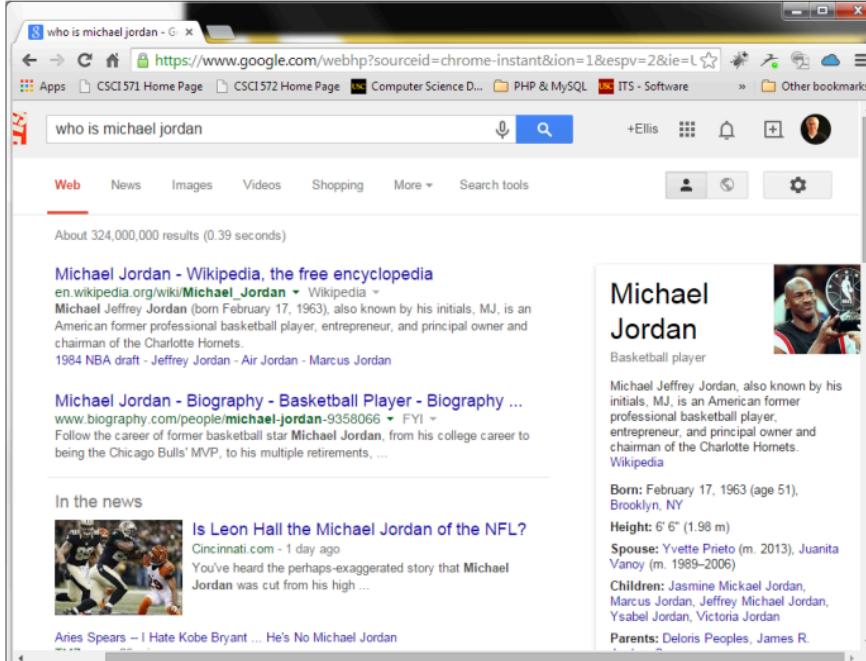
..

University of Southern California  USC

**USC Viterbi**  
School of Engineering

# 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



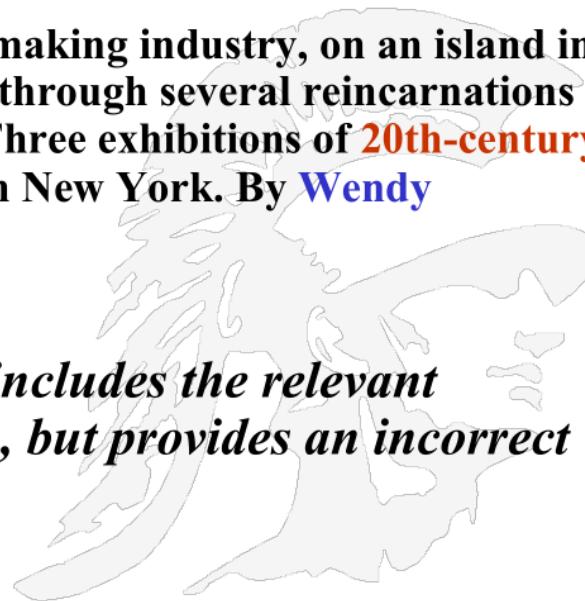
Copyright Ellis Horowitz, 2011-2022

9

..

## 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:** **20<sup>th</sup> Century**
- *the candidate passage below includes the relevant keywords (Wendy's, founded), but provides an incorrect answer*



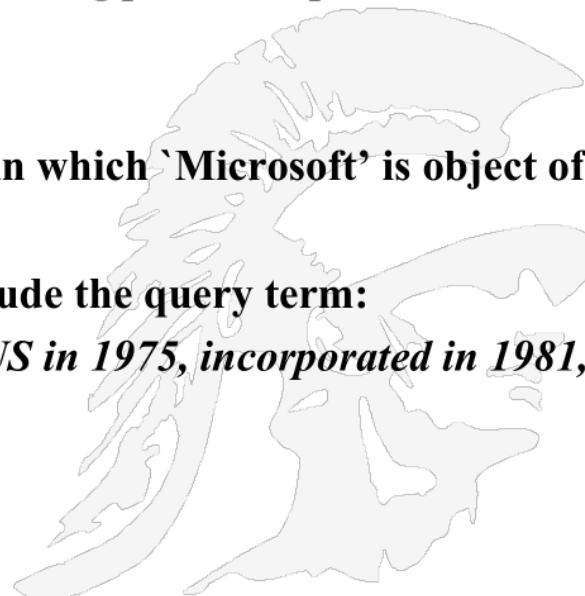
..

## 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.*



••

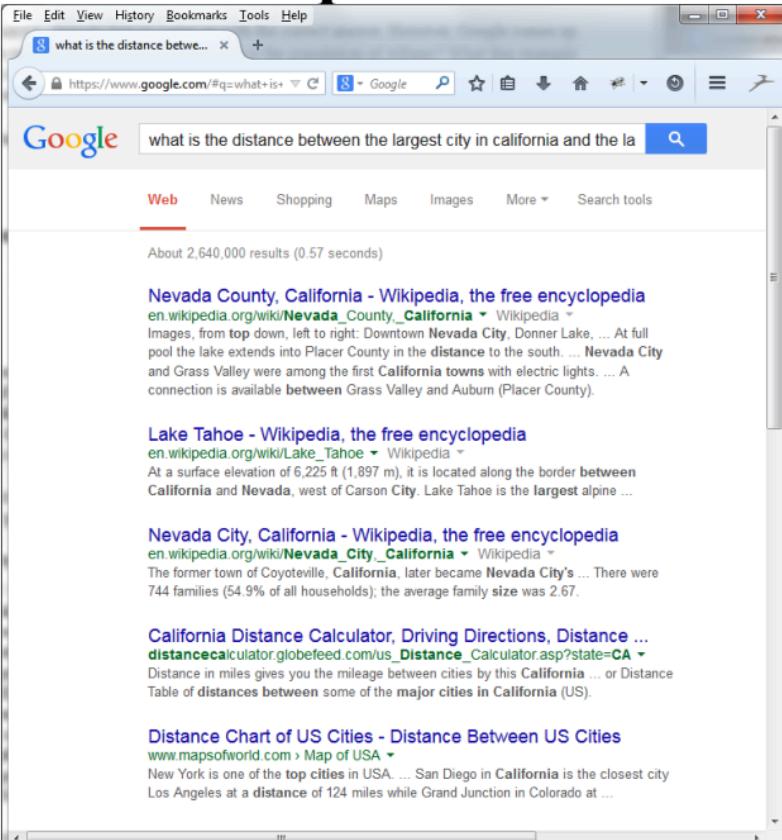
University of Southern California  USC

**Some Questions  
Require Inferences**

**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: Try the query in Google today to see if they have improved their answer



The screenshot shows a Google search results page with the query "what is the distance between the largest city in California and the largest city in Nevada?". The results are as follows:

- Nevada County, California - Wikipedia, the free encyclopedia**  
[en.wikipedia.org/wiki/Nevada\\_County,\\_California](https://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](https://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](https://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](http://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](http://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 ...

Copyright Ellis Horowitz 2011-2022

• •

University of Southern California  USC

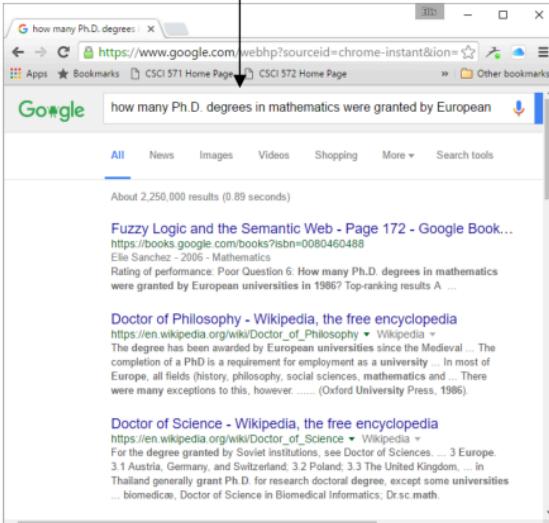
**USC Viterbi**  
School of Engineering

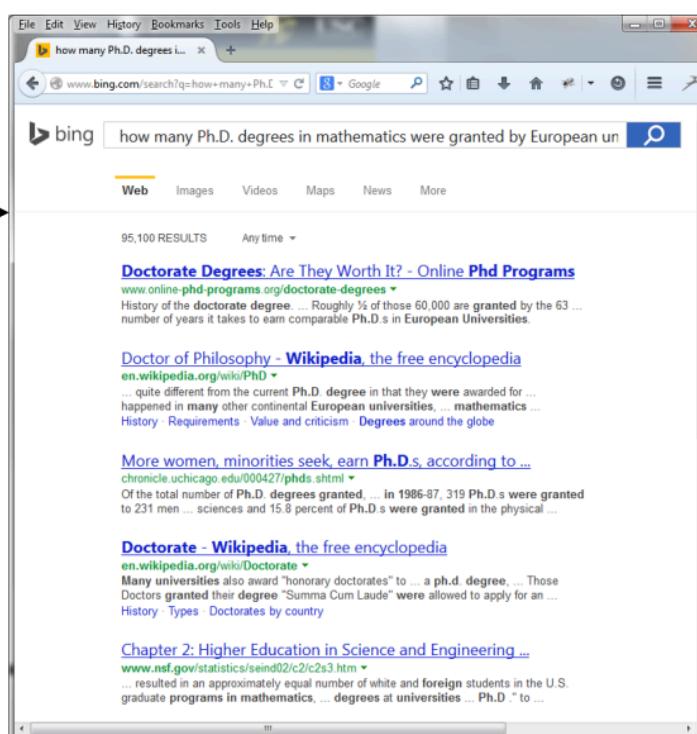
**In Some Cases  
the Data May Not Exist**

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



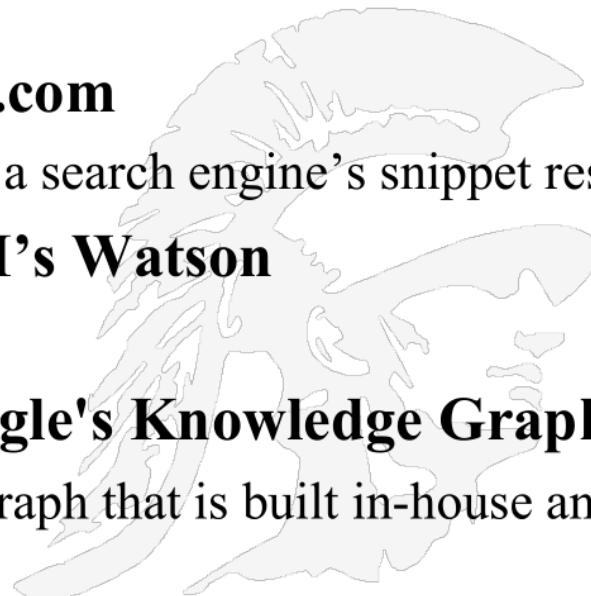


ght Ellis Horowitz 2011-2022

..

## 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 snippet 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 that is built in-house and infer the answer

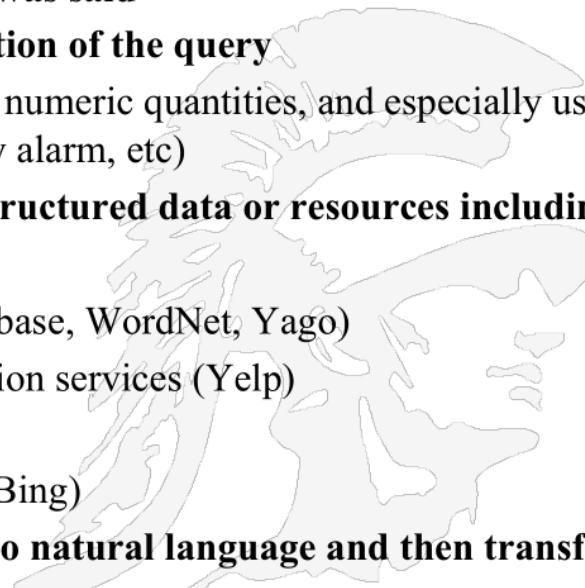


• •



## Approach used by Siri: Knowledge-Based Approach

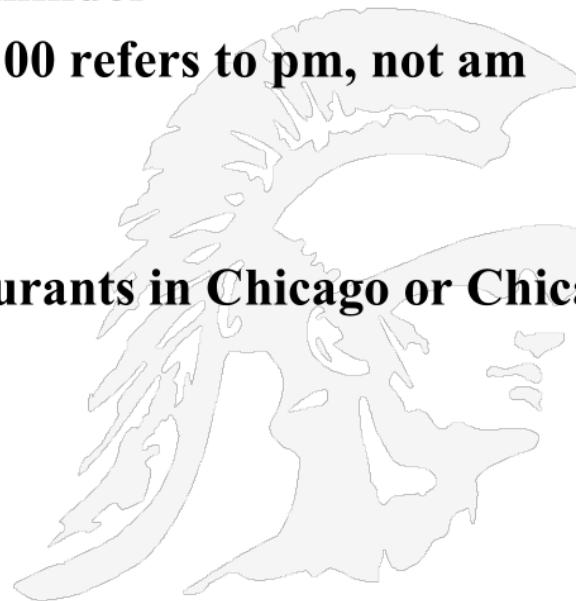
- Siri was begun as a DARPA project called CALO/PAL (Personalized Assistant that Learns)
- 1. First your voice query is put through a recognizer and a language model and Siri comes up with an interpretation of what was said
- 2. Second Siri builds a semantic representation of the query
  - Extract times, 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



..

## Context and Conversation in Virtual Assistants like Siri

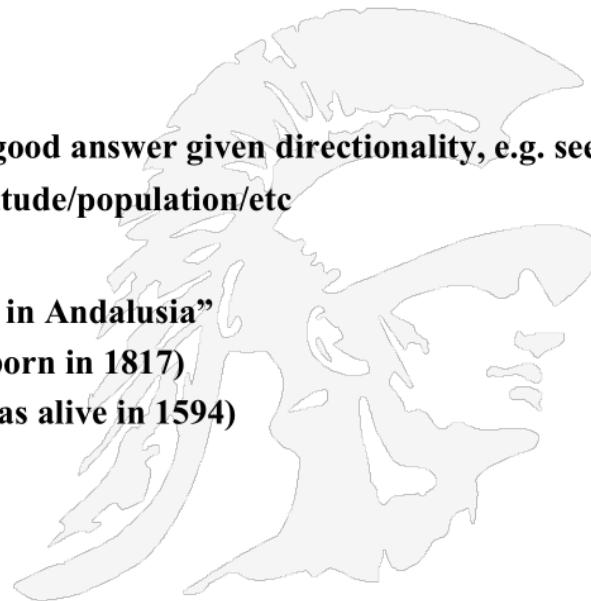
- Coreference helps resolve ambiguities
- U: “book a table at Il Fornaio at 7:00 with my mom”
- U: “also send her an email reminder”
- “her” refers to “my mom”; 7:00 refers to pm, not am
- Clarification questions:
- U: “chicago pizza”
- S: “Did you mean pizza restaurants in Chicago or Chicago-style pizza?”



..

## CANDIDATE ANSWER SCORING IN IBM WATSON

- Each candidate answer gets scores from > 50 components
- From unstructured text, semi-structured text, triple stores
- Logical form (parse) match between question and candidate
- Passage source reliability
- Geospatial location
  - Denver is “southwest of Montana” is a good answer given directionality, e.g. see
  - geonames.org which gives latitude/longitude/population/etc
- Temporal relationships
  - “In 1594 he took a job as a tax collector in Andalusia”
  - Candidates: Thoreau is a bad answer (born in 1817)
  - Candidates: Cervantes is possible (he was alive in 1594)
- Taxonomic classification

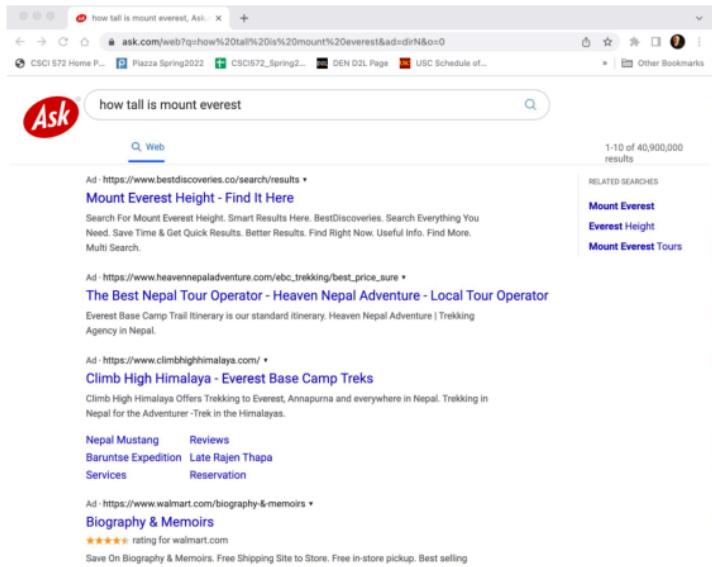
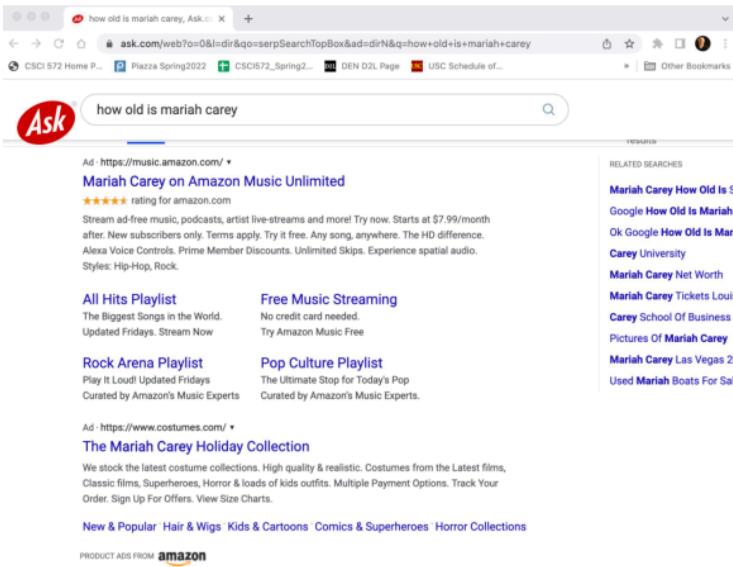


•

University of Southern California  USC

**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



How old is Mariah Carey  
Snapshot taken 04/2022

How tall is mount Everest  
Snapshot taken 04/2022

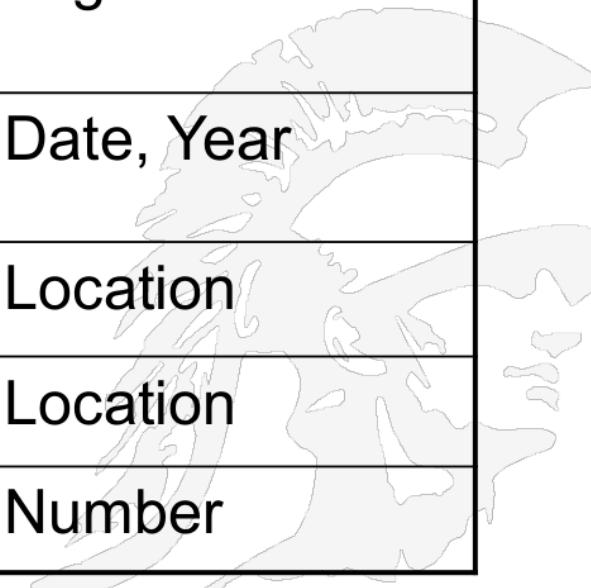
Copyright Ellis Horowitz 2011-2022

..



## Question Types: Many Questions Fall into Distinct Categories

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



..

## 3 Main Phases for Question/Answering

### 1. QUESTION PROCESSING

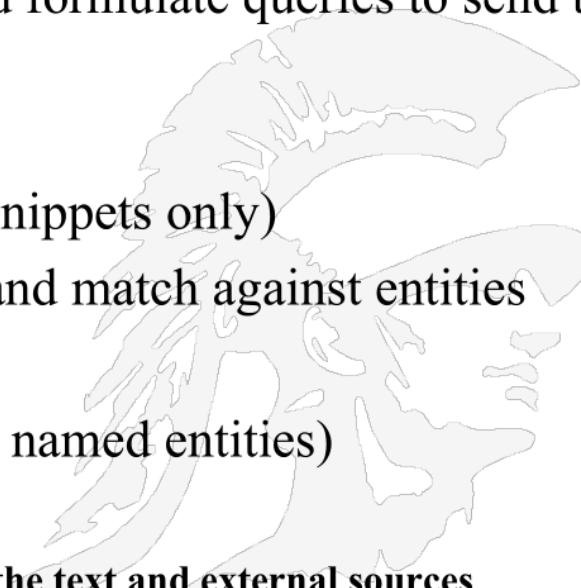
- Detect question type (who, what, when, where, etc)
- Identify important entities and formulate queries to send to a search engine

### 2. PASSAGE RETRIEVAL

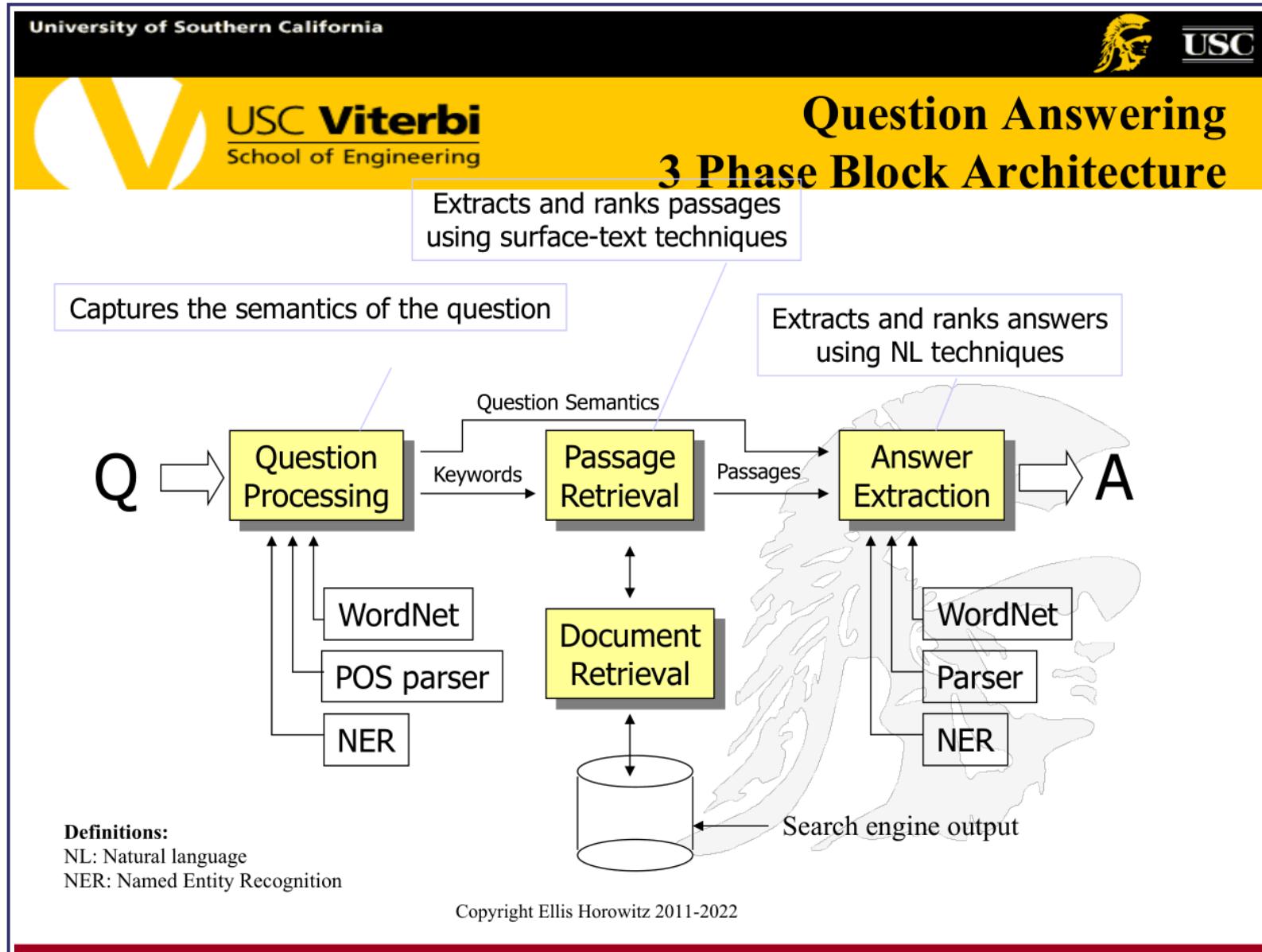
- Retrieve ranked documents (snippets only)
- Break into suitable passages and match against entities

### 3. ANSWER PROCESSING

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

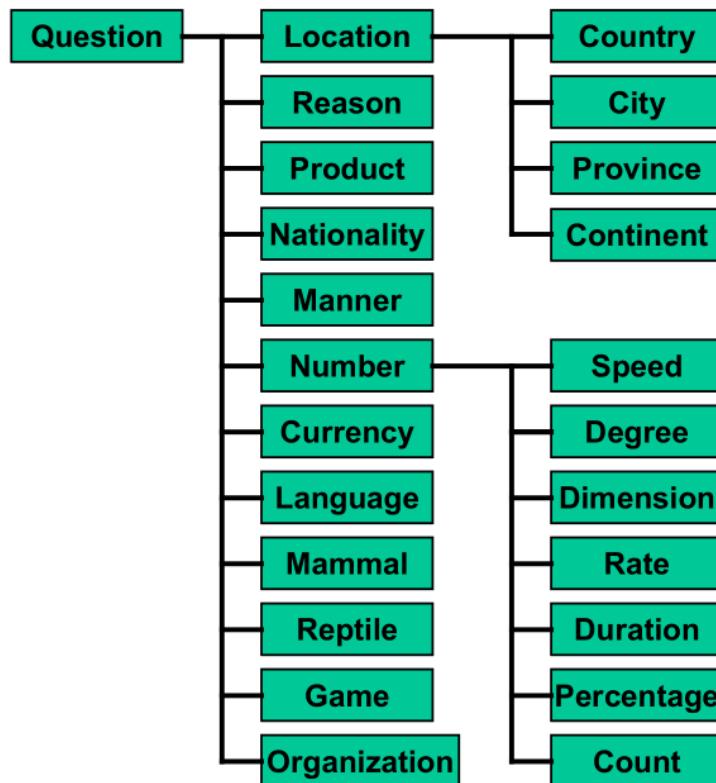


••



• •

## Question Taxonomy



Researchers have tried to organize all question types, e.g.



• •

- **Factoid Questions**
  - Who, where, when, how many
  - The answers fall into a limited and somewhat predictable set of categories, see e.g. to the right
- 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
- This approach argues for a template applied to the query

## However Question Taxonomies Can Get Very Large

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?



• •

## More Question Types and Examples

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?
speed	What is the speed of the Mississippi River?
temp	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?

• •

## Some General Capabilities for Question-Answering Systems

- **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](https://en.wikipedia.org/wiki/Part-of-speech_tagging)

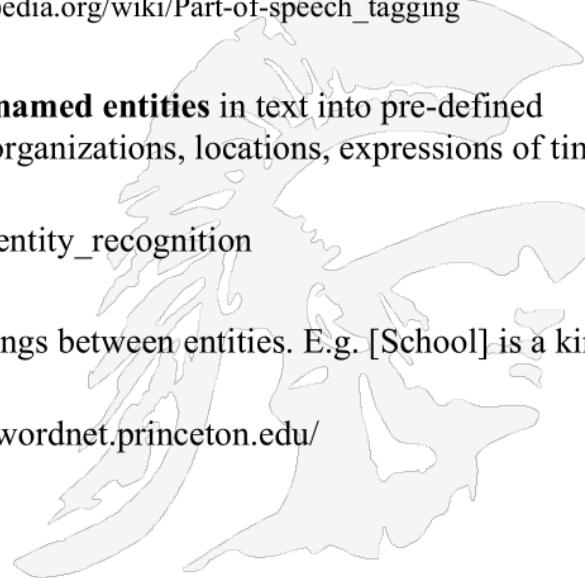
- **Named Entity Extraction**

- 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](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/Thesauri**





## Question Processing Tool Named Entity Recognizer Example

contentSkip to site indexPoliticsSubscribeLog InSubscribeLog InToday's PaperAdvertisementSupported ORG byF.B.I. Agent Peter Strzok PERSON , Who Criticized Trump PERSON in Texts, Is FiredImagePeter Strzok, a top F.B.I. GPE counterintelligence agent who was taken off the special counsel investigation after his disparaging texts about President Trump PERSON were uncovered, was fired. CreditT.J. Kirkpatrick PERSON for The New York TimesBy Adam Goldman ORG and Michael S. SchmidtAug PERSON . 13 CARDINAL , 2018WASHINGTON CARDINAL — Peter Strzok PERSON , the F.B.I. GPE senior counterintelligence agent who disparaged President Trump PERSON in inflammatory text messages and helped oversee the Hillary Clinton PERSON email and Russia GPE investigations, has been fired for violating bureau policies, Mr. Strzok PERSON 's lawyer said Monday DATE .Mr. Trump and his allies seized on the texts — exchanged during the 2016 DATE campaign with a former F.B.I. GPE lawyer, Lisa Page — in PERSON assailing the Russia GPE investigation as an illegitimate "witch hunt." Mr. Strzok PERSON , who rose over 20 years DATE at the F.B.I. GPE to become one of its most experienced counterintelligence agents, was a key figure in the early months DATE of the inquiry.Along with writing the texts, Mr. Strzok PERSON was accused of sending a highly sensitive search warrant to his personal email account.The F.B.I. GPE had been under immense political pressure by Mr. Trump PERSON to dismiss Mr. Strzok PERSON , who was removed last summer DATE from the staff of the special counsel, Robert S. Mueller III PERSON . The president has repeatedly denounced Mr. Strzok PERSON in posts on

- The process of recognizing information units like names, including persons, organizations, location names, and numeric expressions including time, date, money and percent expressions from unstructured text.
- This is an example of supervised learning as training sets are first created
- See <https://nlp.stanford.edu/software/> for a Java program and explanation

••

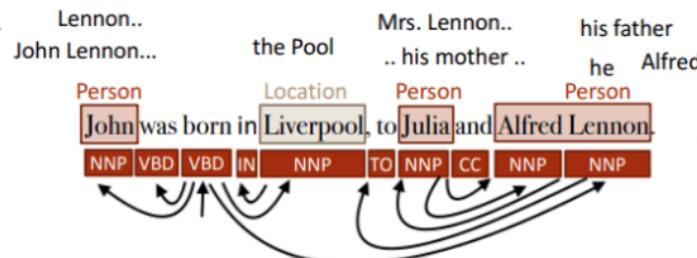


## Example of NLP Extraction Used to Build a Knowledge Graph:

John was born in Liverpool, to Julia and Alfred Lennon.

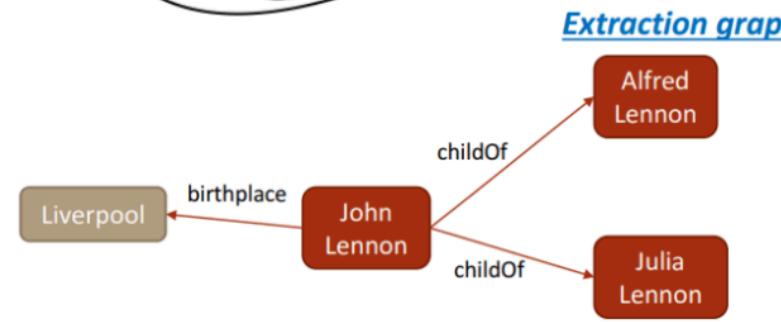
[Text](#)

NLP



[Annotated text](#)

Information  
Extraction



[Extraction graph](#)

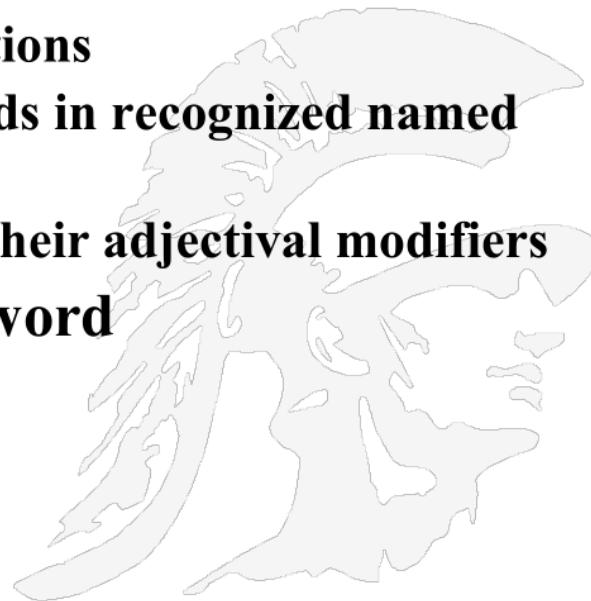
Nouns are (usually) entities and verbs are (usually) relationships among the entities and adjectives (usually) describe the relationships

[https://kgtutorial.github.io/wsdm-slides/Part2\\_knowledge-extraction.pdf](https://kgtutorial.github.io/wsdm-slides/Part2_knowledge-extraction.pdf)  
Copyright Ellis Horowitz 2011-2022

..

## General Keyword Selection Algorithm

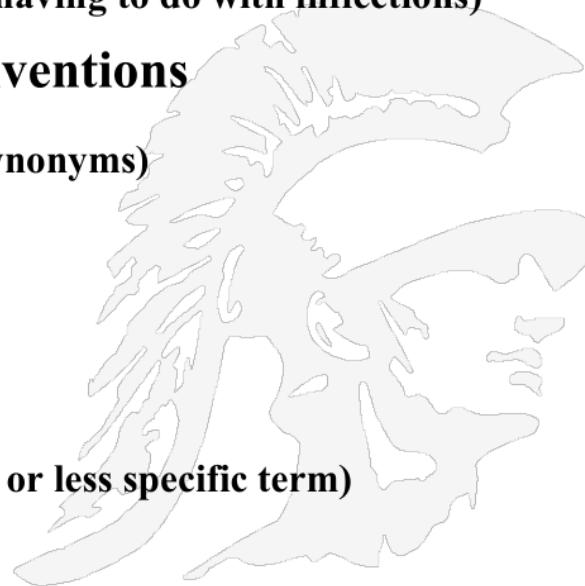
- 1. Use the part-of-speech recognizer to identify all
  - nouns
  - verbs
  - non-stopwords in quotations
  - NNP (proper noun) words in recognized named entities
  - complex nominals with their adjectival modifiers**
- 2. Select the answer type word**



..

## Expanding the Keyword Set Using Variants

- There are 3 distinct ways to expand the keyword set determined by the keyword selection algorithm
- Morphological variants (having to do with inflections)
  - invented → inventor → inventions
- Lexical variants (similar to synonyms)
  - killer → assassin
  - far → distance
- Semantic variants
  - like → prefer (a more specific or less specific term)



..

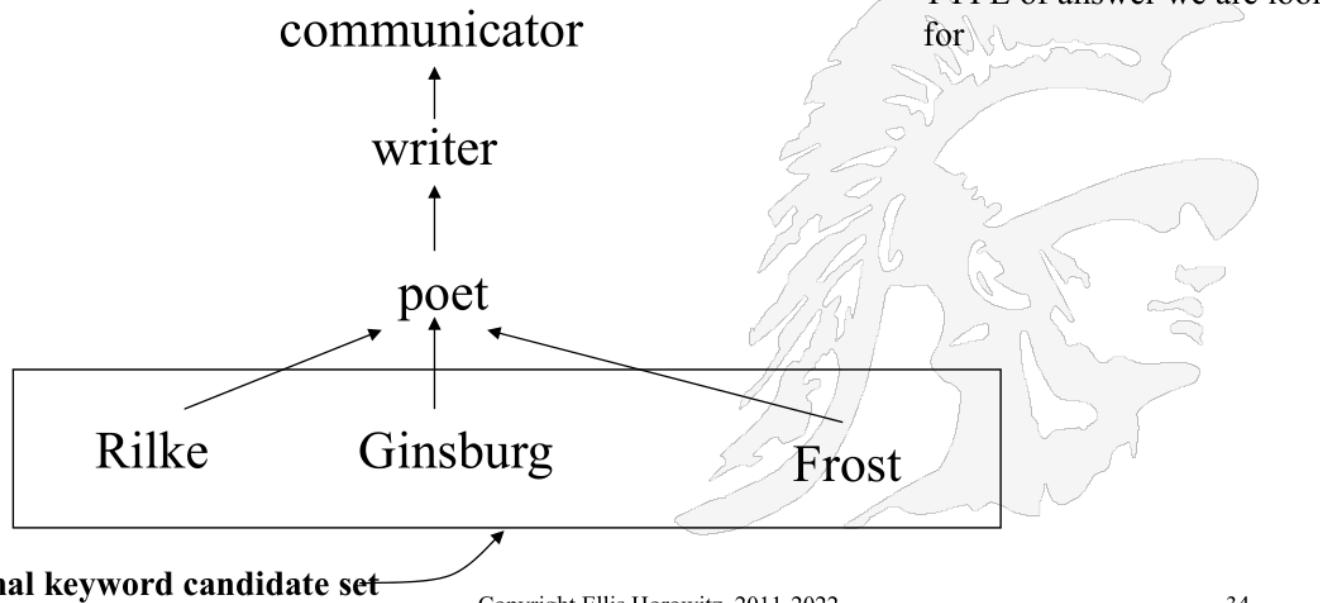
## Use WordNet for Type Identification

We have already seen the use of WordNet, a lexical database of English nouns, verbs, adjectives, adverbs

**“What 20<sup>th</sup> century poet wrote Howl?”**

WordNet permits refinement  
of poet to specific instances

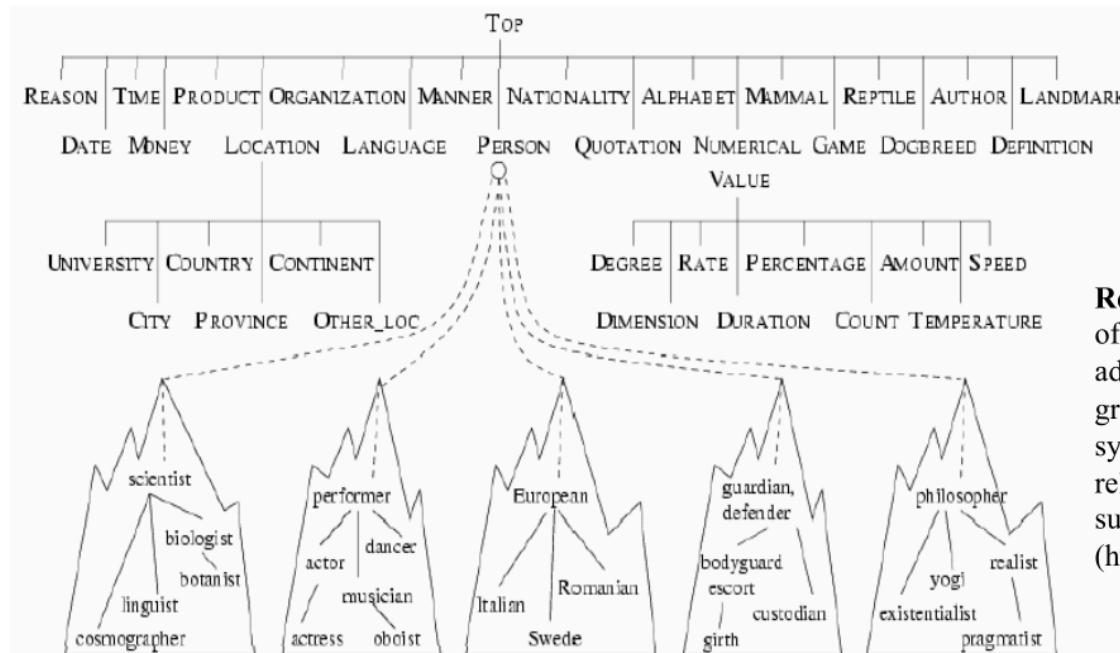
WordNet also helps determine the  
TYPE of answer we are looking  
for





# Answer Type Taxonomy

- Use WordNet to merge 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)

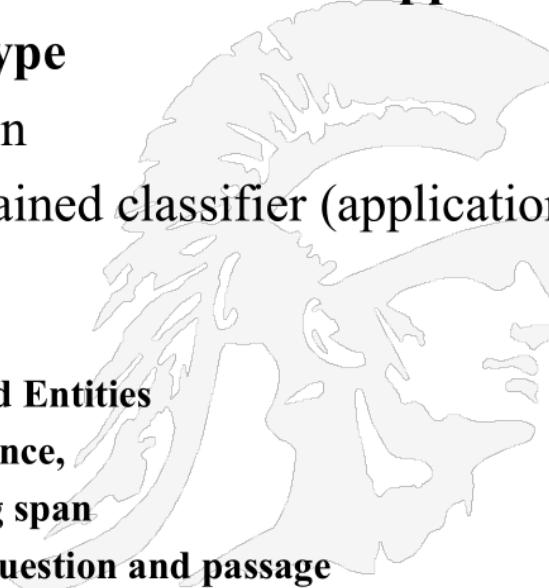
If you know the answer should be a person  
WordNet helps determine what sort of person

..



## Part 2: Passage Retrieval

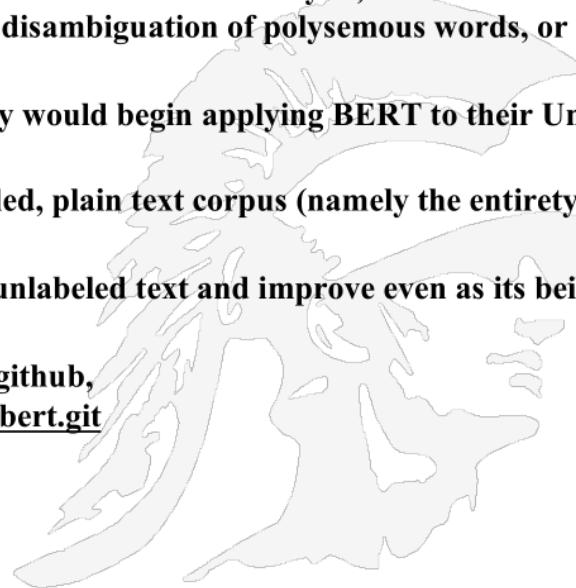
- Once we have formulated queries using tools like NER, POS, variant expansion and WordNet results
- Send queries to a search engine and retrieve snippet results
- Filter the results for correct type
  - use answer type classification
  - Rank passages based on a trained classifier (application of machine learning)
    - Features:
      - Question keywords, Named Entities
      - Longest overlapping sequence,
      - Shortest keyword-covering span
      - N-gram overlap between question and passage



• •

## What is BERT

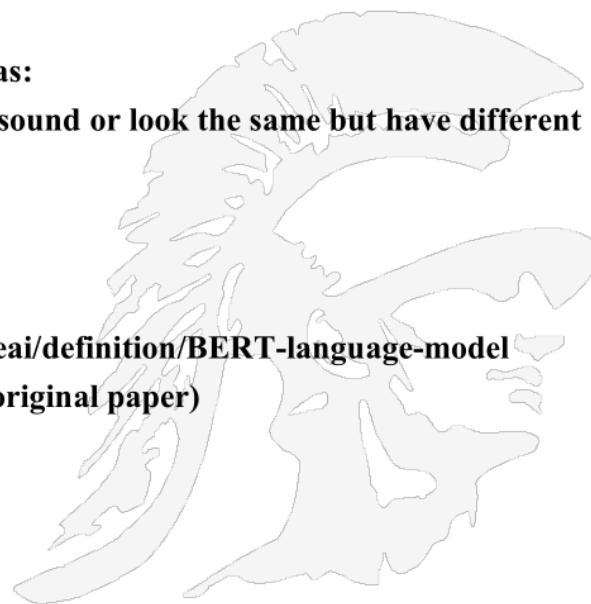
- **Bidirectional Encoder Representations from Transformers**
- In 2018, Google introduced and open-sourced BERT
  - BERT is designed to help computers understand the meaning of ambiguous language in text by using surrounding text to establish context
  - It achieved strong results on problems such as sentiment analysis, semantic role labeling, sentence classification and the disambiguation of polysemous words, or words with multiple meanings
- In October 2019, Google announced that they would begin applying BERT to their United States based production search algorithms.
- BERT was pre-trained using only an unlabeled, plain text corpus (namely the entirety of the English Wikipedia).
- It continues to learn unsupervised from the unlabeled text and improve even as its being used in practical applications
- the BERT Github repository is available on github,  
git clone <https://github.com/google-research/bert.git>



• •

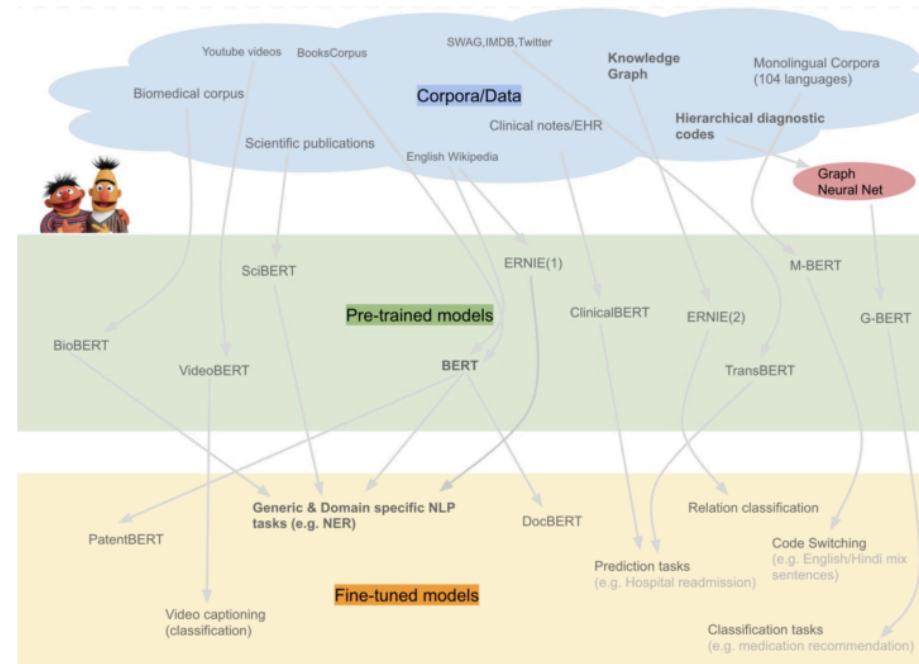
# What is BERT used for

- Sequence-to-sequence based language generation tasks such as:
  - Question answering
  - Abstract summarization
  - Sentence prediction
  - Conversational response generation
- Natural language understanding tasks such as:
  - Polysemy and Coreference (words that sound or look the same but have different meanings) resolution
  - Word sense disambiguation
  - Natural language inference
  - Sentiment classification
- <https://www.techtarget.com/searchenterpriseai/definition/BERT-language-model>
- <https://csci572.com/papers/BERT.pdf> (the original paper)



## BERT Has Many Pre-Trained Models

- BERT encoders have larger feedforward networks (768 and 1024 nodes in Base and Large respectively) and more attention heads (12 and 16 respectively).**
- BERT was trained on Wikipedia and other datasets**
- To the right you can see a diagram of additional variants of BERT pre-trained on specialized corpora**



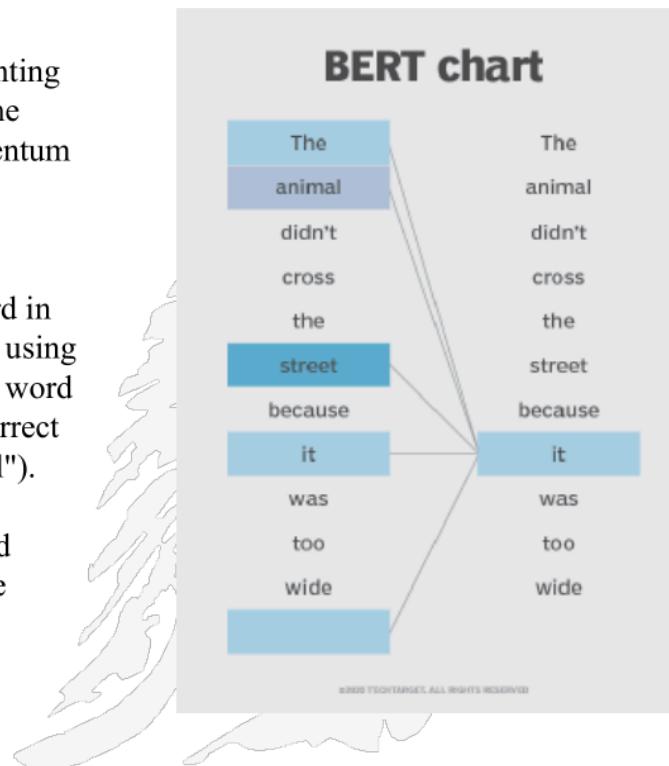
<https://towardsdatascience.com/bert-for-dummies-step-by-step-tutorial-fb90890ffe03>





# How BERT works

- BERT functions by reading bidirectionally, accounting for the effect of all other words in a sentence on the focus word and eliminating the left-to-right momentum that biases words towards a certain meaning as a sentence progresses.
- At the right BERT is determining which prior word in the sentence the word "it" is referring to, and then using its attention mechanism to weigh the options. The word with the highest calculated score is deemed the correct association (i.e., "it" refers to "street", not "animal").
- If this phrase was a search query, the results would reflect this subtler, more precise understanding the BERT reached.



• •

## BERT for Question-Answering

- question answering is just a prediction task
- on receiving a question as input, the goal of the application is to identify the right answer from some corpus
- given a question and a context paragraph, the model predicts a start and an end token from the paragraph that most likely answers the question.

- Input Question:

Where do water droplets collide with ice crystals to form precipitation?

- Input Paragraph:

... Precipitation forms as smaller droplets coalesce via collision with other rain drops or ice crystals within a cloud. ...

- Output Answer:

within a cloud



Here is a thorough explanation of BERT.