PURDUE UNIVERSITY

CS54701: Information Retrieval

Question Answering

5 April, 2016

Prof. Chris Clifton

Web Search in 2020?

The web, it is a changing.

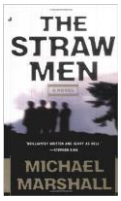
What will people do in 2020?

- Type key words into a search box?
- Use the Semantic Web?
- Speak to your computer with natural language search?
- Use social or “human powered” search?

Getting information

The common person's view? [From a novel]

> "I like the Internet. Really, I do. Any time I need a piece of shareware or I want to find out the weather in Bogota … I'm the first guy to get the modem humming. But as a source of information, it sucks. You got a billion pieces of data, struggling to be heard and seen and downloaded, and anything I want to know seems to get trampled underfoot in the crowd."



Michael Marshall. *The Straw Men.* HarperCollins,

"Information retrieval"

- 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

The future?

- Some scattered ideas from the corporate world
  Courtesy Chris Manning

Google (2014-2015) What's been happening?

- "Mobilegeddon" (Apr 21, 2015):
  - "Mobile friendliness" as a major ranking signal
- "Pigeon" update (July 2014):
  - More use of distance and location in ranking signals
- "App Indexing" (Android, iOS support May 2015)
  - Search results can take you to an app
- Why?
  - About half of all searches are now from mobile
  - Making/wanting good changes, but obvious self-interest in trying to keep people using mobile web rather than apps

Google
What's been happening? 2014

- New search index at Google: "Hummingbird"
  - http://www.forbes.com/sites/roberthof/2013/09/26/google-just-revamped-search-to-handle-your-long-questions/
- Answering long, "natural language" questions better
  - Partly to deal with spoken queries on mobile
- More use of the Google Knowledge Graph
  - Concepts versus words

What's been happening

- Google Knowledge Graph
- Facebook Graph Search
- Bing's Satori
- Things like Wolfram Alpha

Common theme: Doing graph search over structured knowledge rather than traditional text search

What's been happening

- More semi-structured information embedded in web pages
  - schema.org


```
275     <p class="byline-column">
276                 On Tennis               </p>
277     <p class="byline">By                    <span itemprop="author creator" itemscope
itemtype="http://schema.org/Person"
itemid="http://topics.nytimes.com/top/reference/timestopics/people/c/christopher_clarey/index.html">
278             <a
href="http://topics.nytimes.com/top/reference/timestopics/people/c/christopher_clarey/index.html" rel="author"
title="More Articles by CHRISTOPHER CLAREY">                    <span class="byline-author" data-byline-
name="Christopher Clarey" itemprop="name">CHRISTOPHER CLAREY</span>
279             </a>            </span>
280         </p>
```

What's been happening

- Move to mobile favors a move to speech which favors "natural language information search"
  - Will we move to a time when over half of searches are spoken?

Towards intelligent agents

Two goals

- Things not strings
- Inference not search

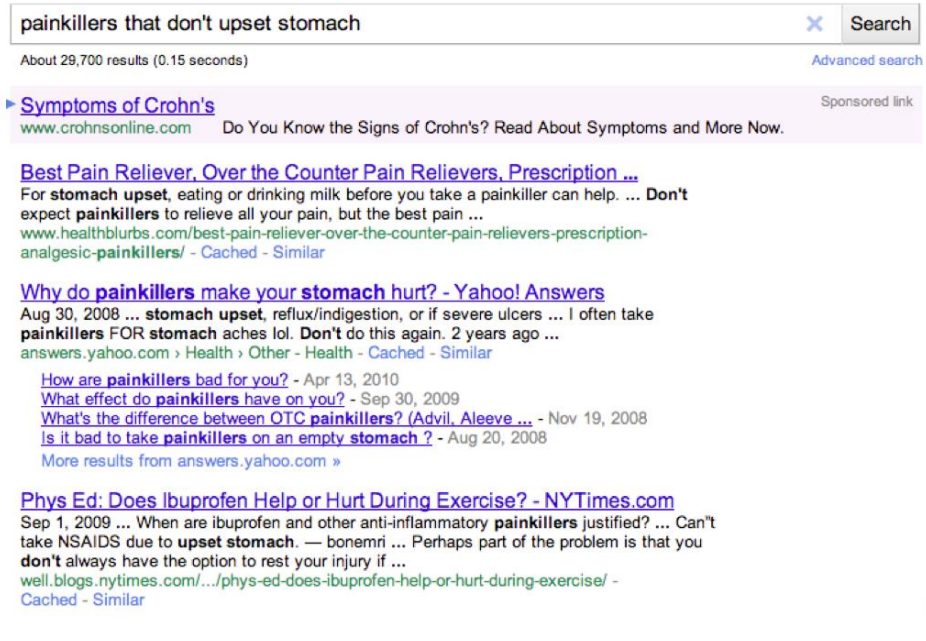
Two paradigms for question answering

- Text-based approaches
  - TREC QA, IBM Watson
- Structured knowledge-based approaches
  - Apple Siri, Wolfram Alpha, Facebook Graph Search

(And, of course, there are hybrids, including some of the above.)

At the moment, structured knowledge is back in fashion, but it may or may not last

Example from Fernando Pereira (GOOG)



class negation synonyms

painkillers that don't upset stomach Search

About 29,700 results (0.15 seconds) Advanced search

Sponsored link

Symptoms of Crohn's
www.crohnsonline.com Do You Know the Signs of Crohn's? Read About Symptoms and More Now.

Best Pain Reliever, Over the Counter Pain Relievers, Prescription ...
For **stomach upset**, eating or drinking milk before you take a painkiller can help. ... **Don't** expect **painkillers** to relieve all your pain, but the best pain ...
www.healthblurbs.com/best-pain-reliever-over-the-counter-pain-relievers-prescription-analgesic-**painkillers**/ - Cached - Similar

Why do **painkillers** make your **stomach** hurt? - Yahoo! Answers
Aug 30, 2008 ... **stomach upset**, reflux/indigestion, or if severe ulcers ... I often take **painkillers** FOR **stomach** aches lol. **Don't** do this again. 2 years ago ...
answers.yahoo.com › Health › Other - Health - Cached - Similar

How are **painkillers** bad for you? - Apr 13, 2010
What effect do **painkillers** have on you? - Sep 30, 2009
What's the difference between OTC **painkillers**? (Advil, Aleeve ... - Nov 19, 2008
Is it bad to take **painkillers** on an empty **stomach** ? - Aug 20, 2008
More results from answers.yahoo.com »

Phys Ed: Does Ibuprofen Help or Hurt During Exercise? - NYTimes.com
Sep 1, 2009 ... When are ibuprofen and other anti-inflammatory **painkillers** justified? ... Can't take NSAIDS due to **upset stomach**. — bonemri ... Perhaps part of the problem is that you **don't** always have the option to rest your injury if ...
well.blogs.nytimes.com/.../phys-ed-does-ibuprofen-help-or-hurt-during-exercise/ - Cached - Similar

Related concepts may not match question

"painkillers such as * tylenol" Search

About 19,200 results (0.21 seconds) Advanced search

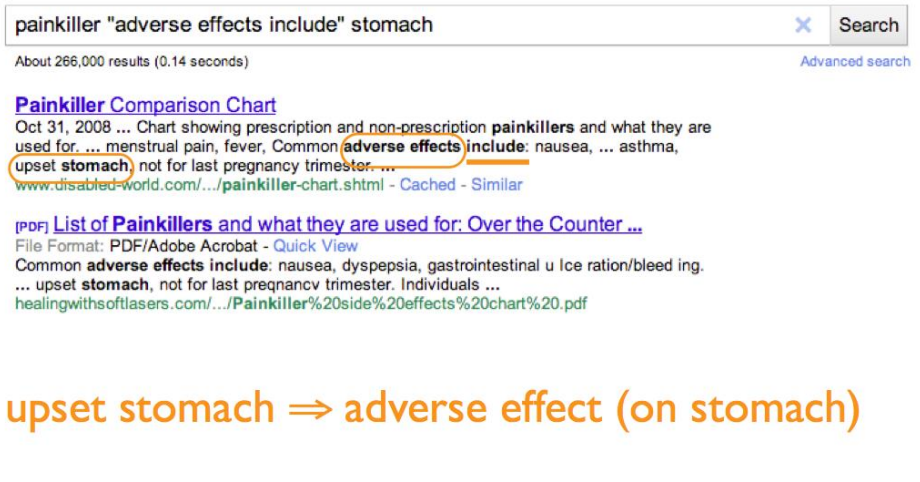
Sponsored link

TYLENOL ® Official Site
Tylenol.com Pain Relief from TYLENOL. Get Information and Coupons Here.
Head & Body Pain - Arthritis Pain - Sinus - Cold & Flu

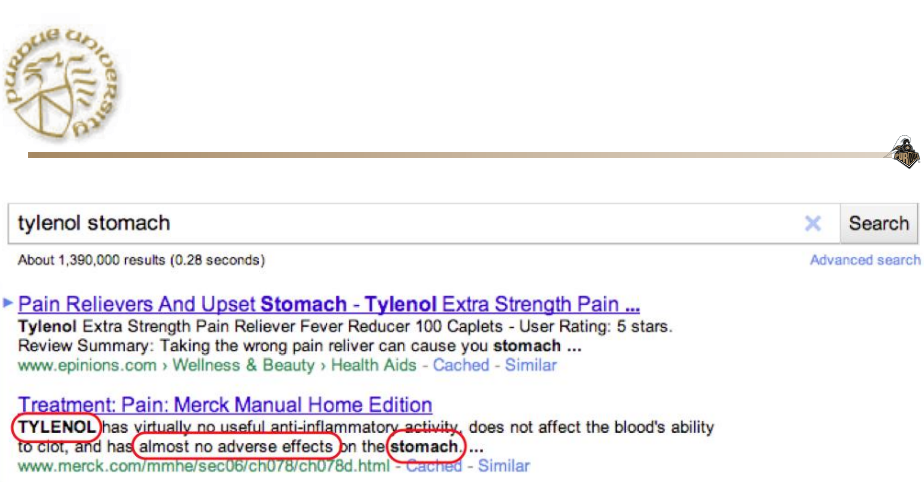
Ovarian hyperstimulation syndrome (OHSS): Treatments and drugs ...
To ease abdominal discomfort, take over-the-counter **painkillers such as acetaminophen** (**Tylenol**, others) or ibuprofen (Advil, Motrin, others). ...
www.mayoclinic.com/.../DSECTION=treatments-and-drugs - Cached - Similar

tylenol ⇒ painkiller

Text can identify synonyms



upset stomach ⇒ adverse effect (on stomach)

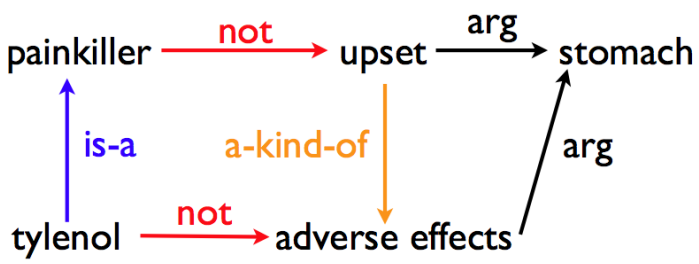


tylenol ⇒ no adverse effects (on stomach)
⇒ no upset stomach

This took us from Query

painkiller upset stomach

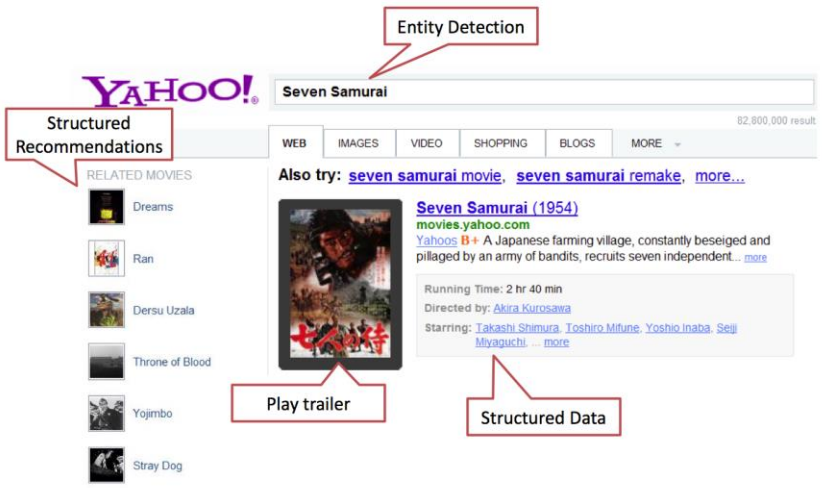
To Answer

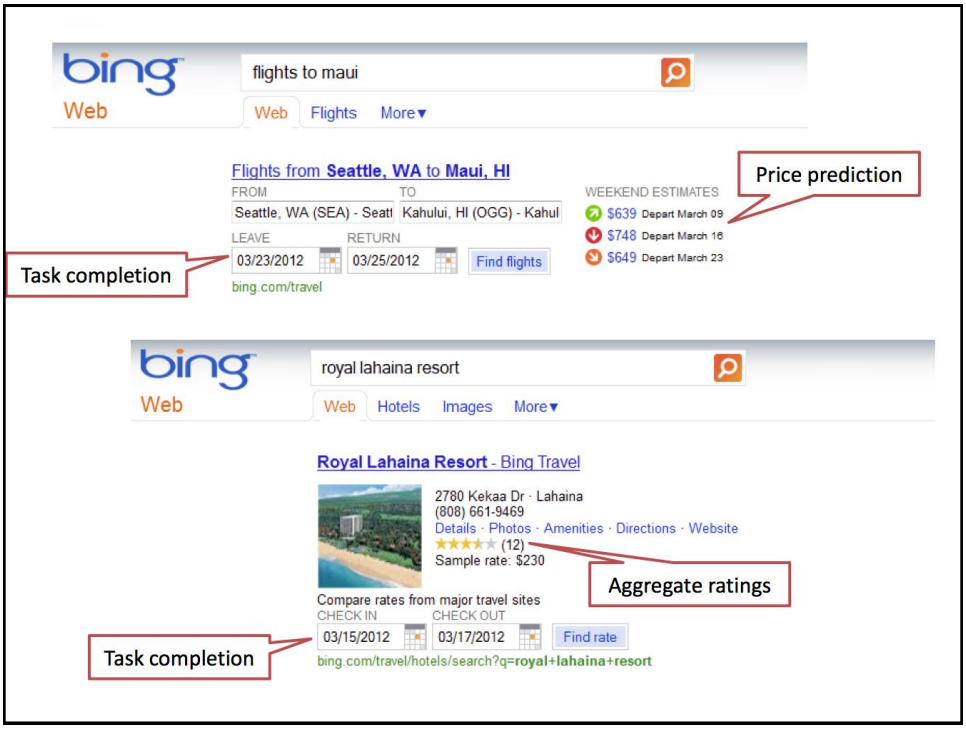


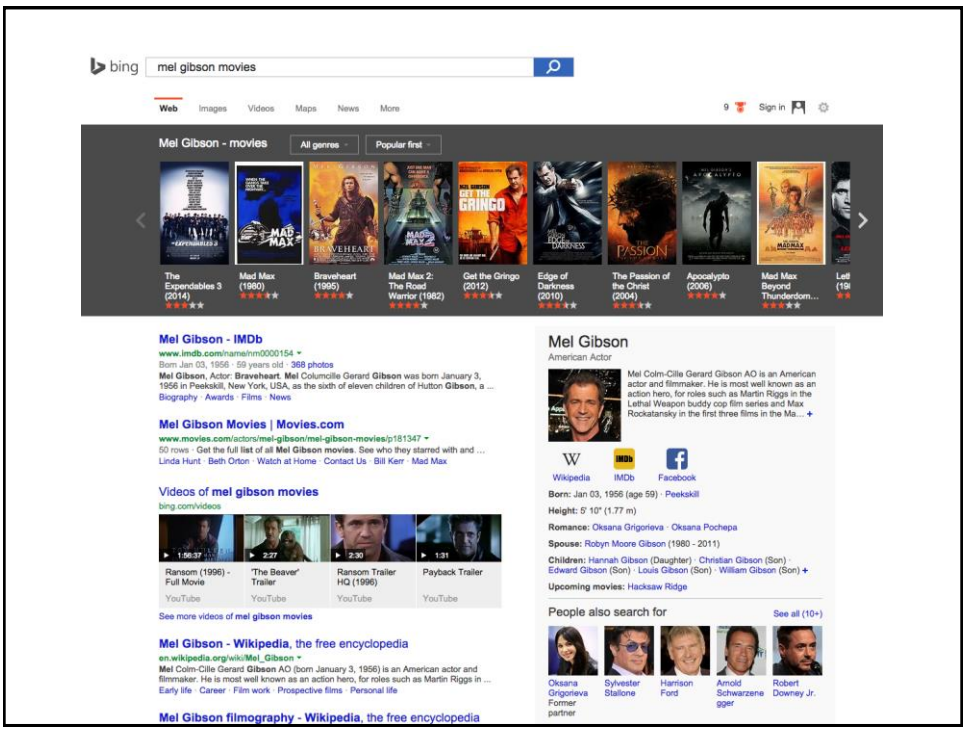
"Things, not strings"

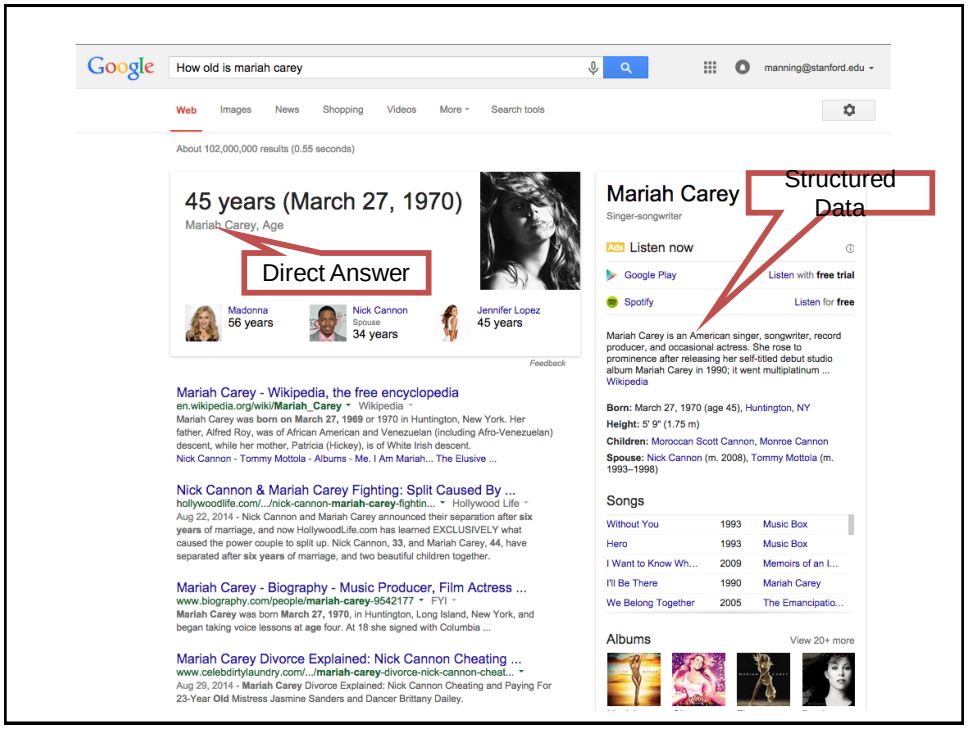
From	To	Requires
Term	Concept	Parsing, disambiguation, coreference
Term identity	Entailment	Concept relations
Co-occurrence	Syntactic relation	Document structure, parsing
Term index	Semantic index	Concept disambiguation, inference

Slides from Patrick Pantel (MSFT)

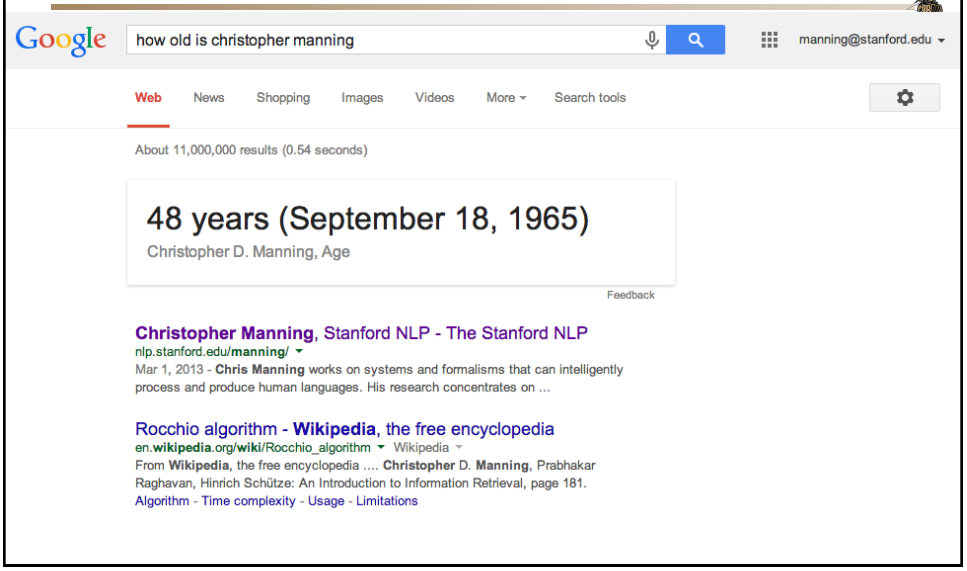




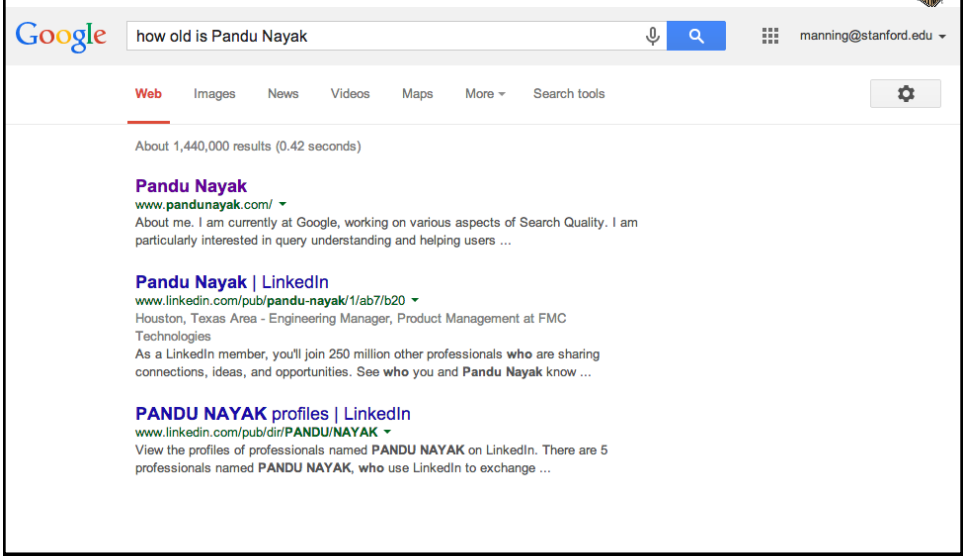




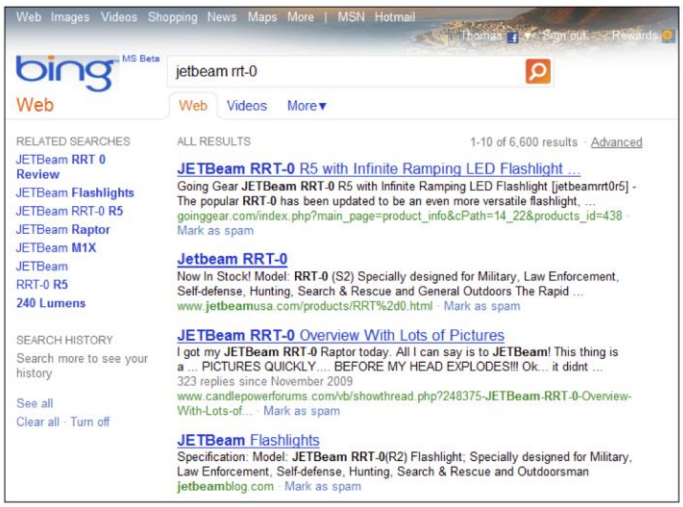
Query Answering



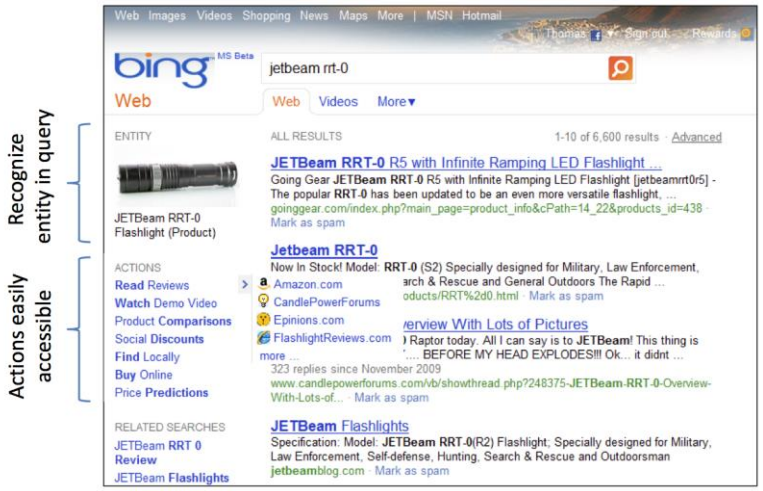
But not always…



Patrick Pantel talk
(Then) Current experience

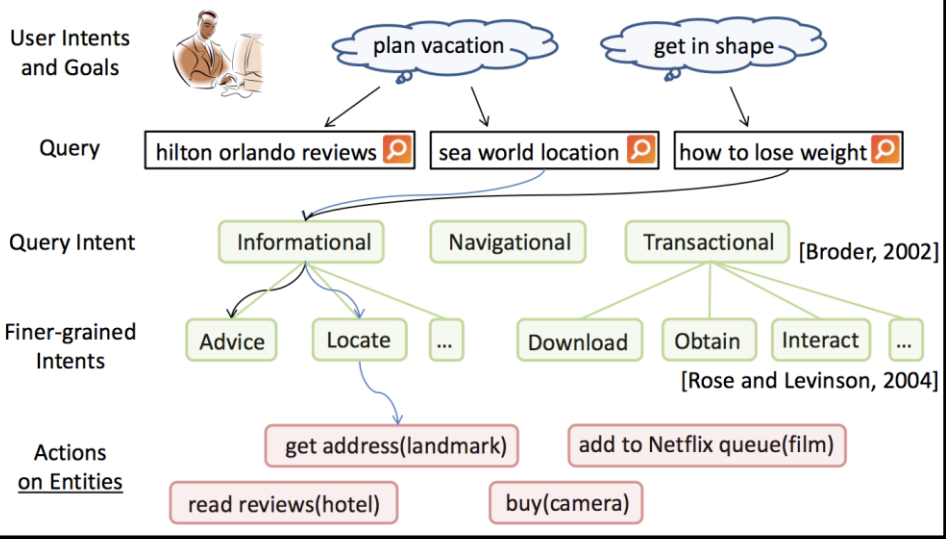


Desired experience: Towards actions



Politician

Actions vs. Intents



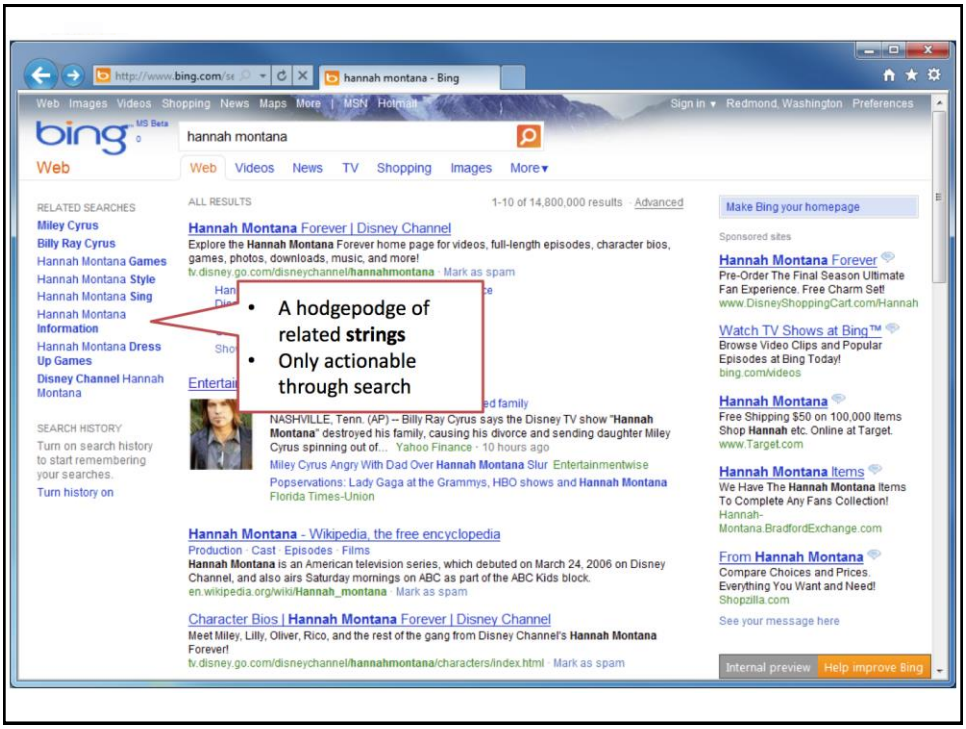
Learning actions from web usage logs

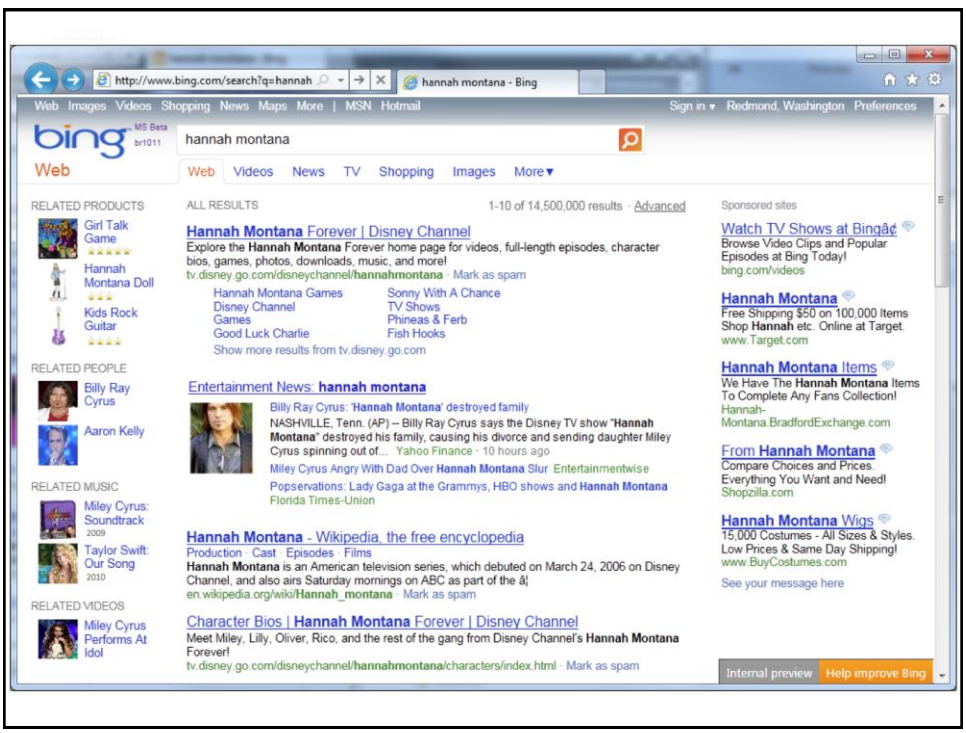
- Three months of us-en web logs
- Annotate with Freebase entities

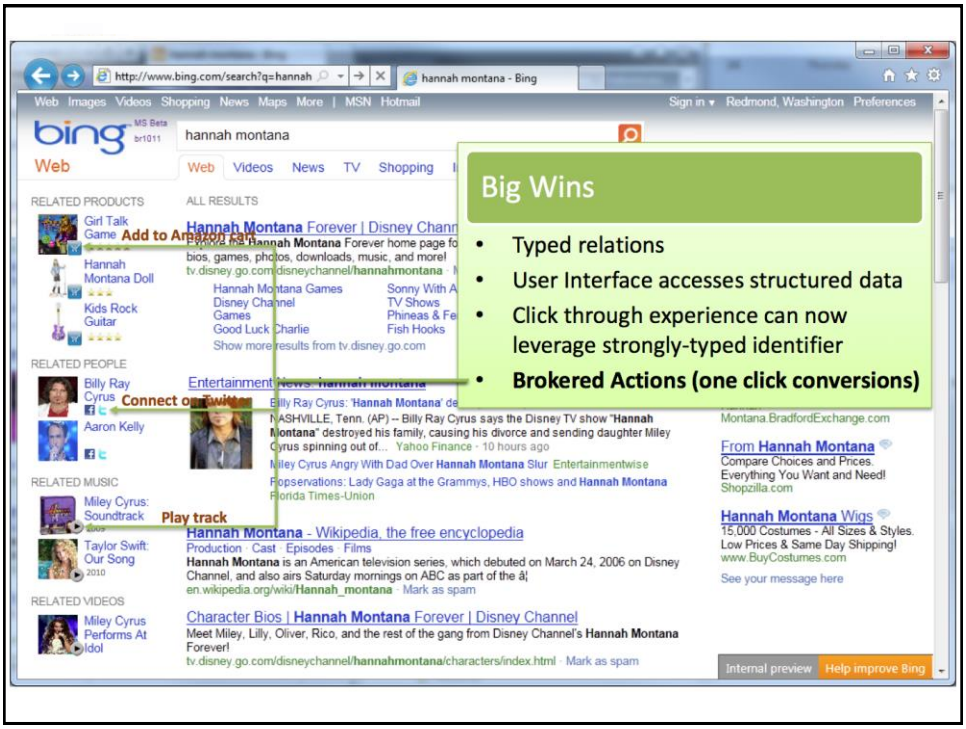
2,164,579 (query, host) pairs over 3 months	235,385 entities	21 types
	129,088 contexts	58,123 hosts

- Keep queries with an entity in set of 21 types
- Filter out navigational queries
- Filter out clicked hosts that weren't clicked at least 100 times

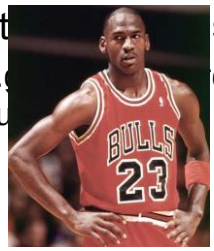
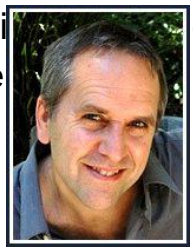
get reviews	←	Orlando hotel reviews
read biography		Does Hope Solo have a boyfriend?
download software		Free Winzip download
watch shows online		watch family guy online



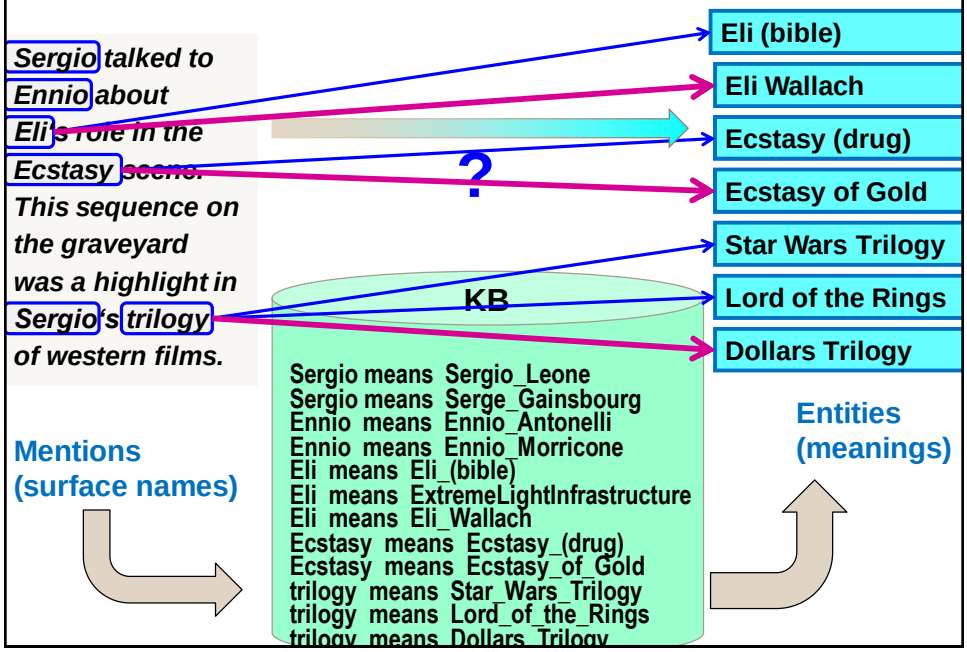




Entity disambiguation and linking

- Key requirement is that entities get identified
  - Named entity recognition (e.g., Stanford NER!)
- and disambiguated
  - Entit  sometimes  cation")
    - e.g., Michael Jordan the basketball player or the ML researcher

Mentions, Meanings, Mappings



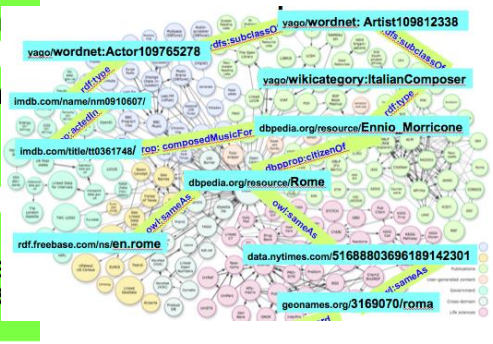
Incorporate structured info?

- and linked to a canonical reference
  - Freebase, dbPedia, Yago2, (WordNet)

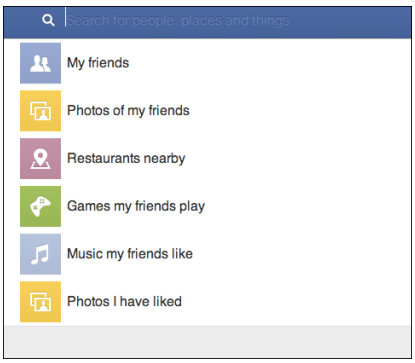
http://.../wiki/Dollars_Trilogy
http://.../wiki/The_Good,_the_Bad,
http://../wiki/Clint_Eastwood
http://.../wiki/Honorary_Academ

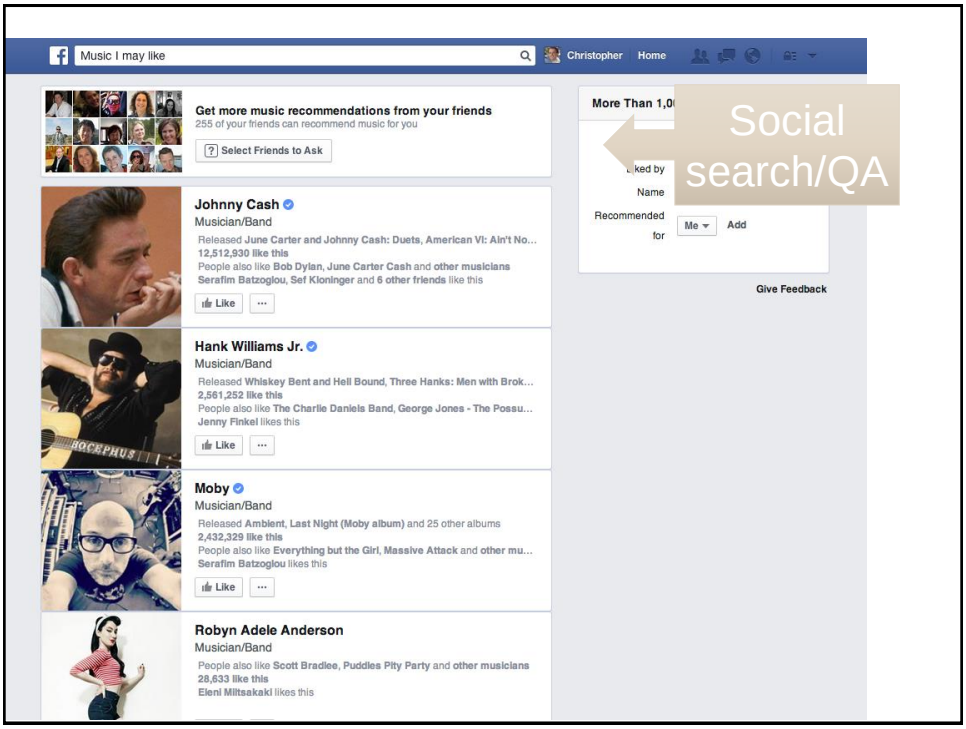
http://.../wiki/The_Good,_the_Ba
http://.../wiki/Metallica
http://.../wiki/Bellagio_(casino)
http://.../wiki/Ennio_Morricone

http://.../wiki/Sergio_Leone
http://.../wiki/The_Good,_the_Ba
http://.../wiki/For_a_Few_Dollars
http://.../wiki/Ennio_Morricone



Facebook Graph Search

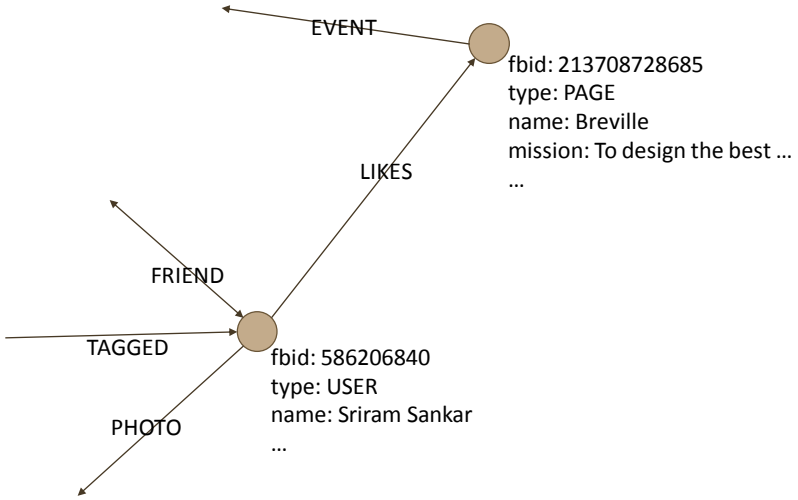


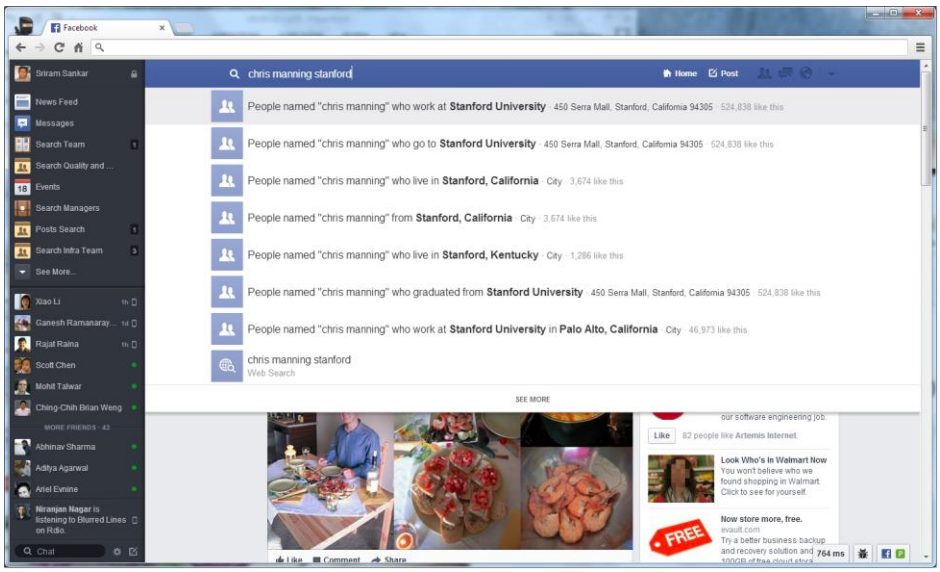


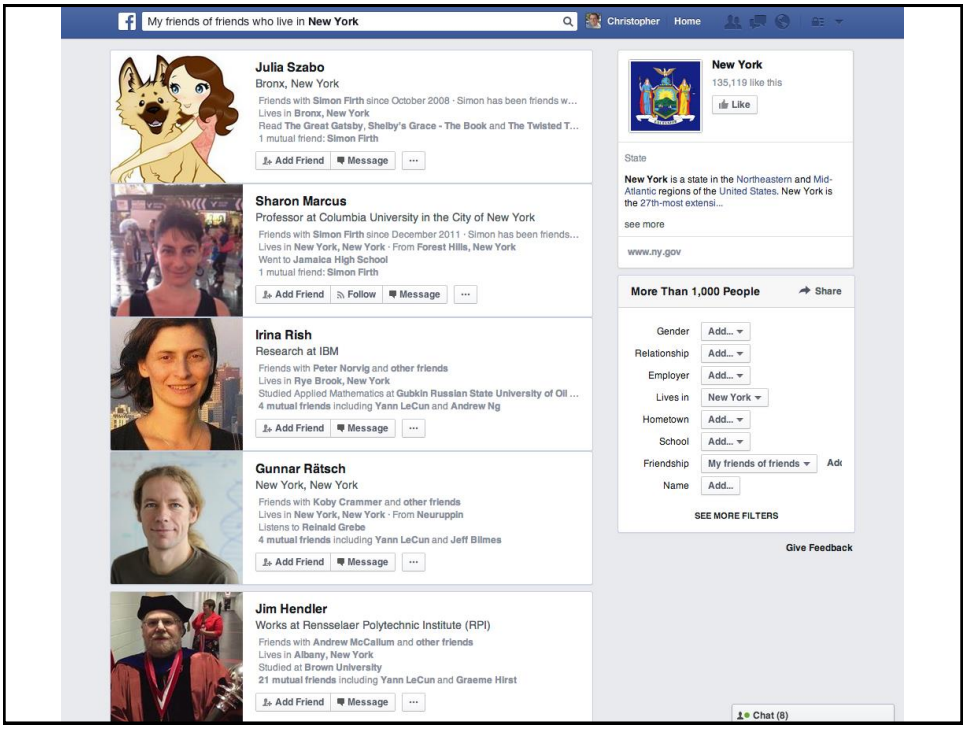
The Facebook Graph

- Collection of entities and their relationships
- Entities (users, pages, photos, etc.) are *nodes*
- Relationships (friendship, checkins, tagging, etc.) are *edges*
- Nodes and edges have metadata
- Nodes have a unique id – the *fbid*

Facebook Graph Snippet



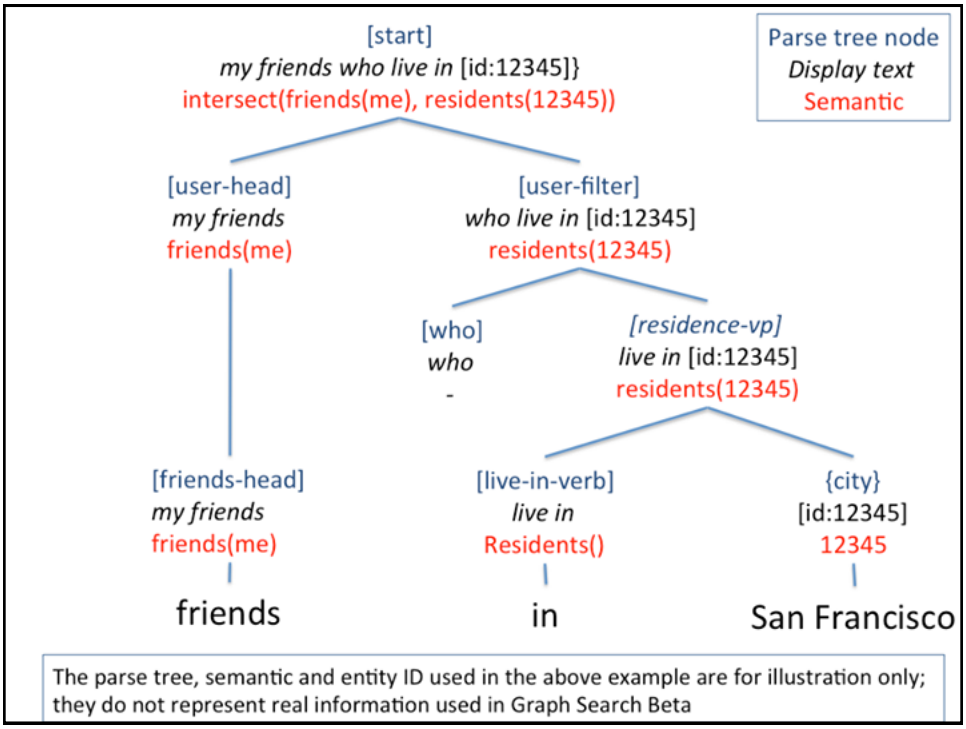




Facebook Graph Search

- Uses a weighted context free grammar (WCFG) to represent the Graph Search query language:
- [start] => [users] $1
- [users] => my friend friends(me)
- [users] => friends of [users] friends($1)
- [users] => {user} $1
- [start] => [photos] $1
- [photos] => photos of [users] photos($1)
- A terminal symbol can be an entity, e.g., {user}, {city}, {employer}, {group}; it can also be a word/phrase, e.g., friends, live in, work at, members, etc. A parse tree is produced by starting from [start] and expanding the production rules until it reaches terminal symbols.

https://www.facebook.com/notes/facebook-engineering/under-the-hood-the-natural-language-interface-of-graph-search/10151432733048920
http://spectrum.ieee.org/telecom/internet/the-making-of-facebooks-graph-search



3 approaches to question answering: Knowledge-based approaches (Siri)

- Build a semantic representation of the query
  - Times, dates, locations, entities, numeric quantities
- Map from this semantics to query structured data or resources
  - Geospatial databases
  - Ontologies (Wikipedia infoboxes, dbPedia, WordNet, Yago)
  - Restaurant review sources and reservation services
  - Scientific databases
  - Wolfram Alpha

54

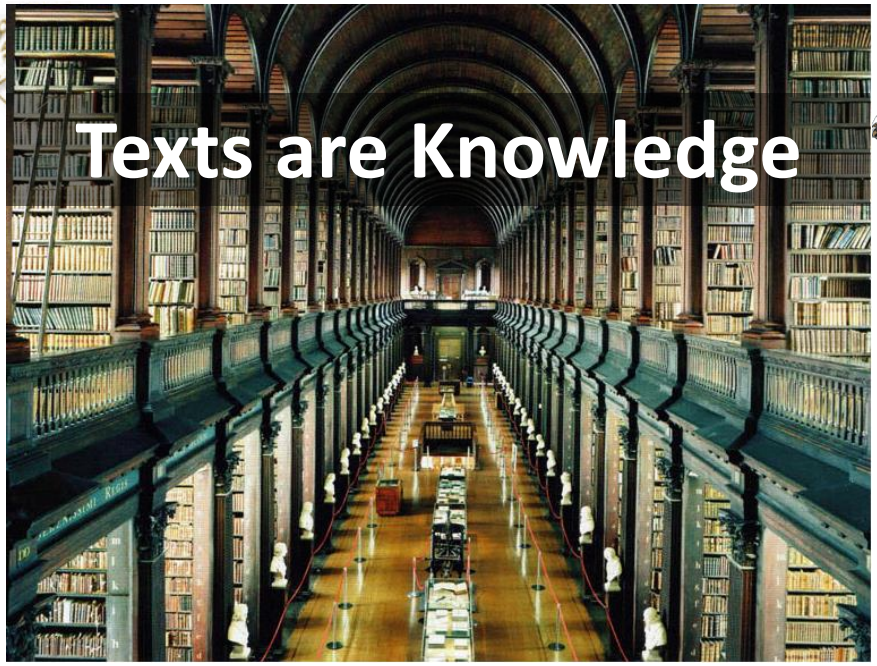
Text-based (mainly factoid) QA

- QUESTION PROCESSING
  - Detect question type, answer type, focus, relations
  - Formulate queries to send to a search engine
- PASSAGE RETRIEVAL
  - Retrieve ranked documents
  - Break into suitable passages and rerank
- ANSWER PROCESSING
  - Extract candidate answers (as named entities)
  - Rank candidates
    - using evidence from relations in the text and external sources

Hybrid approaches (IBM Watson)

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

56



Is the goal to go from language to knowledge bases?

- For humans, going from the largely unstructured language on the web to actionable information is effortlessly easy
- But for computers, it's rather **difficult!**
- This has suggested to many that if we're going to produce the next generation of intelligent agents, which can make decisions on our behalf
  - Answering our routine email
  - Booking our next trip to Fiji

  then we still first need to construct **knowledge bases**
- To go *from languages to information*

Knowledge: Not just semantics but pragmatics

- Pragmatics = taking account of context in determining meaning
- Search engines are great because they inherently take into account pragmatics ("associations and contexts")
  - [the national] → The National (a band)
  - [the national ohio] → The National - Bloodbuzz Ohio – YouTube
  - [the national broadband] → www.broadband.gov

Inference directly in text: Natural Logic

(van Benthem 2008, MacCartney & Manning 2009)

Q *Who is Beyoncé Knowles's husband?*
Q' *Beyoncé Knowles's husband is* ***X****.*

A *Beyoncé's marriage to rapper Jay-Z and portrayal of Etta James in Cadillac Records (2008) influenced her third album I Am... Sasha Fierce (2008).*

Task – Answer Sentence Selection

- Given a factoid question, find the sentence that
  - Contains the answer
  - Can sufficiently support the answer

Q
S_1

S_2

Scott Wen-tau Yih (ACL 2013) paper

Word Alignment for Question Answering
TREC QA (1999-2005)

- What is the fastest car in the world?

- The Jaguar XJ220 is the dearest, fastest and most sought after car on the planet.

Full NLP QA: LCC (Harabagiu/Moldovan) [below is the architecture of LCC's QA system circa 2003]