

Christopher Manning  
Pandu Nayak

## Introduction to Information Retrieval

CS276: Information Retrieval and Web Search  
Lecture 19: Web Question Answering

Introduction to Information Retrieval

## “Information retrieval”

The name **information retrieval** is standard, but as traditionally practiced, it’s not really right

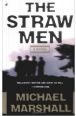
All you get is **document retrieval**, and beyond that the job is up to you

Introduction to Information Retrieval

## 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, 2002.

Introduction to Information Retrieval

## Web Search in 2025?

The web, it is a changing.

What will people do in 2025?

- Type key words into a search box?
- Use the Semantic Web?
- Ask questions to their computer in natural language?
- Use social or “human powered” search?

Introduction to Information Retrieval

## What do we know that’s happening?

- Much of what is going on is in the products of companies, and there isn’t exactly careful research explaining or evaluating it
- So most of this is my own meandering observations giving voice over to slides from others

Introduction to Information Retrieval

## Google

### What’s been happening? 2013–2019

- Many updates a year ... and 3rd party sites try to track them
  - e.g., <https://moz.com/google-algorithm-change> by & aimed at SEOs
- I just mention a few changes here
- New search index at Google: “Hummingbird” (2013)
  - <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 (2014)
  - Concepts versus words
- RankBrain (second half of 2015):
  - A neural net helps in document matching for the long tail

Introduction to Information Retrieval

## Google

### What's been happening? 2013–2019

- “Pigeon” update (July 2014):
  - More use of distance and location in ranking signals
- “Mobilegeddon” (Apr 21, 2015):
  - “Mobile friendliness” as a major ranking signal
- “App Indexing” (Android, iOS support May 2015)
  - Search results can take you to an app
- Mobile-friendly 2 (May 12, 2016):
  - About half of all searches are now from mobile
- “Fred” (1st quarter 2017)
  - Various changes discounting spammy, clickbaity, fake? sites

Introduction to Information Retrieval

## Google

### What's been happening? 2013–2019

- Longer snippets in results pages (Nov 2017)
- Mobile-first Index (Mar 2018)
  - Index mobile version of websites in preference to desktop!
- Revert snippet length in results pages (May 2018)
- “Medic” update (Aug 2018)
  - More emphasis on expertise, authoritativeness, trust
  - Big changes for diet, nutrition, medical products sites
- Core Algorithm Update (Mar 2019)
  - Seems kind of like “Medic 2”
- 2019 seems to have been kinda quiet so far ...

Introduction to Information Retrieval

## The role of knowledge bases


- 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

Introduction to Information Retrieval

## What's been happening

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




```

276 <p class="byline-column">
277   On Tennis
278 </p>
279 <p class="byline">By
280   <span itemprop="author creator" itemscope
281     itemType="http://schema.org/Person"
282     itemId="http://topics.nytimes.com/top/reference/linetopics/people/c/christopher_clarey/index.html">
283     <a href="http://topics.nytimes.com/top/reference/linetopics/people/c/christopher_clarey/index.html" rel="author">
284       More Articles by CHRISTOPHER CLAREY</a>
285     <span class="byline-author" data-byline-name="Christopher Clarey" itemprop="name">CHRISTOPHER CLAREY</span>
286   </p>
            
```

Introduction to Information Retrieval

## Mobile

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?

Introduction to Information Retrieval

## Mobile

- Mobile proved importance of NLU/QA
- [What is the best time for wildflowers in the bay area]

[Where to See Wildflowers in the Bay Area - Bay Area Parent](#)  
[www.bayarearent.com/Article/Where-to-See-Wildflowers-in-the-Bay-Area/](#) ▾  
 This is a great time of year to find wildflowers in the Bay Area. Call parks ... Loop Trail, known for being flat, moderate in difficulty and best for viewing the blooms.

[Best Bay Area hikes into wildflower country - Bay Nature](#)  
[baynature.org/article/best-bay-area-hikes-into-wildflower-country/](#) ▾ Bay Nature ▾  
 Mar 28, 2013 - It's spring, and all you nature lovers know what that means ... time to go see wildflowers! You've probably already noticed them popping in little ...

[Bay Area Wildflower Hikes | EveryTrail](#)  
[www.everytrail.com/guide/bay-area-wildflower-hikes](#) ▾  
 ★★★★★ Rating: 5 - 2 votes  
 Download the Bay Area Wildflower Hikes mobile travel guide to your iPhone or ... Even in lean years, wildflower season offers hikers beautiful reminders of the ... The best bloom is on the Chisnantuk Peak Trail, which is deep in the park and ...

## Information quality

- There have always been concerns about **information provenance** (the source) and **information reliability**, especially among “information professionals” (reporters, lawyers, spies, ...)
- It wasn't ignored on the web: ideas like PageRank were meant to find good content, and there has been a decade of work targeting link farms, etc.
- However, a lot of recent events have shown the limited effectiveness of that work, and how “fake” information easily gets upvoted and spreads

## Towards intelligent agents

- Two goals
  - Things not strings
  - Inference not search

## Two paradigms for question answering

- Text-based approaches
  - TREC QA, IBM Watson, DrQA
- 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)

Introduction to Information Retrieval

Search: "painkillers such as \* tylenol"

About 19,200 results (0.21 seconds)

**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) or ibuprofen (Advil, Motrin, others) ...

tylenol ⇒ painkiller

Introduction to Information Retrieval

Search: "painkiller "adverse effects include" stomach"

About 266,000 results (0.14 seconds)

**Painkiller Comparison Chart**  
 Oct 31, 2008 ... Chart showing prescription and non-prescription painkillers and what they are used for ... menstrual pain, fever. Common adverse effects include: nausea, ... asthma, upset stomach not for last pregnancy trimester...

upset stomach ⇒ adverse effect (on stomach)

Introduction to Information Retrieval

Search: "tylenol stomach"

About 1,390,000 results (0.28 seconds)

**Pain Relievers And Upset Stomach - Tylenol Extra Strength Pain ...**  
 Tylenol Extra Strength Pain Reliever Fever Reducer 100 Caplets - User Rating: 5 stars.  
 Review Summary: Taking the wrong pain reliever can cause you stomach ...

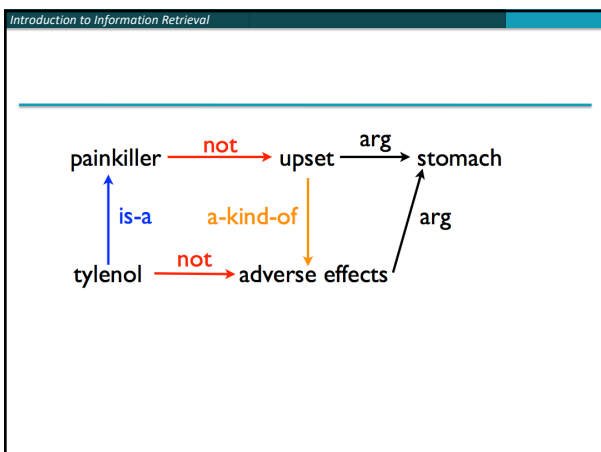
Treatment: Pain: Merck Manual Home Edition  
 (TYLENOL) has virtually no useful anti-inflammatory activity, does not affect the blood's ability to clot, and has almost no adverse effects on the stomach...

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

Introduction to Information Retrieval

Diagram showing relationships between terms:

painkiller                  upset                  stomach



Introduction to Information Retrieval

“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

Introduction to Information Retrieval

# Slides from Patrick Pantel (MSFT)

**Entity Detection**: Points to the search term 'Seven Samurai'.

**Structured Recommendations**: Points to the 'RELATED MOVIES' section listing titles like 'Dreams', 'Ran', 'Dersu Uzala', etc.

**Play trailer**: Points to the 'Play trailer' button for the movie 'Seven Samurai (1954)'.

**Structured Data**: Points to the movie details section, including the director 'Akira Kurosawa', running time, and cast.

Introduction to Information Retrieval

**Task completion**: Points to the search bar and filters for the flight search.

**Price prediction**: Points to the 'WEEKEND ESTIMATES' section showing flight prices for different dates.

**Aggregate ratings**: Points to the 'Compare rates from major travel sites' section for the Royal Lahaina Resort.

Introduction to Information Retrieval

**Direct Answer**: Points to the '45 years (March 27, 1970)' snippet at the top of the search results.

**Structured Data**: Points to the detailed information box for Mel Gibson, including his birth date, profession, and a list of his movies.

Introduction to Information Retrieval

**Direct Answer**: Points to the '45 years (March 27, 1970)' snippet at the top of the search results.

**Structured Data**: Points to the detailed information box for Mariah Carey, including her birth date, profession, and a list of her songs.

Introduction to Information Retrieval

**Direct Answer**: Points to the '53 years' snippet at the top of the search results.

**Structured Data**: Points to the detailed information box for Christopher D. Manning, including his profession as a computer scientist, his index score, and a list of his publications.

Introduction to Information Retrieval

**Direct Answer**: Points to the search results for 'pandu nayak', which includes a snippet from his Quora profile.

**Structured Data**: Points to the detailed information box for Pandu Nayak, including his profession as a computer scientist, his index score, and a list of his publications.

Introduction to Information Retrieval

## Patrick Pantel talk (Then) Current experience

A screenshot of a Bing search result for 'jetbeam mt-0'. The page shows search results with a list of related searches on the left and a main result area on the right. The main result is for 'JETBeam RRT-0 R5 with Infinite Ramping LED Flashlight'. The page is annotated with a vertical label on the left: 'Recognize entity in query' pointing to the search term and 'Actions easily accessible' pointing to the list of actions like 'Read Reviews', 'Watch Demo Video', etc.

Introduction to Information Retrieval

## Desired experience: Towards actions

A screenshot of a Bing search result for 'jetbeam mt-0', similar to the previous one. It is annotated with a vertical label on the left: 'Recognize entity in query' pointing to the search term and 'Actions easily accessible' pointing to the list of actions. The annotations are more prominent, highlighting the transition from a simple search to a list of actionable items.

Introduction to Information Retrieval

## Politician

A screenshot of a Bing search result for 'Obama'. The page shows search results for Barack Obama, including his official website and a Wikipedia entry. The page is annotated with a vertical label on the left: 'Recognize entity in query' pointing to the search term and 'Actions easily accessible' pointing to the list of actions like 'Watch Weekly Address', 'Follow Social Updates', etc.

Introduction to Information Retrieval

## Actions vs. Intents

A diagram titled 'Actions vs. Intents'. It shows a flow from 'User Intents and Goals' (plan vacation, get in shape) to 'Query' (hilton orlando reviews, sea world location, how to lose weight). Below the query, it shows 'Query Intent' (Informational, Navigational, Transactional) and 'Finer-grained Intents' (Advice, Locate, Download, Obtain, Interact). At the bottom, it shows 'Actions on Entities' (get address(landmark), add to Netflix queue(film), read reviews(hotel), buy(camera)).

Introduction to Information Retrieval

## Learning actions from web usage logs

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

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

- 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

Introduction to Information Retrieval

A screenshot of a Bing search result for 'hannah montana'. The page shows search results for the Disney Channel show. A red box highlights a section of the results with the text: 'A hodgepodge of related strings' and 'Only actionable through search'. The page is annotated with a vertical label on the left: 'Recognize entity in query' pointing to the search term and 'Actions easily accessible' pointing to the list of actions like 'Watch TV Shows at Bing', etc.

Introduction to Information Retrieval

hannah montana - Bing

Web Images Videos Shopping News Maps More | MSN | Hotmail

Web Videos News TV Shopping Images More

RELATED PRODUCTS

Get Talk Game  
Hannah Montana Doll  
Kids Rock Guitar  
Billy Ray Cyrus  
Aron Kelly

RELATED PEOPLE

Billy Ray Cyrus  
Aron Kelly

RELATED MUSIC

Miley Cyrus Soundtrack  
Taylor Swift Our Song  
Miley Cyrus Performs At Idol

ALL RESULTS 1-10 of 14,500,000 results

Hannah Montana Forever | Disney Channel  
Explore the Hannah Montana Forever home page for videos, full-length episodes, character bios, games, photos, downloads, music, and more!

Hannah Montana Games  
Disney Channel Games  
Good Luck Charlie  
Show more results from tv.disney.go.com

Entertainment News: hannah montana  
Billy Ray Cyrus 'Hannah Montana' destroyed family  
NASHVILLE, Tenn. (AP) — Billy Ray Cyrus says the Disney TV show 'Hannah Montana' destroyed his family, causing his divorce and sending daughter Miley Cyrus spinning out of... Yahoo Finance - 10 hours ago

Hannah Montana - Wikipedia, the free encyclopedia  
Production: Cast Episodes: Films  
Hannah Montana is an American television series, which debuted on March 24, 2006 on Disney Channel, and also airs Saturday mornings on ABC as part of the 4i

Character Bios | Hannah Montana Forever | Disney Channel  
Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's Hannah Montana Forever!

Watch TV Shows at Bing®  
Browse Video Clips and Popular Episodes at Bing Today!

Hannah Montana  
Free Shipping \$50 on 100,000 Items  
Shop Hannah etc. Online at Target  
www.target.com

Hannah Montana Items  
We Have The Hannah Montana Items To Complete Any Fans Collection!  
Hannah Montana BradfordExchange.com  
Compare Choices and Prices  
Everything You Want and Need! Shopzilla.com

From Hannah Montana  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

Hannah Montana Wigs  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

is says the Disney TV show 'Hannah Montana' destroyed his family and sending daughter Miley Cyrus spinning out of... Yahoo Finance - 10 hours ago

Montana BradfordExchange.com  
Compare Choices and Prices  
Everything You Want and Need! Shopzilla.com

From Hannah Montana  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

Character Bios | Hannah Montana Forever | Disney Channel  
Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's Hannah Montana Forever!

Internal preview Help improve Bing

Introduction to Information Retrieval

hannah montana - Bing

Web Images Videos Shopping News Maps More | MSN | Hotmail

Web Videos News TV Shopping Images More

RELATED PRODUCTS

Get Talk Game  
Hannah Montana Doll  
Kids Rock Guitar  
Billy Ray Cyrus  
Aron Kelly

RELATED PEOPLE

Billy Ray Cyrus  
Aron Kelly

RELATED MUSIC

Miley Cyrus Soundtrack  
Taylor Swift Our Song  
Miley Cyrus Performs At Idol

ALL RESULTS 1-10 of 14,500,000 results

Hannah Montana Forever | Disney Channel  
Explore the Hannah Montana Forever home page for videos, full-length episodes, character bios, games, photos, downloads, music, and more!

Hannah Montana Games  
Disney Channel Games  
Good Luck Charlie  
Show more results from tv.disney.go.com

Entertainment News: hannah montana  
Billy Ray Cyrus 'Hannah Montana' destroyed family  
NASHVILLE, Tenn. (AP) — Billy Ray Cyrus says the Disney TV show 'Hannah Montana' destroyed his family, causing his divorce and sending daughter Miley Cyrus spinning out of... Yahoo Finance - 10 hours ago

Hannah Montana - Wikipedia, the free encyclopedia  
Production: Cast Episodes: Films  
Hannah Montana is an American television series, which debuted on March 24, 2006 on Disney Channel, and also airs Saturday mornings on ABC as part of the 4i

Character Bios | Hannah Montana Forever | Disney Channel  
Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's Hannah Montana Forever!

Watch TV Shows at Bing®  
Browse Video Clips and Popular Episodes at Bing Today!

Hannah Montana  
Free Shipping \$50 on 100,000 Items  
Shop Hannah etc. Online at Target  
www.target.com

Hannah Montana Items  
We Have The Hannah Montana Items To Complete Any Fans Collection!  
Hannah Montana BradfordExchange.com  
Compare Choices and Prices  
Everything You Want and Need! Shopzilla.com

From Hannah Montana  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

Hannah Montana Wigs  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

is says the Disney TV show 'Hannah Montana' destroyed his family and sending daughter Miley Cyrus spinning out of... Yahoo Finance - 10 hours ago

Montana BradfordExchange.com  
Compare Choices and Prices  
Everything You Want and Need! Shopzilla.com

From Hannah Montana  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

Character Bios | Hannah Montana Forever | Disney Channel  
Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's Hannah Montana Forever!

Internal preview Help improve Bing

**Big Wins**

- Typed relations
- User Interface accesses structured data
- Click through experience can now leverage strongly-typed identifier

Introduction to Information Retrieval

hannah montana - Bing

Web Images Videos Shopping News Maps More | MSN | Hotmail

Web Videos News TV Shopping Images More

RELATED PRODUCTS

Get Talk Game  
Hannah Montana Doll  
Kids Rock Guitar  
Billy Ray Cyrus  
Aron Kelly

RELATED PEOPLE

Billy Ray Cyrus  
Aron Kelly

RELATED MUSIC

Miley Cyrus Soundtrack  
Taylor Swift Our Song  
Miley Cyrus Performs At Idol

ALL RESULTS 1-10 of 14,500,000 results

Hannah Montana Forever | Disney Channel  
Explore the Hannah Montana Forever home page for videos, full-length episodes, character bios, games, photos, downloads, music, and more!

Hannah Montana Games  
Disney Channel Games  
Good Luck Charlie  
Show more results from tv.disney.go.com

Entertainment News: hannah montana  
Billy Ray Cyrus 'Hannah Montana' destroyed family  
NASHVILLE, Tenn. (AP) — Billy Ray Cyrus says the Disney TV show 'Hannah Montana' destroyed his family, causing his divorce and sending daughter Miley Cyrus spinning out of... Yahoo Finance - 10 hours ago

Hannah Montana - Wikipedia, the free encyclopedia  
Production: Cast Episodes: Films  
Hannah Montana is an American television series, which debuted on March 24, 2006 on Disney Channel, and also airs Saturday mornings on ABC as part of the 4i

Character Bios | Hannah Montana Forever | Disney Channel  
Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's Hannah Montana Forever!

Watch TV Shows at Bing®  
Browse Video Clips and Popular Episodes at Bing Today!

Hannah Montana  
Free Shipping \$50 on 100,000 Items  
Shop Hannah etc. Online at Target  
www.target.com

Hannah Montana Items  
We Have The Hannah Montana Items To Complete Any Fans Collection!  
Hannah Montana BradfordExchange.com  
Compare Choices and Prices  
Everything You Want and Need! Shopzilla.com

From Hannah Montana  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

Hannah Montana Wigs  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

is says the Disney TV show 'Hannah Montana' destroyed his family and sending daughter Miley Cyrus spinning out of... Yahoo Finance - 10 hours ago

Montana BradfordExchange.com  
Compare Choices and Prices  
Everything You Want and Need! Shopzilla.com

From Hannah Montana  
15,000 Costumes - All Sizes & Styles  
Low Prices & Same Day Shipping!  
www.BuyCostumes.com  
See your message here

Character Bios | Hannah Montana Forever | Disney Channel  
Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's Hannah Montana Forever!

Internal preview Help improve Bing

**Big Wins**

- Typed relations
- User Interface accesses structured data
- Click through experience can now leverage strongly-typed identifier
- Brokered Actions (one click conversions)

Introduction to Information Retrieval

## Entity disambiguation and linking

- Key requirement is that entities get identified
  - Named entity recognition (e.g., Stanford NER!)
- and disambiguated
  - Entity linking (or sometimes "Wikification")
    - e.g., Michael Jordan the basketballer or the ML guy

Introduction to Information Retrieval

## Mentions, Meanings, Mappings [G. Weikum]

Sergio talked to Ennio about Eli's role in the Ecstasy scene. This sequence on the graveyard was a highlight in Sergio's trilogy of western films.

**Mentions (surface names)**

**KB**

Sergio means Sergio Leone  
Sergio means Serge Gainsbourg  
Ennio means Ennio Antonelli  
Ennio means Ennio Morricone  
Eli means Eli (bible)  
Eli means ExtremeLightInfrastructure  
Eli means Eli Wallach  
Ecstasy means Ecstasy (drug)  
Ecstasy means Ecstasy of Gold  
trilogy means Star Wars Trilogy  
trilogy means Lord of the Rings  
trilogy means Dollars Trilogy

**Entities (meanings)**

Eli (bible)  
Eli Wallach  
Ecstasy (drug)  
Ecstasy of Gold  
Star Wars Trilogy  
Lord of the Rings  
Dollars Trilogy

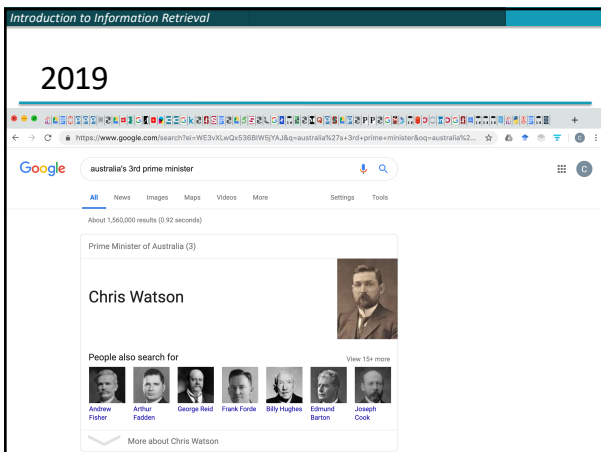
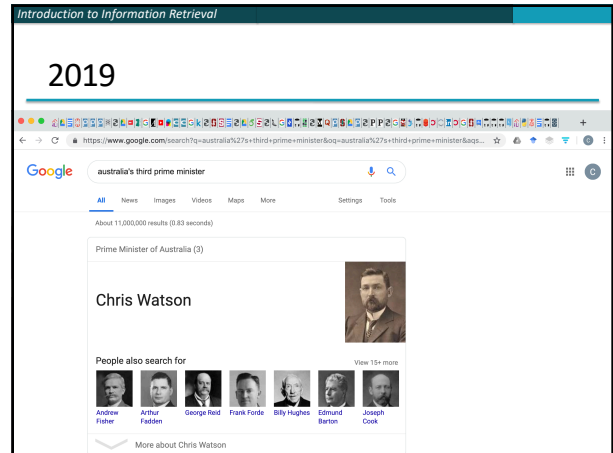
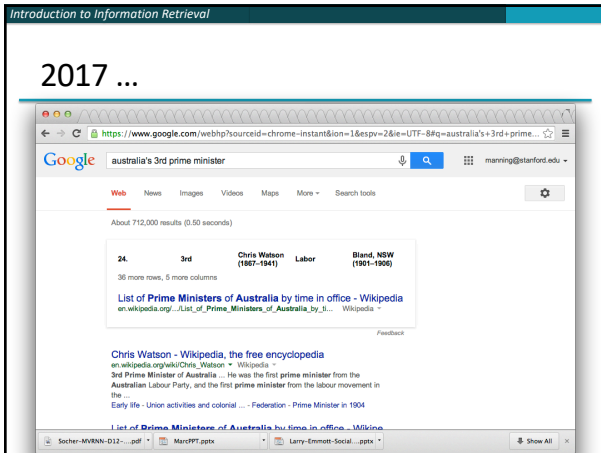
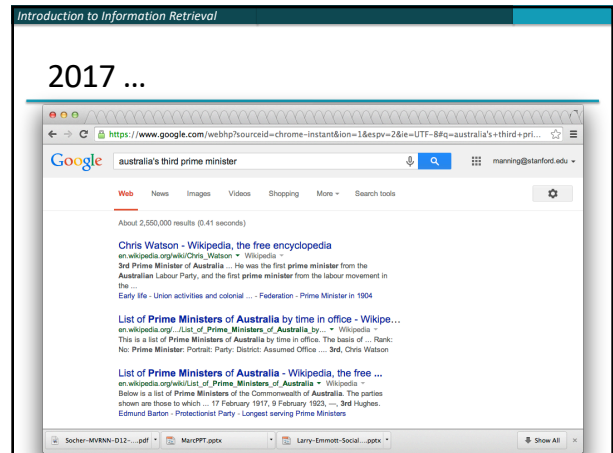
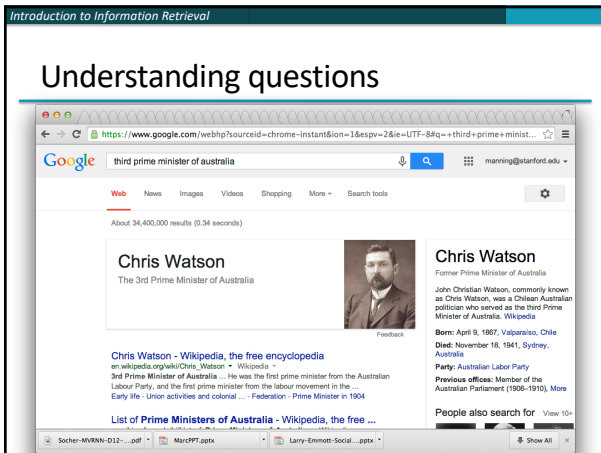
Introduction to Information Retrieval

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

http://.../wiki/Dollars\_Triology  
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\_B  
http://.../wiki/For\_a\_Few\_Dollar  
http://.../wiki/Ennio\_Morricone



**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



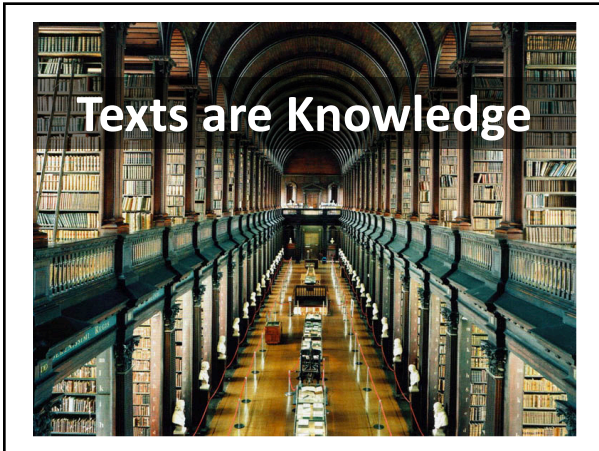
## Text-based (mainly factoid) QA

- 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

50



## Knowledge: Jeremy Zawodny says ...



## 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*
  - But should we rather just have computers work with language?



## Knowledge: Not just semantics but pragmatics

Pragmatics = taking account of context in determining meaning  
A natural part of language understanding and use

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](http://www.broadband.gov)

Scott Wen-tau Yih (ACL 2013) paper

Microsoft Research

bing MSN

Who won the best actor Oscar in 1973?

Lemmon was awarded the Best Supporting Actor Oscar in 1956 for *Mister Roberts* (1955) and the Best Actor Oscar for *Save the Tiger* (1973), becoming the first actor to achieve this rare double...  
Source: Jack Lemmon - Wikipedia

Jack Lemmon

John Uhler "Jack" Lemmon III was an American actor and musician. He starred in more than 60 films, including *Tom...*  
en.wikipedia.org

Jack Lemmon - Wikipedia, the free encyclopedia

### Who won the best actor Oscar in 1973?

American actor and musician. He starred in more than 60 films including ...

Jack Lemmon - IMDb

www.imdb.com/name/nm0000493

Born Feb 08, 1925 · Died Jun 27, 2001 · News · 341 photos · Biography

Actor: *The Apartment* (1960) · *Some Like It Hot* (1959) · *Clangor Glen Ross* (1992) · *Grumpy Old Men* (1993) · **Jack Lemmon's** father was the ...

Awards: Academy Award for Actor in a Leading Role · +

Images of Jack Lemmon

bing.com/images

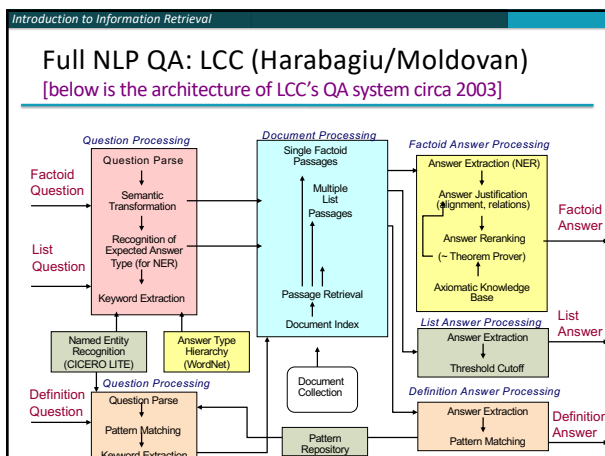
Microsoft Research

## 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.  
[Harabagiu & Moldovan, 2001]

- Assume that there is an underlying alignment
  - Describes which words in and can be associated
  - See if the (syntactic/semantic) relations support the answer



### DrQA: Open-domain Question Answering

(Chen, et al. ACL 2017) <https://arxiv.org/abs/1704.00051>

WIKIPEDIA

ask me anything

58

### Open-domain Question Answering

WIKIPEDIA

- SQuAD**

Q: How many of Warsaw's inhabitants spoke Polish in 1933? A: 833,500
- TREC**

Q: What U.S. state's motto is "Live free or Die"? A: New Hampshire
- WebQuestions** (Berant et al, 2013) 

Q: What part of the atom did Chadwick discover? A: neutron
- WikiMovies** (Miller et al, 2016) 

Q: Who wrote the film Gigli? A: Martin Brest

59

Q: How many of Warsaw's inhabitants spoke Polish in 1933?

WIKIPEDIA

Document Retriever

Document Reader

833,500

60

(Chen et al, 2017)

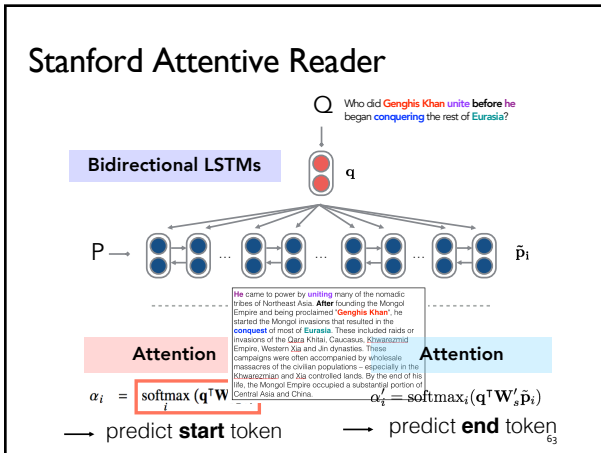
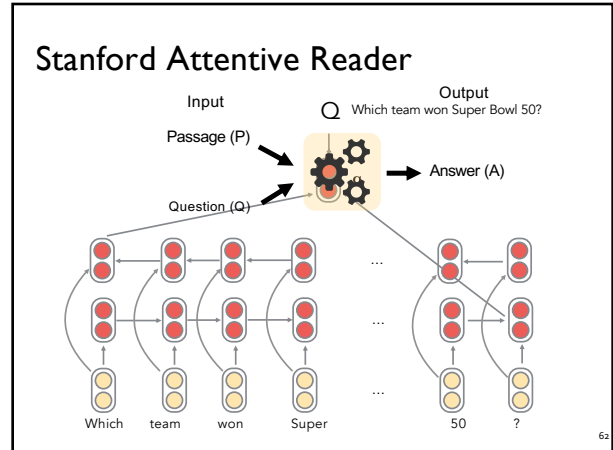
# Document Retriever

Dataset	Wiki Search	Doc. plain	Retriever +bigrams
SQuAD	62.7	76.1	<b>77.8</b>
CuratedTREC	81.0	85.2	<b>86.0</b>
WebQuestions	73.7	<b>75.5</b>	74.4
WikiMovies	61.7	54.4	<b>70.3</b>

Traditional tf.idf inverted index + efficient bigram hash

**70-86%** of questions we have that the answer segment appears in the **top 5** articles

63



## SQuAD Results (single model)

Model	F1
Logistic regression	51.0
Fine-Grained Gating (Carnegie Mellon U)	73.3
Match-LSTM (Singapore Management U)	73.7
DCN (Salesforce)	75.9
BiDAF (UW & Allen Institute)	77.3
Multi-Perspective Matching (IBM)	78.7
ReasoNet (MSR Redmond)	79.4
DrQA (Chen et al. 2017)	79.4
r-net (MSR Asia) [Wang et al., ACL 2017]	79.7
Human performance	91.2

64

## General questions

Combined with **Web search**, we can answer **57.5%** of **trivia questions** correctly

**Q:** The Dodecanese Campaign of WWII that was an attempt by the Allied forces to capture islands in the Aegean Sea was the inspiration for which acclaimed 1961 commando film?

**A:** The Guns of Navarone

**Q:** American Callan Pinckney's eponymously named system became a best-selling (1980s-2000s) book/video franchise in what genre?

**A:** Fitness

65

## Demo

Hi!

Hello! Please ask a question.

What is question answering?

a computer science discipline within the fields of information retrieval and natural language processing

Who was the winning pitcher in the 1956 World Series?

Don Larsen

What is the answer to life, the universe, and everything?

42

66