
Question Answering and Knowledge Graph

Dekang Lin
[@naturali.io](mailto:lindekang@gmail.com)

Overview

- ◆ Question answering
 - Knowledge graph based approaches
 - Search-based approaches
- ◆ Knowledge graph population / completion
 - Information extraction based methods
 - Question answering based methods

Question Answering: three approaches

♦ Question-to-Question

- Map user questions to questions with known answers in a database of question-answer pairs.

♦ With structured knowledge base

- Translate users' questions to knowledge base retrieval queries

♦ With unstructured text

- Use IR to identify top documents / passages
- Extract answers from the retrieved text

Question to Question: an example



哈利波特里罗恩是谁演的



百度一下

[网页](#) [新闻](#) [贴吧](#) [知道](#) [音乐](#) [图片](#) [视频](#) [地图](#) [文库](#) [更多»](#)

百度为您找到相关结果约7,030个

[搜索工具](#)

问 哈利波特里罗恩是谁演的:

• [哈利波特里罗恩扮演者是谁?_百度知道](#)

鲁伯特·格林特 中文名: 鲁伯特·格林特 英文名: Rupert Grint 全名:Rupert Alexander Lloyd Grint 性别: 男 生日: 1988-8-24 身高:1.78米 出生地:...

来自百度知道 | 1个回答 | 2015-04-25

• [哈利波特里罗恩的扮演者_百度知道](#)

中文名: 鲁珀特·格林特 英文名: Rupert Grint 全名:Rupert Alexander Lloyd Grint 假的

来自百度知道 | 21个回答 | 2009-01-30

Question Answering: three approaches

♦ Question-to-Question

- Map user questions to questions with known answers in a database of question-answer pairs.

♦ With structured knowledge base

- Translate users' questions to knowledge base retrieval queries

♦ With unstructured text

- Use IR to identify top documents / passages
- Extract answers from the retrieved text

QA with Structured Knowledge Base

- ◆ Knowledge base: a collection of triples: $(E_{\text{subj}}, \text{Rel}, E_{\text{obj}})$
 - ❖ Any relation can be decomposed into a set of triples
 - ❖ Example
 - ❖ (“Pulp Fiction /m/0f4_l”, DirectedBy, “Quentin Tarantino /m/0693l”)
- ◆ Knowledge bases can be accessed through formal query language
 - ❖ E.g. SQL, Lambda calculus,
 - ❖ However, user queries are in natural language

NL Relation vs KB Relation

- ♦ NL relations and KB relations may not directly correspond to each other.
- ♦ Example:
 - Natural language:
 - Mavis Leno is the spouse of Mavis Leno
 - Freebase:
 - (“Jay Leno /m/046lt”, Marriage, Internal_ID:Marriage1234)
 - (Internal_ID:Marriage1234, Spouse, “Mavis Leno /m/06vbjz”)
 - (Internal_ID:Marriage1234, From, “11/30/1980”)

Semantic Parsing

- ♦ Translate user queries / questions into database retrieval query
- ♦ Example:
 - “Who was Tom Cruise’s wife in 2007?”
 - $\lambda x. \exists e. \text{Marriage}(x, e) \wedge \text{Spouse}(e, \text{TomCruise}) \wedge \text{Year}(\text{MarriageFrom}(e)) \leq 2007 \wedge \text{Year}(\text{MarriageTo}(e)) \geq 2007$



Who was Tom Cruise's wife in 2007



All

News

Images

Videos

Shopping

More ▼

Search tools

About 9,540,000 results (1.44 seconds)

Tom Cruise / Spouse (2007)

Katie Holmes

m. 2006–2012



Movies and TV shows and overview

Feedback



Who was Tom Cruise's wife in 2007



Web

Images

Videos

Maps

News

Explore

3,720,000 RESULTS

Any time ▼

Including results for *who was tom **cruise** wife in 2007.*

Do you want results only for Who was Tom Cruise's wife in 2007?

Tom Cruise - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/Tom_Cruise ▼

Cruise met his second **wife**, ... Rathbun claims that **wife** Nicole Kidman was wiretapped on **Tom Cruise's** ... October 17, 2007. ^ a b "**Tom Cruise's** Irish ...

[Early life](#) · [Career](#) · [Relationships](#) · [Scientology](#) · [Legacy](#) · [Litigation](#)

Tom Cruise on "Valkyrie" Set - Photo 1 - Pictures - CBS News

www.cbsnews.com/pictures/tom-cruise-on-valkyrie-set ▼

Tom Cruise on "Valkyrie" Set. Facebook; Twitter; ... 2007. Suri and **Cruise's wife**, ... 2007. Tom Cruise stars in the film, ...



Tom Cruise's wife



Web

Images

Videos

Maps

News

Explore

3,560,000 RESULTS

Any time ▼

Tom Cruise - Wife

Katie Holmes

m. 2006 - 2012



Nicole Kidman

m. 1990 - 2001



Mimi Rogers

m. 1987 - 1990



Semantic Parsing

- ◆ Grammar induction with supervised examples
 - Too costly for create training data
 - Not enough precision for users
- ◆ Manually constructed rules
- ◆ Distant supervision

Manually Constructed Rules

♦ Example:

- “?(Who is | was) \$PERSON ’s wife ?(in \$YEAR)”
 - $\lambda x. \exists e. \text{Marriage}(x, e) \wedge \text{Spouse}(e, \$\text{PERSON}) \wedge \text{Year}(\text{MarriageFrom}(e)) \leq \$\text{YEAR} \wedge \text{Year}(\text{MarriageTo}(e)) \geq \YEAR

♦ The good: predictable, precise

♦ The bad: costly, low coverage

- But bearable with very high query traffic

♦ Rule-based NLP has been dead for a long time in academia, but is very much alive in industry.

Indiana Jones vs Swordman



Semantic Parsing

- ◆ Grammar induction with supervised examples
 - Too costly for create training data
 - Not enough precision for users
- ◆ Manually constructed rules
- ◆ Distant supervision

Distant Supervision

- ◆ Instead of manually label examples, extract examples from an unlabeled corpus with a heuristic rule.
- ◆ Supervised learning with the extracted (weakly-labeled) examples.

Mappings between KB and NL Text

- ✦ Relations in KB and NL text are specified differently
 - infidelity.participant vs. “cheated on”
 - marriage.spouse vs “wife”
- ✦ Connection can be established through aliases of entities
 - KB entities have canonical names and aliases in natural language

Distant Supervision

- ♦ Goal:
 - Given a KB relation (e.g, Marriage.Spouse), find natural language text that expresses relation?
- ♦ Heuristic (assumption):
 - If two entities in a KB relation co-occur in the same sentence (passage), their relation is the same as the KB relation.
- ♦ Extract training examples using the heuristic for supervised training.

Distant Supervision: An Example

◆ KB instance:

- (“冯小刚”, [Marriage.Spouse](#), “徐帆”)

◆ Text containing both entities:

- 1999年，徐帆与导演冯小刚结婚。
- 冯小刚发表长微博，深情表白妻子徐帆。
- 据悉，这篇文章是14年前冯小刚写给妻子的，当时给徐帆看完，她笑得合不拢嘴。
- 冯小刚和现任老婆徐帆为什么不生子内幕.

Distant Supervision: An Example

◆ KB instance:

- (“冯小刚”, [Marriage.Spouse](#), “徐帆”)

◆ Text containing both entities:

- 1999年，徐帆与导演冯小刚结婚。
- 冯小刚发表长微博，深情表白妻子徐帆。
- 据悉，这篇文章是14年前冯小刚写给妻子的，当时给徐帆看完，她笑得合不拢嘴。
- 冯小刚和现任老婆徐帆为什么不生子内幕。

Distant Supervision: An Example

◆ KB instance:

- (“李连杰”, [Marriage.Spouse](#), “利智”)

◆ Text containing both entities:

- 李连杰在接受媒体采访时透露，这几年全身心的投入到慈善公益事业中， ... 他也坦言财产全部由妻子利智打理
- 提到李连杰的妻子，在多数人印象中会第一时间想到利智，
- 时过境迁，李连杰和利智结婚十多年了，育有一对女儿。
- 大家都知道功夫巨星李连杰，谈到李连杰的老婆大家就会想到利智。
- 相信很多朋友都知道，影坛巨星李连杰与利智结婚后，自2000年起相继的生育有两个女儿

Distant Supervision: example

◆ KB instance:

- (“李连杰”, Marriage.Spouse, “利智”)

◆ Text containing both entities:

- 李连杰在接受媒体采访时透露，这几年全身心的投入到慈善公益事业中， ... 他也坦言财产全部由妻子利智打理
- 提到李连杰的妻子，在多数人印象中会第一时间想到利智，
- 时过境迁，李连杰和利智结婚十多年了，育有一对女儿。
- 大家都知道功夫巨星李连杰，谈到李连杰的老婆大家就会想到利智。
- 相信很多朋友都知道，影坛巨星李连杰与利智结婚后，自2000年起相继的生育有两个女儿

Distant Supervision: source of natural language text

- ♦ [Cai & Yates 2013]:
 - Snippets from top search results
- ♦ [Yao & van Durme 2014]:
 - Sentences from ClueWeb

Question Answering: three approaches

♦ Question-to-Question

- Map user questions to questions with known answers in a database of question-answer pairs.

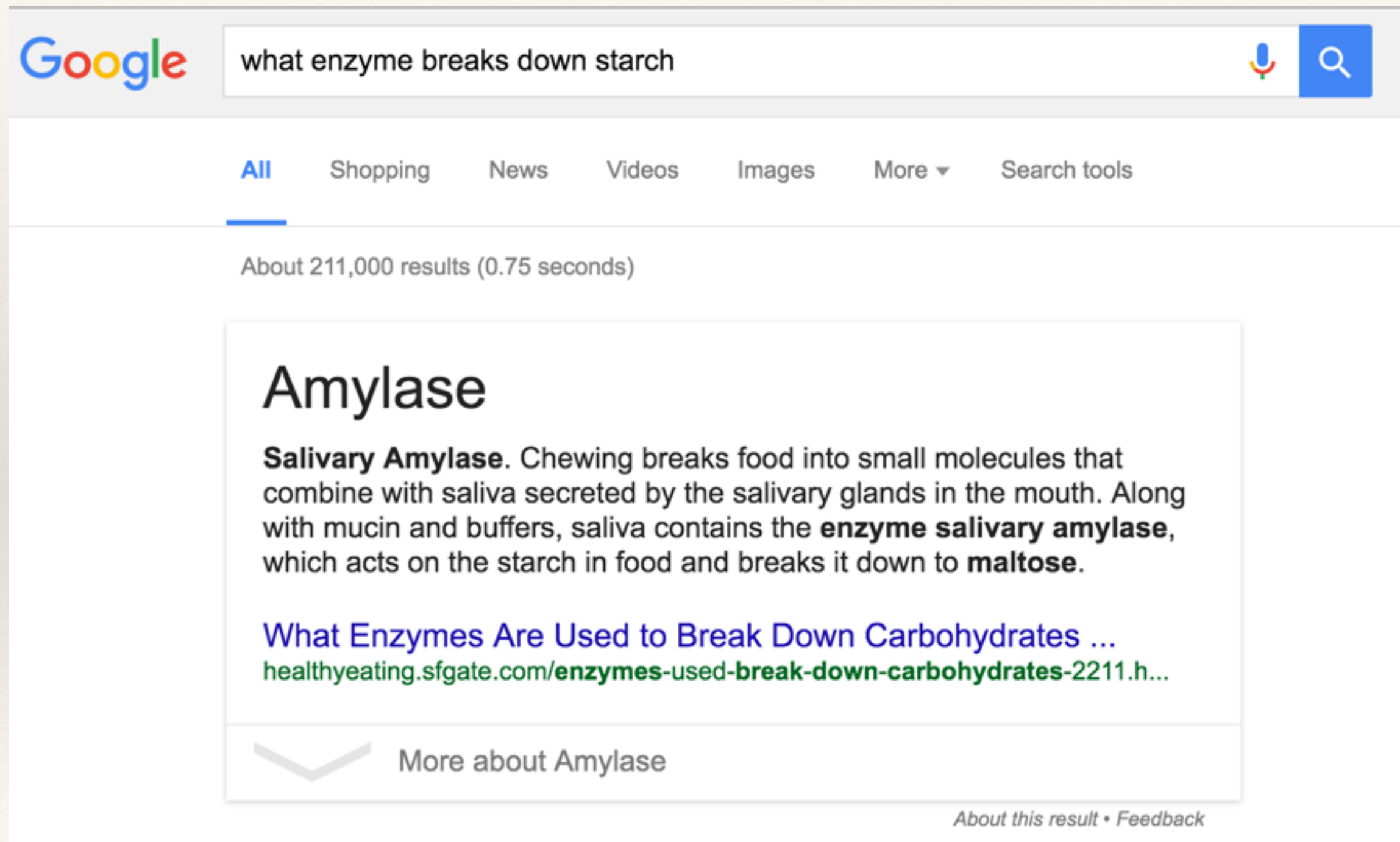
♦ With structured knowledge base

- Translate users' questions to knowledge base retrieval queries

♦ With unstructured text

- Use IR to identify top documents / passages
- Extract answers from the retrieved text

Question Answering with Unstructured Text



The image is a screenshot of a Google search interface. At the top left is the Google logo. To its right is a search bar containing the text "what enzyme breaks down starch". To the right of the search bar are a microphone icon and a blue search button with a white magnifying glass. Below the search bar is a horizontal menu with tabs: "All" (which is underlined in blue), "Shopping", "News", "Videos", "Images", "More" (with a downward arrow), and "Search tools". Below the menu, it says "About 211,000 results (0.75 seconds)". The main content area features a knowledge panel for "Amylase". The panel has the title "Amylase" in large black font. Below the title is a paragraph: "Salivary Amylase. Chewing breaks food into small molecules that combine with saliva secreted by the salivary glands in the mouth. Along with mucin and buffers, saliva contains the **enzyme salivary amylase**, which acts on the starch in food and breaks it down to **maltose**." Below this paragraph is a link in blue text: "What Enzymes Are Used to Break Down Carbohydrates ..." followed by a green URL: "healthyeating.sfgate.com/enzymes-used-break-down-carbohydrates-2211.h...". At the bottom of the panel is a grey chevron icon pointing down and the text "More about Amylase". At the very bottom right of the page, there is a small link: "About this result • Feedback".

Google

what enzyme breaks down starch

All Shopping News Videos Images More Search tools

About 211,000 results (0.75 seconds)

Amylase

Salivary Amylase. Chewing breaks food into small molecules that combine with saliva secreted by the salivary glands in the mouth. Along with mucin and buffers, saliva contains the **enzyme salivary amylase**, which acts on the starch in food and breaks it down to **maltose**.

[What Enzymes Are Used to Break Down Carbohydrates ...](#)
healthyeating.sfgate.com/enzymes-used-break-down-carbohydrates-2211.h...

More about Amylase

About this result • Feedback



annual rainfall of beijing



All

Images

Maps

News

Shopping

More ▼

Search tools

About 216,000 results (0.40 seconds)

December.

Rainfall/ Precipitation in Beijing, China. Beijing receives on average **576.9 mm** (22.7 in) of rainfall per year, or 48.1 mm (1.9 in) per month. On average there are 66 days per year with more than 0.1 mm (0.004 in) of rainfall (precipitation) or 5.5 days with a quantity of rain, sleet, snow etc. per month.

[Rainfall in Beijing, China Average Precipitation and Wet Days](http://www.beijing.climateemps.com/precipitation.php)
www.beijing.climateemps.com/precipitation.php

About this result • Feedback



how much does a mac air weigh



Web

Images

News

Videos

More ▾

Search tools

About 906,000 results (0.30 seconds)

2.96 pounds

The 13" Macbook Air weighs **2.96 pounds** and the 13" Macbook Pro weighs **3.57 pounds**. However, with the 15" Macbook Pro weighing in at 4.46 pounds, the 13" Pro is much more comparable to the 13" Air than it is the 15" Pro.

[Why I Sold My 13" Air for a Macbook Pro Retina Display](#)

briangardner.com/macbook-pro-retina-display/

Feedback

[Apple - MacBook Air - Technical Specifications](#)

<https://www.apple.com/hk/en/macbook-air/specs.html> ▾

View the technical specifications for MacBook Air, see what comes in the box, ...

Search-based Question Answering

Query Analysis:

[what enzyme breaks down starch]

[what **enzyme** breaks down starch]

Candidate Extraction:

Secreted in the saliva, salivary amylase breaks down long-chain and branched carbohydrates into two- and three-molecule sugars called maltose. Once it enters the duodenum, the first portion of the small intestine, pancreatic amylase converts to its active form. ... Sucrase divides sucrose, more commonly known as table sugar, into its glucose and fructose components. ...

Candidate Scoring:

amylase (0.998)
amylase enzymes (0.925)
enzyme maltase (0.881)
salivary amylase (0.860)
protease enzymes (0.765)
protease (0.697)

lipase (0.665)
lipase enzymes (0.663)
sucrase (0.653)
mucin (0.633)
carbohydrase (0.525)
disaccharides (0.474)

Query Analysis: Find the Focus Words in a Query

- ◆ The focus words are words in a query that specify the type of the answer that the user is looking for.
- ◆ A query is answer-seeking if and only if such focus words are found

Focus Words: Examples

♦ Explicit questions

- what **enzyme** breaks down starch?
- 2014 ebola outbreak happened in what **african countries**
- **how much** does a macbook air **weigh**
- what is the **melting point** of paraffin wax
- **what** do hedgehogs **eat**

♦ Implicit questions

- fastest **animal**
- **annual rainfall** of beijing

♦ Non-question:

- when in rome rotten

Finding Focus Words in Explicit Questions

- ♦ Explicit questions have wh-words
- ♦ Pattern matching rules works well
 - Wh-word itself: when / where / who
 - The first noun phrase after which / what
 - what **enzyme** breaks down starch?
 - 2014 ebola outbreak happened in what **african countries**
 - what is the **melting point** of paraffin wax

Implicit Questions

- ✦ Pattern: (what | who) (is | was | are | were) the Attribute of Entity
 - [what is the fastest animal]
 - [who was the annual rainfall of beijing]
- ✦ Pattern: (what | who) (is | was | are | were) Entity's Attribute
 - [what is vietnam's population]

Search-based Question Answering

Query Analysis:

[what enzyme breaks down starch]

[what **enzyme** breaks down starch]

Candidate Extraction:

Secreted in the saliva, salivary amylase breaks down long-chain and branched carbohydrates into two- and three-molecule sugars called maltose. Once it enters the duodenum, the first portion of the small intestine, pancreatic amylase converts to its active form. ... Sucrase divides sucrose, more commonly known as table sugar, into its glucose and fructose components. ...

Candidate Scoring:

amylase (0.998)
amylase enzymes (0.925)
enzyme maltase (0.881)
salivary amylase (0.860)
protease enzymes (0.765)
protease (0.697)

lipase (0.665)
lipase enzymes (0.663)
sucrase (0.653)
mucin (0.633)
carbohydrase (0.525)
disaccharides (0.474)

Expected Answer Type

- ◆ Expected answer type of a question specifies what phrase / text can potentially be an answer (without knowing the context).
- ◆ The expected answer type of a question depends on
 - The focus words of the question
 - The ability of recognize whether something belongs to that type

Coarse-Grain Answer Types

- ♦ A small number of types
 - PERSON, LOCATION, TIME, TEMPERATURE, ...
- ♦ Advantage:
 - use a named entity recognizer to determine what phrase matches the type.
- ♦ Disadvantage:
 - mapping from the focus word in query to an answer type is non-trivial
 - melting point -> TEMPERATURE
 - annual rainfall -> LENGTH
 - assistant coach -> PERSON

Fine-Grain Answer Types

- ♦ IBM Watson
 - ~5000 types
- ♦ The most extreme version of fine-grain answer types:
 - Use as many answer types as focus words
 - Trivial mapping from focus words to answer types
 - Similar types are modeled separately
 - E.g., diameter and width are treated as two different types.

Extract Answer Type Examples with Patterns

♦ Numerical types:

- (has | with) (a | an) TYPE of _____
- Example:
 - has a DIAMETER of 13 cm

♦ Entity types:

- TYPE(s), such as _____
- Example:
 - cities such as Beijing

Answer Type Example: annual rainfall

- ✦ The Zatecka Basin, the driest area, has an annual rainfall of about 18 inches.
- ✦ More than 80% of the continent has an annual rainfall of less than 600 mm;
- ✦ Cuzco with a mean average temperature of 10.7°C, and highest average monthly 12.1, has an annual rainfall of 804 mm.
- ✦ along the western Ghats escarpment, has an annual rainfall of about 5600 mm which is towards the mid-range of rain stations in the area
- ✦ For example, the 'region comprising New Mexico, Arizona, Colorado, Utah, Nevada, and Wyoming has an annual rainfall of from eight to sixteen inches
- ✦ Durban has an annual rainfall of 1,009 millimeters.

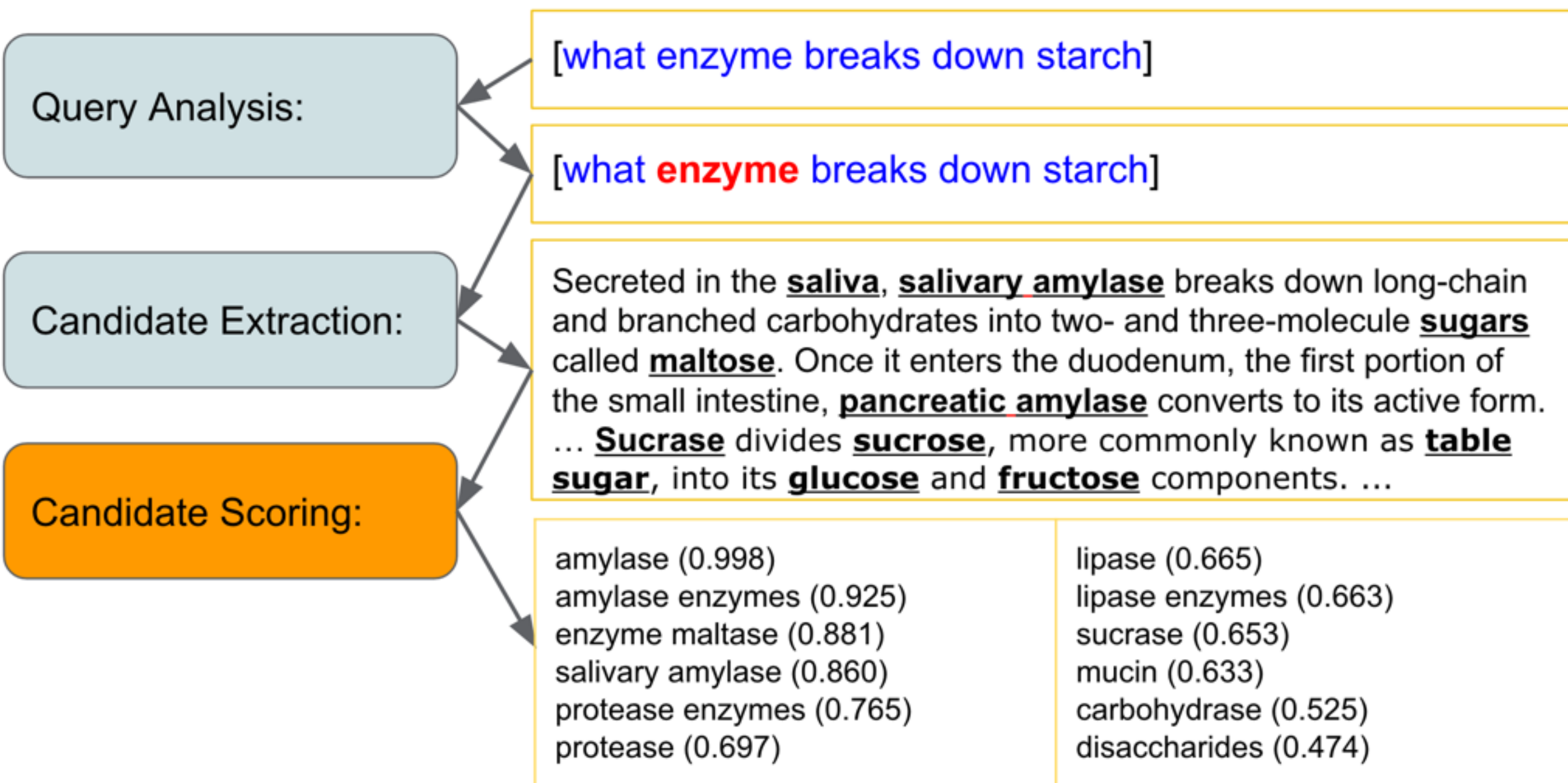
Answer Type Example: melting point

- ✦ Transcript of At room temperature, a substance with a melting point of 40 C.
- ✦ Unlike the trifluoride, gallium(III) chloride exists as dimeric molecules, Ga_2Cl_6 , with a melting point of 78°C.
- ✦ Hafnium nitride, with a melting point of 3300°C
- ✦ Pure Benzene (C_6H_6) has a melting point of 5.5 degrees C.
- ✦ This solid material has a melting point of approximately 106°C and a relative density (RD or SG) of approximately ...

Answer Type Example: sleeping pills

- ◆ There is little evidence that the newer generation **sleeping pills** such as **Ambien** or **Lunesta** are more effective than older **sleeping pills** such as **Dalmane** or ...
- ◆ Diphenhydramine is found in many popular over the counter **sleeping pills** such as **Tylenol PM**, **Excedrin PM** and **Nytol**.
- ◆ Sedative **sleeping pills** such as **Ambien** can nearly double the risk for car accidents among new users compared with nonusers
- ◆ A newer class of non-benzodiazepine **sleeping pills** such as **eszopiclone** or **zolpidem** appears to be safe to use by patients with OSA

Search-based Question Answering



Answer Candidate Scoring

♦ Features

- Answer type score
- Proximity to query keywords
- Popularity
-

Question Answering: Product Features

- ◆ Answer context
- ◆ Robustness
- ◆ Integration with other products
- ◆ Related questions
- ◆ Fun facts

Answer Context

- ✦ Automatic question answering system is not perfect
- ✦ Allow users to assess the credibility of answers
 - The source of the answer
 - The context in which the answer appears



what chemical kills bed bugs



Pyrethrins are botanical insecticides derived from chrysanthemum flowers. **Pyrethroids** are synthetic chemical insecticides that act like **pyrethrins**. Both compounds are lethal to bed bugs and can flush bed bugs out of their hiding places and kill them.

[Pesticides to Control Bed Bugs | Bed Bugs: Get Them Out ...](https://www.epa.gov/bedbugs)

[https://www.epa.gov/bedbu...](https://www.epa.gov/bedbugs) United States Environmental Protection Agency ▼



size of ping pong table

All

Shopping

Images

Videos

News

More ▼

Search tools

About 11,100,000 results (0.57 seconds)

274 cm

Table Tennis Rules. 1.1 The table shall be in surface rectangular, **274 cm.** (9 ft.) in length, **152.5 cm.** (5 ft.) in width. It shall be supported so that its upper surface, termed the playing surface, shall lie in a horizontal plane **76 cm.** (**2 ft. 6 in.**) above the floor.

[Official Ping Pong - Table Tennis Rules | Killerspin](http://www.killerspin.com/spin-info-table-tennis-rules/)

www.killerspin.com/spin-info-table-tennis-rules/

About this result • Feedback

Robustness

- ◆ Different phrasing of the same question should get the same result



what size is ping pong table



All

Shopping

Images

News

Videos

More ▾

Search tools

About 10,700,000 results (0.58 seconds)

Shop for what size is ping pong ta... on Google

Sponsored ⓘ



American
Heritage ...

\$314.03

Wayfair



MD Sports
4-Piece Ping-...

\$133.53

Walmart



Brunswick 8 Ft
Table Tennis ...

\$399.00

Leisure Fitne...



Pottery Barn
Pool Table ...

\$799.00

Pottery Barn



Kettler Indoor
Stockholm GT

\$559.99

Modell's Spor...

Table Tennis Rules. 1.1 The table shall be in surface rectangular, **274 cm.** (9 ft.) in length, **152.5 cm.** (5 ft.) in width. It shall be supported so that its upper surface, termed the playing surface, shall lie in a horizontal plane **76 cm.** (2 ft. 6 in.) above the floor.

Official Ping Pong - Table Tennis Rules | Killerspin

www.killerspin.com/spin-info-table-tennis-rules/

Integration with Other Search/Answer Products

- ◆ Image search
- ◆ Knowledge base answers



alcohol content of tsingtao beer



All

Shopping

Images

News

Maps

More ▼

Search tools

About 8,750 results (0.51 seconds)

Tsingtao Beer, a well-hopped standard pilsner of **4.7%** abv, is the flagship brew, accounting for most of the brewery's production.



Tsingtao Brewery - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/Tsingtao_Brewery Wikipedia ▼

About this result • Feedback



alcohol content of tsingtao beer



All

Shopping

Images

News

Maps

More ▼

Search tools

About 8,750 results (0.51 seconds)

Tsingtao Beer, a well-hopped standard pilsner of **4.7%** abv, is the flagship brew, accounting for most of the brewery's production.



Tsingtao Brewery - Wikipedia, the free encyclopedia

https://en.wikipedia.org/wiki/Tsingtao_Brewery Wikipedia ▼

About this result • Feedback

[GOP \(disambiguation\) - Wikipedia, the free encyclopedia](#)

[https://en.wikipedia.org/wiki/GOP_\(disambiguation\)](https://en.wikipedia.org/wiki/GOP_(disambiguation)) ▼ Wikipedia ▼

GOP, short for Grand Old Party, is a nickname for the Republican Party of the United States. **GOP** or **Gop** may also refer to: ...

[Republican Party \(United States\) - Wikipedia, the free ...](#)

[https://en.wikipedia.org/wiki/Republican_Party_\(United_States\)](https://en.wikipedia.org/wiki/Republican_Party_(United_States)) ▼ Wikipedia ▼

The Republican Party, commonly referred to as the **GOP** is one of the two major contemporary political parties in the United States, the other being its historic ...

People also ask

What is the difference between republican and democratic parties? ▼

Who is the leader of the Republican Party? ▼

Are conservative and Republican the same? ▼

Why was the republican party formed? ▼

[With or Without Ryan, A Convention Savior Unlikely for GOP ...](#)

www.nbcnews.com/.../or-without-ryan-convention-savior... ▼ NBCNews.com ▼

14 hours ago - With House Speaker Paul Ryan's "count me out" declaration Tuesday, the prospects of a "savior" candidate to emerge at the **GOP** convention ...

Fun Facts (I am feeling curious)

- ◆ When the google query is “fun facts” or “i am feeling curious”, the search results is an interesting question and its answer (shown after some delay).



fun facts



All

Images

News

Videos

Shopping

More ▼

Search tools

About 194,000,000 results (0.48 seconds)

Are foxes and wolves related?

Dogs belong to the taxonomic family Canidae (canines) which is divided into two tribes: those related to wolves (Canini) and those related to foxes (Vulpini). A couple of canine species lay outside these two tribes, but hyenas are not canines.

[14 of Your Dog's Wild Relatives | Mental Floss](http://m.mentalfloss.com/article.php?id=30959)

<http://m.mentalfloss.com/article.php?id=30959>

ASK ANOTHER QUESTION



i'm feeling curious



All

News

Shopping

Apps

Videos

More ▼

Search tools

About 54,900,000 results (0.39 seconds)

When was New York the capital of America?

The city and its surroundings came under English control in 1664. New York served as the capital of the United States from 1785 until 1790. It has been the country's largest city since 1790.

[New York City - Wikipedia, the free encyclopedia](#)

http://en.m.wikipedia.org/wiki/New_York_City

ASK ANOTHER QUESTION

Combining Structure and Unstructured Answers



[All](#) [Shopping](#) [Images](#) [Videos](#) [News](#) [More ▾](#) [Search tools](#)

Matt Damon / Movies

Most popular first ▾



The Martian
2015



Good Will Hunting
1997



Jason Bourne
2016



The Bourne Ultimatum
2007



The Bourne Identity
2002



The Departed
2006



Saving Private Ryan
1998



Elysium
2013



Green Zone
2010



The Great Wall
2016

Matt Damon filmography - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Matt_Damon_filmography ▾ Wikipedia ▾

He featured as a con man in Steven Soderbergh's Ocean's Trilogy (2001–07) and played the titular spy Jason Bourne in four of the five films in the Bourne series (2002–16). Damon co-wrote, co-edited and starred in the drama Gerry in 2002, which polarized film critics.

Matt Damon - IMDb
www.imdb.com/name/nm0000354/ ▾ IMDb ▾

Known For. **The Bourne Identity** Bourne. (2002) **Saving Private Ryan** Private Ryan. (1998) **Good Will Hunting** Will Hunting. (1997) **The Departed** Colin Sullivan. (2006)

[The Great Wall](#) · [Matt Damon/Bourne Sequel](#) · [The Zero Theorem](#) · [Awards](#)

Matt Damon
Actor





Matthew Paige "Matt" Damon is an American actor, film producer and screenwriter. He is ranked among Forbes magazine's most bankable stars and is one of the highest-grossing actors of all time. [Wikipedia](#)

Combining Structure and Unstructured Answers

Google 1980s pop singers

All Images Shopping Videos News More Search tools

Singers / 1980s

Pop

Madonna George Michael Prince Michael Jackson David Bowie Cyndi Lauper Tina Turner Whitney Houston Phil Collins

1958–2016 1958–2009 1947–2016

Best 80s Pop Artists | List of Top Pop Stars & Singers of the 1980s
www.ranker.com/crowdranked-list/best-pop-artists-of-the-1980s ▼
The 1980s was a huge decade in terms of breaking pop stars. ... Bon Jovi, Journey, and the late Whitney Houston are examples of '80s singers whose music ...

List of Billboard Hot 100 number-one singles of the 1980s - Wikipedia ...
https://en.wikipedia.org/.../List_of_Billboard_Hot_100_number-one_single... ▼ Wikipedia ▼
This article is about the American Billboard Hot 100 chart held during the 1980s. The Billboard
Note: That for Singer George Michael and Wham! is included, which gives George Michael 8 #1 Songs.
Note: That for Singer Phil Collins, ...
Number ones · Statistics by decade · See also · References

1980s in music - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/1980s_in_music ▼ Wikipedia ▼
Jump to **Pop** - American artists such as Whitney Houston, Michael Jackson, Madonna, Bruce

80517712799630968....jpg distantly_supervised_d....pdf tailKnowledge Luna Do....pdf fb2w.nt.gz 9140-39995-1-PB.pdf



origin of miniature schnauzer



All

Images

News

Shopping

Videos

More ▾

Search tools

About 113,000 results (0.97 seconds)

Miniature Schnauzer / Origin



Germany

Miniature Schnauzers were originally bred to be ratters and guard dogs on farms. They were developed in the mid-to-late 19th century in **Germany** by crossbreeding the Standard Schnauzer with smaller breeds, such as the Miniature Pinscher, Affenpinscher, and perhaps the Poodle or Pomeranian.

[Miniature Schnauzer Dog Breed Information, Pictures ...](https://dogtime.com/dog-breeds/miniature-schnauzer)
dogtime.com/dog-breeds/miniature-schnauzer Dogtime ▾



Destinations and points of interest

Miniature Schnauzers were originally bred to be ratters and guard dogs on farms. They were developed in the mid-to-late 19th century in **Germany** by crossbreeding the Standard Schnauzer with smaller breeds, such as the Miniature Pinscher, Affenpinscher, and perhaps the Poodle or Pomeranian.

[Miniature Schnauzer Dog Breed Information, Pictures ...](#)
dogtime.com/dog-breeds/miniature-schnauzer Dogtime ▾

Capital: [Berlin](#)

Dialing code: +49

Population: 80.62 million (2013) World Bank

Chancellor: [Angela Merkel](#)

President: [Joachim Gauck](#)

Destinations

[View 10+ more](#)



[Berlin](#)



[Munich](#)



[Frankfurt](#)



[Hamburg](#)



[Heidelberg](#)



[Dresden](#)

Points of interest

[View 40+ more](#)



[Brandenb...
Gate](#)



[Neuschwa...
Castle](#)



[Berlin Wall](#)



[Cologne
Cathedral](#)



[Reichstag
building](#)



[Europa-P...](#)

 Show less

Knowledge Base Population and Completion

♦ Motivation:

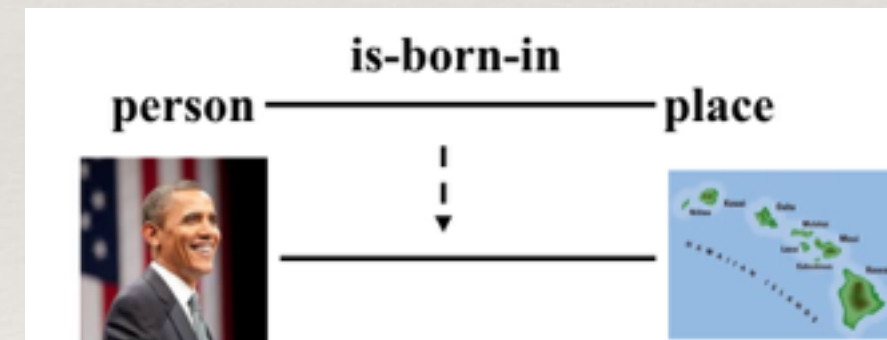
- Large-scale knowledge bases—e.g., Freebase , NELL , and YAGO are still woefully incomplete in many ways

Relation	Percentage unknown	
	<i>All 3M</i>	<i>Top 100K</i>
PROFESSION	68%	24%
PLACE OF BIRTH	71%	13%
NATIONALITY	75%	21%
EDUCATION	91%	63%
SPOUSES	92%	68%
PARENTS	94%	77%
CHILDREN	94%	80%
SIBLINGS	96%	83%
ETHNICITY	99%	86%

Table 1: Incompleteness of Freebase for some relations that apply to entities of type PERSON. *Left:* all 3M Freebase PERSON entities. *Right:* only the 100K most frequent PERSON entities.

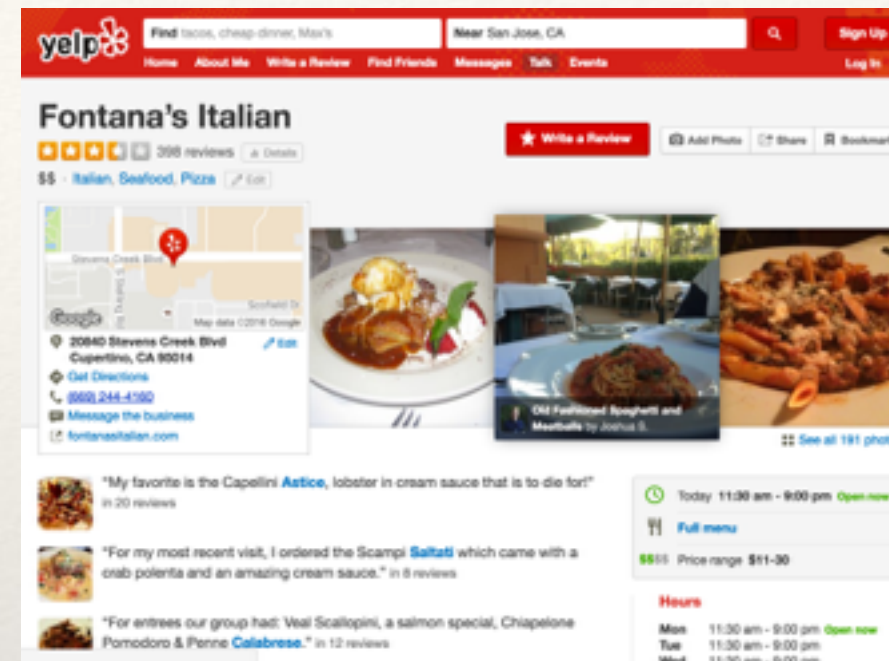
Turn Web into a Knowledge Base

- ◆ Taxonomic knowledge
- ◆ Relational knowledge
 - Information Extraction-based methods
 - Question Answering-based methods



KB Population by Information Extraction: Sources

- ◆ Unstructured text
- ◆ Semi-structured text
- ◆ Tables



BHUTAN	
Capital and largest city	Thimphu 27°28.0'N 89°38.5'E
Official languages	Dzongkha
Religion	Vajrayana Buddhism
Demonym	Bhutanese
Government	Unitary parliamentary constitutional monarchy
• King	Jigme Khesar Namgyel Wangchuck
• Prime Minister	Tshering Tobgay
Legislature	Parliament
• Upper house	National Council
• Lower house	National Assembly
Formation	Early 17th century
• House of Wangchuck	17 December 1907
• Indo-Bhutan Treaty	8 August 1949
• Constitutional monarchy	18 July 2008
Area	

Brand	ABV Alcohol % by Volume	Calories Per Serving	C.F.A. Calories From Alcohol Higher is better	Carbs Per Serving
Natural Ice	5.9%	130 calories per 12 oz	Score: A- 86.49%	4.00
 King Cobra	6%	134 calories per 12 oz	Score: A- 85.33%	4.30
Molson Canadian 67	3%	67 calories per 12 oz	Score: A- 85.33%	NA
 Pig's Eye Lean Light	4.2%	95 calories per 12 oz	Score: B+ 84.70%	2.60
 Sleeman Clear	4%	90 calories per 12 oz	Score: B+ 84.70%	2.50
 Michelob Ultra	4.2%	95 calories per 12 oz	Score: B+ 84.25%	2.60
				

Hearst Patterns in Chinese

◆ Patterns:

(像 | 在 | 将 | 及) NP_{inst} (一样的 | 这样的 | 这种) NP_{type} (, | 。 | , | 和)

- 即使是像姚晨这样的明星，也不能.....

(、 | , | 和) NP_{inst} (等 | 等等) NP_{type} (, | 。 | , | 和)

- 临靠上海，苏州等城市。

Example Type: 食品 (food)

水果 2254, 西瓜 1650, 鱼 1553, 饼干 1511, 蛋 1419, 蔬菜 1315, 玉米 1180, 牛奶 1127, 芝麻 1126, 黑芝麻 1124, 肉 1068, 面包 1065, 鸡蛋 1033, 瘦肉 1026, 花生 976, 方便面 926, 饮料 899, 巧克力 884, 豆类 852, 肉类 836, 糕点 833, 豆制品 815, 糖果 797, 海鲜 787, 老婆饼 778, 辛辣 746, 杏仁饼 725, 羊肉 722, 酒类 656, 蛋类 603, 蜂蜜 601, 咖啡 601, 海带 584, 罐头 578, 蛋糕 547, 小吃 546, 木耳 534, 核桃 530, 紫菜 524, 过咸 484, 干粮 473, 虾 462, 萝卜 459, 甘蔗 458, 冰冻 451, 黑木耳 416, 果汁 413, 牛肉 408, 香菇 406, 冰淇淋 389, 梨 387, 鱼类 385, 蛋黄 383, 苹果 372, 荸荠 361, 瓜子 361, 番茄 360, 矿泉水 357, 油 353, 牛羊肉 345, 生梨 337, 月饼 335, 白茅根 334, 韭菜 329, 菠菜 325, 奶 317, 香肠 314, 豆腐 309, 香蕉 295, 面条 289, 果冻 289, 煎饼果子 283, 馒头 282, 奶酪 278, 肉夹馍 277, 山楂 275, 鸡 271, 果酱 270, 火腿肠 268, 酒 267,

食品

Rank	Inst	Freq	Rank	Inst	Freq
1	水果	2254	2201	泡椒	3
201	麻辣烫	102	2401	马哈鱼	2
401	酥油茶	42	2601	肉铺	2
601	李子	25	2801	水豆腐	2
801	橙子	16	3001	全麦粉	2
1001	白薯	11	3201	青油	1
1201	调味粉	8	3401	蚝	1
1401	通心粉	6	3601	细面	1
1601	燻烤	5	3801	猪内脏	1
1801	維力	4	4001	沙冰	1
2001	面鱼	3	4201	应急灯	1
			4401	叉燒包	1

Example Type: 品牌 (brand)

三星 1124, 松下 559, 美的 443, 联想 441, 海尔 344, 摩托罗拉 342, 索尼 338, 多普达 321, 东芝 272, 佳能 238, 华硕 233, 奥迪 228, 丰田 225, 宝马 221, 海信 216, 雪佛兰 201, 李宁 198, 惠普 197, 长城 188, 涉 179, 三菱 179, 西门子 170, 现代 168, 耐克 164, 雀巢 161, 日立 160, 夏普 160, 索爱 159, 志高 156, 诺基亚 155, 富士 150, 大众 150, 戴尔 143, 兰蔻 142, 奔驰 141, 阿迪达斯 139, 本田 138, 碧欧泉 136, 格力 135, 方正 135, 长虹 134, 尼康 134, 别克 133, 三洋 132, 飞利浦 130, 比亚迪 122, 康佳 121, 富士通 121, 九阳 121, 欧莱雅 119, 福特 118, 化妆品 117, 苹果 113, 凯迪拉克 113, 香奈儿 111, 兰博基尼 110, 沃尔沃 109, 中兴 109, 鲁花 108, 百丽 107, 安踏 106, 浪琴 104, 达利园 102, 创维 102, 爱普生 101, 日产 100, 明基 99, 奇瑞 96, 宏基 94, 奥克斯 93, 施耐德 91, 保时捷 91, 彪马 90, 华为 90, 上海大众 89, 雷克萨斯 87, ...

品牌

Rank	Inst	Freq	Rank	Inst	Freq
1	三星	1124	4201	小叮当	2
301	神州行	31	4501	健身	2
601	百雀羚	17	4801	香玲	1
901	舒龙	11	5101	郑记	1
1201	全顺	9	5401	英国保罗	1
1501	安安	7	5701	筒射灯	1
1801	鹏程	5	6001	特种旅游	1
2101	三木	5	6301	水景	1
2401	新乐	4	6601	新奇	1
2701	银河	3	6901	山田	1
3001	曼联	3	7201	天创	1
3301	五华	3	7501	卡利亚里	1
3601	蓝带啤酒	2	7801	优乐	1
3901	润洁	2	8101	yamato	1

Errors in Hearst Pattern Outputs

♦等 means *etc.* instead of *such as*:

- 用于摔跤、举重等运动员。
- 晚会还请来罗大佑、赵薇、小沈阳、张靓颖等现场演出，

♦Frequent words and segmentation errors: e.g., in 品牌

Rank	Inst	Freq	Rank	Inst	Freq
136	等	56	1078	医院	10
666	中	16	1178	天	9
760	同	14	1230	香港	8
851	日	12	1289	标准	8
909	美	11	1313	建设	8
921	文化	11	1345	北京	8
1031	服务	10	1481	成功	7

Filtering Extracted Instances with PMI

PMI: Point-wise Mutual Information

$$PMI(inst, context) = \log \frac{P(inst|context)}{P(inst)}$$

$P(inst | context)$: how likely does $inst$ appear in $context$

$$P(inst|context) = \frac{N(inst, context)}{\sum_{inst'} N(inst', context)}$$

$P(inst)$: how likely does $inst$ appear in any context

$$P(inst) = \frac{N(inst)}{\sum_{ngram} N(ngram)}$$

Filtering out Errors in 品牌 with PMI

Rank	Inst	Freq	PMI
136	等	56	-0.11
666	中	16	-2.71
760	同	14	-1.69
851	日	12	-1.98
909	美	11	-1.84
921	文化	11	-0.10
1031	服务	10	-0.85
1078	医院	10	-0.13
1178	天	9	-2.49
1230	香港	8	-0.08
1289	标准	8	-0.20
1313	建设	8	-0.52
1345	北京	8	-0.89
1481	成功	7	-0.09

品牌

Rank	Inst	Freq	PMI	Rank	Inst	Freq	PMI
1	三星	1124	6.57559	4201	小叮当	2	4.34962
301	神州行	31	6.35509	4501	健身	2	-0.055421
601	百雀羚	17	7.26131	4801	香玲	1	3.58952
901	舒龙	11	10.187	5101	郑记	1	4.61972
1201	全顺	9	5.77639	5401	英国保罗	1	4.64069
1501	安安	7	3.4437	5701	筒射灯	1	6.50914
1801	鹏程	5	4.4682	6001	特种旅游	1	3.88327
2101	三木	5	5.14057	6301	水景	1	0.566763
2401	新乐	4	4.27255	6601	新奇	1	-0.815131
2701	银河	3	2.00848	6901	山田	1	0.651701
3001	曼联	3	2.41141	7201	天创	1	1.53205
3301	五华	3	3.83703	7501	卡利亚里	1	3.07658
3601	蓝带啤酒	2	6.74887	7801	优乐	1	1.66726
3901	润洁	2	5.0417	8101	yamato	1	3.62627

Extracting Chinese Verbs

◆Pattern:

(你|你们|他|他们|她|她们|我|我们)____(了什么|了谁|不了什么|着什么|着谁)

正在____的时候

◆Extracted Instances:

做 7513, 说 5595, 活 4534, 执 2939, 爱 2662, 看 2511, 睡 2457, 穿 2393, 带 2352, 吃 2119, 用 2112, 死 2016, 抱 1981, 干 1733, 坐 1676, 忘 1556, 买 1520, 拿 1505, 走 1395, 有 1189, 得 1064, 看到 1040, 站 1039, 深爱 959, 犯 944, 想 919, 杀 875, 开 820, 写 777, 握 767, 去 754, 出 750, 老 738, 下 737, 打 722, 背 716, 笑 714, 留 704, 想到 674, 躺 668, 跟 664, 醒 624, 发现 614, 输 610, 错 605, 悬 570, 牵 567, 提 562, 想起 557, 发 557, 丢 557, 戴 550, 装 545, 知道 541, 哭 515, 抓 509, 赢 502, 做错 494, 讲 464, 指 464, 累 460, 拉 459, 得到 442, 听 440, 身 439, 看见 439, 学 417, 见 415, 受 411, 等 405, 选 403, 选择 401, 盯 397, 删 382, 沉 379, 忘记 379, 失去 379, 遇到 378, 忽略 372, 饿 369, 遇见 364, 挂 361, 给 355, 找 352, 久违 345, 空 336, 怀 335, 靠 334, 点 332, 找到 332, 错过 322, 放 318, 问 317, 明白 317, 懂 317, 习惯 316, ...

Extracted Verbs

Rank	Verb	Freq	PMI
1	做	7513	3.1869
201	凭	135	1.96404
401	燃烧	61	2.44851
601	打死	37	2.87521
801	冻僵	27	5.88245
1001	驾驶	20	0.6504
1201	疏漏	16	4.33751
1401	算出	13	1.59965
1601	玷污	11	3.64278
1801	说明白	9	3.44448
2001	捉弄	8	3.04751

Rank	Verb	Freq	PMI
2201	指点	7	0.321181
2401	汇款	6	1.07645
2601	表述	5	0.863935
2801	吮吸	5	2.9011
3001	私奔	4	1.97992
3201	学精	4	1.44567
3401	选修	3	0.600118
3601	痊愈	3	0.698633
3801	执著	3	1.1455
4001	厮守	3	2.66336

Filtered Out by PMI

Rank	Word	Freq	PMI
75	中	385	-1.48024
87	上	352	-1.41482
119	生	272	-1.41914
141	动	230	-1.01046
142	好	229	-1.34366
189	所	159	-1.22409
211	信	144	-1.09295
222	不	138	-2.89807
238	管	129	-1.10651
260	定	121	-1.56103
270	当	113	-1.15121
285	进	107	-1.52745
295	交	103	-1.01106
310	安	96	-1.42013
315	回	95	-1.36784
319	关	94	-1.73741

Extracting Chinese Adjectives

◆Pattern

(非常|很|极为)____的

◆Extracted adjectives

好 1059676, 多 701219, 大 689071, 高 315613, 强 239468, 快 217135, 小 175076, 棒 124127, 重要 118283, 不错 117132, 久 111041, 少 88682, 深 82579, 长 73855, 简单 68313, 低 65693, 漂亮 64153, 明显 54999, 美 50661, 多人 48936, 重 47562, 自然 46610, 有名 39456, 喜欢 38670, 方便 38061, 远 37327, 普通 35675, 正常 35393, 强 34975, 短 34751, 近 34728, 容易 34542, 有趣 34462, 奇怪 33616, 差 33410, 可爱 33356, 好看 32527, 好用 31207, 难 30105, 酷 30074, 强大 28817, 贵 28036, 厉害 27562, 清楚 27393, 熟悉 27371, 严重 27189, 流行 25512, 老 24917, 认真 24838, 有用 24611, 经典 24540, 努力 24221, 实用 23706, 特别 23056, 明显 22961, 便宜 22866, 轻松 22767, 不错 22437, 舒服 21630, 火 21042, 特别 20364, 累 20261, 复杂 19857, 好玩 19590, 薄 19551, 有效 19454, 牛 19052, 厚 19005, 浓 18909, 慢 18765, 好吃 18518, 开心 18502, 优秀 18373, 常见 18366, 长 18038, 可爱 17982, 帅 17795,, ...

Extracted Adjectives

Rank	Word	Freq	PMI
1	好	1059676	3.43374
501	另类	3274	3.60151
1001	固执	1453	3.81543
1501	主动	841	2.31612
2001	狗腿	551	4.6551
2501	低俗	384	3.02372
3001	鬱悶	281	4.06296
3501	風趣	205	4.0695
4001	确信	157	1.85926
4501	擔憂	121	2.09594
5001	高心	94	2.76404
5501	粗狂	74	3.94664
6001	有責任感	58	3.6874
6501	多阅读	46	3.1972
7001	靠杯	36	3.85453
7501	發燒	29	0.237939

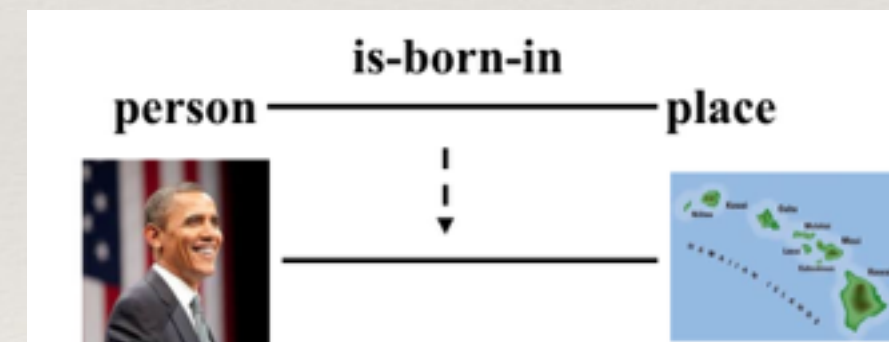
Rank	Word	Freq	PMI
8001	大作为	24	2.03742
8501	多条腿	20	4.38836
9001	豐裕	16	3.31148
9501	信誓旦旦	14	0.639124
10001	贪嘴	11	2.71323
10501	鬼馬	9	2.20363
11001	惶惑	8	1.12299
11501	哈韓	7	1.91163
12001	吸味	6	2.10525
12501	属世	5	0.740774
13001	玩闹	4	0.0130895
13501	个性独特	4	1.50635
14001	易燃烧	3	0.684135
14501	spicy	3	0.989788

Filtered out by PMI

Rank	Word	Freq	PMI
234	是	6330	-2.94521
665	真	2477	-1.62784
696	对	2366	-2.49729
733	直	2191	-1.43322
794	0	1996	-3.36792
840	水	1868	-2.25281
932	系统	1641	-1.27762
935	有	1635	-4.06556
974	配	1583	-1.40165
981	平	1569	-2.09849
1027	密	1471	-1.03327
1045	人	1430	-4.00915
1065	二	1394	-2.3545
1122	想	1300	-2.57462
1127	开	1286	-2.86118
1155	经济	1247	-1.16509

Turn Web into a Knowledge Base

- ◆ Taxonomic knowledge
- ◆ Relational knowledge
 - Information Extraction-based methods
 - Question Answering-based methods



KB Population by Information Extraction: Steps

- ◆ Triple extraction
 - identify (Subject, Predicate, Object) triples
- ◆ Entity linking
 - Map the Subjects and Objects to entities in KBs
- ◆ Predicate linking
 - Map the predicates to relations in KBs

KB Population: Triple Extraction

♦ Example

- Emma Thompson graduated from Cambridge with a B.A. in English literature, and soon after won an Academy Award for screenwriting.



- (Emma_Thompson, graduated_from, Cambridge)
- (Emma_Thompson, graduated_with, B.A._in_English_literature)

KB Population: Entity Linking

♦ Examples:

- Emma_Thompson → /m/0159h6
- Cambridge → /m/378a2
- B.A._in_English_literature → /m/28df71

♦ Method:

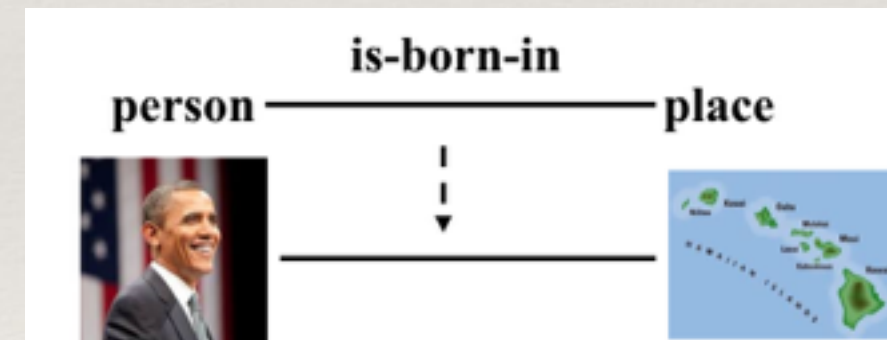
- Disambiguation to Wikipedia

KB Population: Predicate Linking

- ♦ A predicate in text often does not correspond directly to a relation in a knowledge base.
- ♦ Triple: (Person, graduated_from, University)
- ♦ KB Schema:
 - (UniversityAttendance, student, Person)
 - (UniversityAttendance, school, University)
 - (UniversityAttendance, from, Time)
 - (UniversityAttendance, degree, Degree)
 -

Turn Web into a Knowledge Base

- ◆ Taxonomic knowledge
- ◆ Relational knowledge
 - Information Extraction-based methods
 - Question Answering-based methods



Knowledge Base Completion by Question Answering

- ◆ Input: (Subject, Predicate) pairs
- ◆ Output: previously unknown objects
- ◆ Example:

(/m/02whj FRANK ZAPPA, /m/01x3gb5 PAERENTS) →

(/m/01xxvkq Francis Zappa, /m/01xxvky Rose Marie Colimore)

KB Completion via Web-based Question Answering

♦ Steps:

- Obtain query templates for (subject, relation) via distant supervision.
- User query augmentation to come up with better queries
- Run Web-QA with augmented queries and obtain objects

♦ Advantages:

- Leverage web search to obtain reliable, fresh sources of information.
- Collect more facts users are looking for.

Query Template Generation

- ◆ Collect single entity queries (the entity is resolved using both the query and the search result).
 - who is the brother of Ben Affleck (/m/0151w_)
 - where was Noam Chomsky born
- ◆ Use a web-QA system to find the answer and resolve the entity reference
 - Casey Affleck (/m/032_jg)
 - Philadelphia (/m/01wz6kq)

Query Template Generation (continued)

- ♦ Find the relation between the query entity and the answer entity in Freebase
 - (/m/0151w_ /person/sibling, /m/032_jg)
- ♦ Collect relation template pairs:
 - (/person/sibling, “who is the brother of _____”)

Query Augmentation

♦ Motivation:

- use co-located facts to help identify passages

e.g., Frank Zapper mother Baltimore

_____ mother PLACE_OF_BIRTH

- resolve ambiguous terms

e.g., Michael Jackson place of birth World Guide to Beer

_____ place of birth WRITTEN_WORK

Example: birthplace of Michael Jackson (author)

Subject–relation pair:

(MICHAEL JACKSON (WRITER), PLACE OF BIRTH)

True answer: LEEDS

Template: (*birthplace of* __, WORKS WRITTEN)

Query: *birthplace of Michael Jackson World Guide to Beer*

Example: birthplace of Michael Jackson (author)

♦ Without augmentation:

[1] *Michael Jackson – Wikipedia, the free encyclopedia*

Michael Jackson was born on August 29, 1958, in **Gary, Indiana**. He was the eighth of ten children in an African-American working-class family [...] in Gary, an industrial city near **Chicago**.

[2] *Michael Jackson's House – Gary, IN – Yelp*

8 Reviews of Michael Jackson's House "WTF. this place is kind of a bummer. Streets aren't labeled, potholes aren't filled. The house is the only place on the block that was properly painted."

♦ With augmentation:

[1] *Michael Jackson (writer) – Wikipedia, the free encyclopedia*

Jackson was born in **Leeds, West Yorkshire** and spent his early years in nearby **Wetherby**. [...] Jackson, Michael (1977). *The World Guide to Beer* [...]

[2] *The Unique Michael Jackson | Philly Beer Scene*

He was born in **Wetherby** in the city of **Leeds**. [...] Compensation eventually was awarded in 1988, when his agent [...] negotiated a fee on the re-write as *The New World Guide to Beer*.

Summary

- ♦ Answering questions with knowledge graph
- ♦ Enhancing knowledge graph with question answering