

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

- “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 ...

The role of knowledge bases

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

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.bayareaparent.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 beat 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

Introduction to Information Retrieval

Introduction to Information Retrieval

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

Introduction to Information Retrieval

Example from Fernando Pereira (GOOG)

painkillers that don't thin blood Search

painkillers that don't thin blood
painkillers that don't cause constipation
painkillers that don't cause nausea
painkillers that don't upset stomach frequent query
painkillers that don't

About 89,300 results (0.14 seconds) Advanced search

Introduction to Information Retrieval

class negation synonyms

painkillers that don't upset stomach

About 28,700 results (0.15 seconds)

Symptoms of Crohn's

www.crohnsline.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.healthblurbz.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, Aleve... - 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. — bonem... — 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

"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) others) or ibuprofen (Advil, Motrin, others). ... www.mayoclinic.com/.../DSECTION=treatments-and-drugs - Cached - Similar

tylenol \Rightarrow painkiller

painkiller "adverse effects include" stomach

About 266,000 results (0.14 seconds)

Painkiller Comparison Chart
Oct 31, 2009 ... 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. ... www.disabled-world.com/.../painkiller-chart.shtml - Cached - Similar

POV: List of Painkillers and what they are used for: Over the Counter ...
File Format: PDF/Acrobat / Quick View
Common adverse effects include: nausea, dyspepsia, gastrointestinal ulcers/bleeding. ... upset stomach, not for last pregnancy trimester. Individuals ... healingwithsoftlasers.com/.../Painkiller%20side%20effects%20chart%20.pdf

upset stomach \Rightarrow adverse effect (on stomach)

tylenol stomach

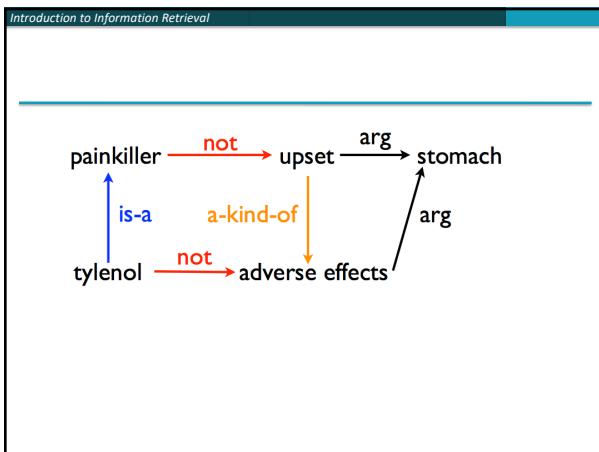
About 1,390,000 results (0.28 seconds)

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** ... www.opinions.com / Wellness & Beauty : Health Aids - Cached - Similar

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** ... www.merck.com/mmhe/sec06/ch07/8/ch078d.html - Cached - Similar

tylenol \Rightarrow no adverse effects (on stomach)
 \Rightarrow no upset stomach

painkiller upset stomach



“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)

Introduction to Information Retrieval

Patrick Pantel talk (Then) Current experience

Web Images Videos Shopping News Maps Mode 1 MSN Home

 **bing** HS Beta

jetbeam rrt-0

Web Videos More ▾

RELATED SEARCHES

- JETBeam RRT 0 Review
- JETBeam Flashlights
- JETBeam RRT-0 R5
- JETBeam Raptor
- JETBeam MX
- JETBeam
- RRT-0 R5
- 240 Lumens

SEARCH HISTORY

- Search more & see your history
- See all
- Clear all - Turn off

ALL RESULTS 1-10 of 6,600 results Advanced

JETBeam RRT-0 R5 with Infinite Ramping LED Flashlight

Going Gear JETBeam RRT-0 R5 with Infinite Ramping LED Flashlight [beammr05] - The original RRT-0 has been updated to be an even more versatile flashlight ... [getproductdetails.php?item=page_product_id&cat=rrt-14_229&product_id=438](#)

Mark as spam

Jetbeam RRT-0

Now In Stock Model: RRT-0 (S2) Specially designed for Military, Law Enforcement, Self-defense, Hunting, Search & Rescue and General Outdoors. The Rapid ... [www.jetbeam.com/jetbeam/RRT200.html](#)

Mark as spam

JETBeam RRT-0 Overview With Lots of Pictures

I got my JETBeam RRT-0 Raptor today. All I can say is JETBeam! This thing is a PICTURES QUICKLY... BEFORE MY HEAD EXPLODES! Ok, it didn't... 322 replies since November 2009 [www.electronics-forums.com/showthread.php?244375-JETBeam-RRT-0-Overview-With-Lots-of-Pics](#)

Mark as spam

JETBeam Flashlights

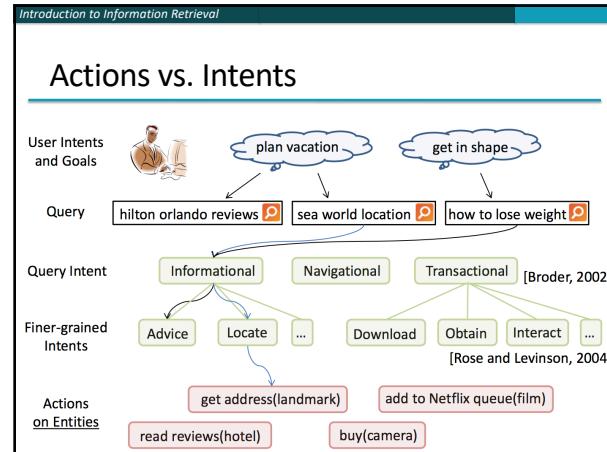
Specification Model: JETBeam RRT-0(R2) Flashlight; Specially designed for Military, Law Enforcement, Self-defense, Hunting, Search & Rescue and Outdoorsman [jetbeamzing.com](#) [Mark as spam](#)

Introduction to Information Retrieval

Desired experience: Towards actions

Recognize entity in query

Actions easily accessible



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

<http://www.bing.com/search?q=hannah>

hannah montana - Bing

Web Images Videos Shopping News Maps More | MSN Hotmail

Sign in Redmond, Washington Preferences

MSN Beta v1.11

bing hannah montana

Web Videos News TV Shopping Images More ▾

RELATED PRODUCTS

Girl Talk
Hannah Montana Doll
Kids Rock
Guitar

RELATED PEOPLE

Billy Ray Cyrus
Aaron Kelly

RELATED MUSIC

Miley Cyrus Soundtrack
Taylor Swift Our Song

RELATED VIDEOS

Miley Cyrus Performs At 400

ALL RESULTS 1-10 of 14,450,000 results Advanced Sponsored sites Watch TV Shows at Bing! ▾
Browse Video Clips and Popular Episodes at Bing! ▾
Shop TV Shows on Bing! ▾
Shop Hannah Montana Items on Bing! ▾
Compare Prices on 100,000 Items
Shop Hannah et. Online at Target.
www.Target.com

Hannah Montana Forever Disney Channel

Watch the Hannah Montana Forever home page for video, full-length episodes, character bios, photos, downloads, music, and more!

to Disney Channel | Hannah Montana Mark as spam

Disney Channel Games TV Shows

Games Phineas & Ferb

Goofy Luck Charlie Full episodes

Show more results from Disney.go.com

Entertainment News: **Hannah Montana**

By Bill Clegg, [Bill Clegg](#) [Entertainment Weekly](#)

NASHVILLE, Tenn. (AP) — Billy Ray Cyrus says the Disney TV show "Hannah Montana" destroyed his family, causing his divorce and sending daughter Miley into the arms of her estranged mother, actress Tish Cyrus.

Miley Cyrus' Aug 12th Divorce With Over Hannah Montana Star Entertainment News Headlines Lady Gaga at the Grammys, HBO shows and Hannah Montana Florida Times-Union

Hannah Montana - Wikipedia, the free encyclopediawiki

Production Cast Episodes Filmography

Hannah Montana is an American television series, which debuted on March 24, 2006 on Disney Channel, and stars Miley Cyrus as Miley Stewart, a teenage girl coming on as part of the all-new wikipedia.org/Hannah_montana Mark as spam

Character Bios | Hannah Montana Forever | Disney Channel

Meet Miley, Lily, Oliver, Rico, and the rest of the gang from Disney Channel's **Hannah Montana**! Go to disneychannel.hannahmontana.com/characters/index.html Mark as spam

Internal preview Help improve this

Big Wins

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

Montana Bradford Exchange.com
From Hannah Montana®
Compare Choices and Prices.
Everything You Want and Need!
Need!

Hannah Montana Wigs
1500+ Choices - All Sizes & Styles
Low Prices & Same Day Shipping!
www.BuyCostumes.com

See your message here

Introduction to Information Retrieval

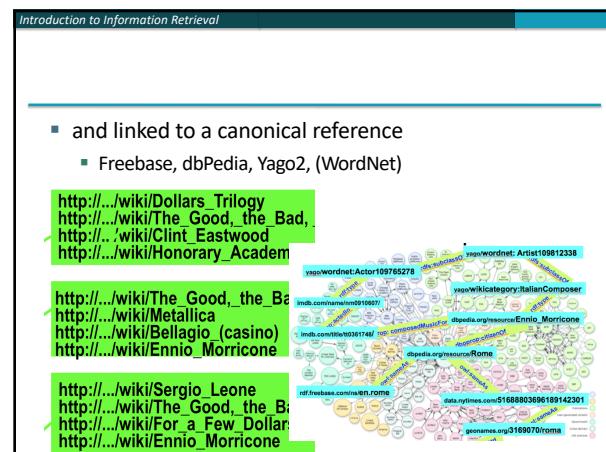
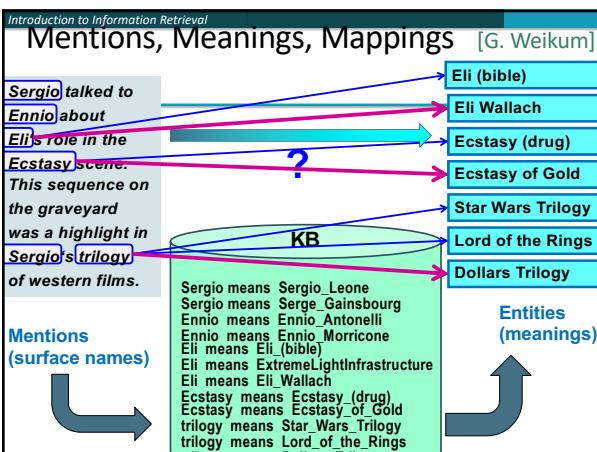
The screenshot shows a search results page from Bing. The query 'hannah montana - Bing' is entered in the search bar. The results are categorized under 'Web', 'Images', 'Videos', 'Shopping', 'News', 'Maps', 'More', and 'MSN'. The top result is a link to the official Disney Channel Hannah Montana website. Below it, there are several other links related to Hannah Montana, including her official YouTube channel, fan sites, and news articles. A sidebar on the right is titled 'Big Wins' and lists five bullet points: 'Typed relations', 'User interface accesses structured data', 'Click through experience can now leverage strongly-typed identifier', and 'Brokered Actions (one click conversions)'. At the bottom of the sidebar, there's a message from 'Montana Bradford' about exchange.com. The overall layout is typical of a web browser interface.

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

Understanding questions

Google search results for "third prime minister of australia". The top result is a Wikipedia page for Chris Watson, the 3rd Prime Minister of Australia. The page includes his portrait, basic bio, and a link to 'List of Prime Ministers of Australia'.

Introduction to Information Retrieval

2017 ...

Google search results for "australia's third prime minister". The top result is a Wikipedia page for Chris Watson, the 3rd Prime Minister of Australia. The page includes his portrait, basic bio, and a link to 'List of Prime Ministers of Australia'.

Introduction to Information Retrieval

2017 ...

Google search results for "australia's 3rd prime minister". The top result is a Wikipedia page for Chris Watson, the 3rd Prime Minister of Australia. The page includes his portrait, basic bio, and a link to 'List of Prime Ministers of Australia'.

Introduction to Information Retrieval

2019

Google search results for "australia's third prime minister". The top result is a Wikipedia page for Chris Watson, the 3rd Prime Minister of Australia. The page includes his portrait, basic bio, and a link to 'List of Prime Ministers of Australia'.

Introduction to Information Retrieval

2019

Google search results for "australia's 3rd prime minister". The top result is a Wikipedia page for Chris Watson, the 3rd Prime Minister of Australia. The page includes his portrait, basic bio, and a link to 'List of Prime Ministers of Australia'.

Introduction to Information Retrieval

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

48

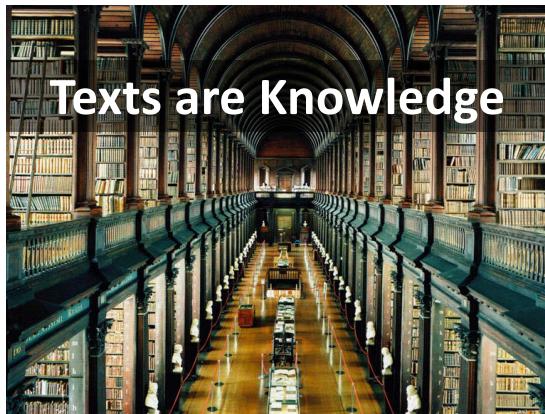
Text-based (mainly factoid) QA

- 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

Scott Wen-tau Yih (ACL 2013) paper

bing Who won the best actor Oscar in 1973?

Leanon 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 American actor and musician. He starred in more than 60 films including *Som...
www.imdb.com*

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/nm000493 ...
Born Feb 08, 1917 | Died Jun 27, 2001 | News | 341 photos | Biography | Actor: *The Apartment* (1960) | *Some Like It Hot* (1959) | *Gentlemen Prefer Blondes* (1953) | *Jack Lemmon's father was the ...*

Images of jack lemm bing.com/images

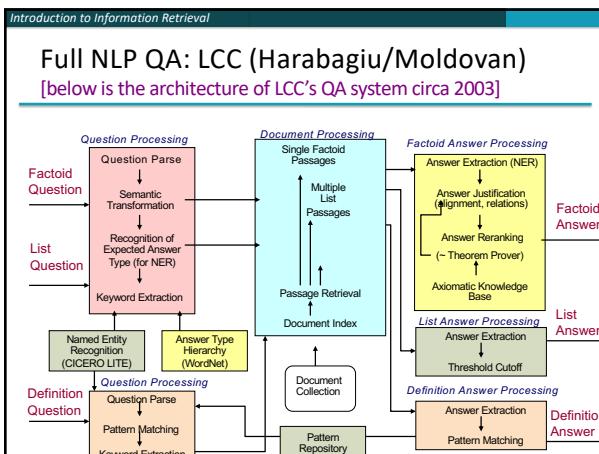

Microsoft Research

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

58

Open-domain Question Answering

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

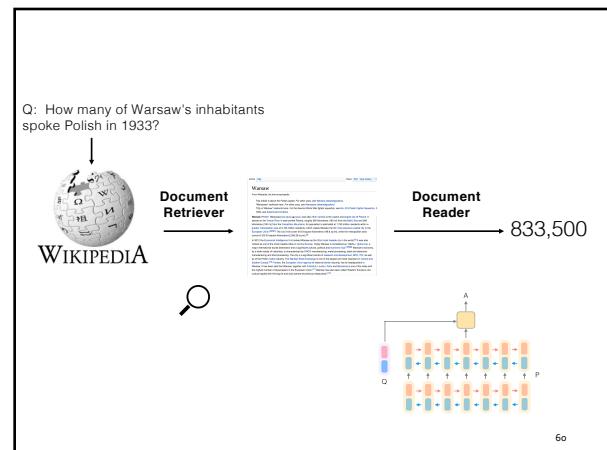
TREC
Q: What U.S. state's motto is "Live free or Die"? A: New Hampshire

WebQuestions (Berant et al, 2013) **Freebase**
Q: What part of the atom did Chadwick discover? A: neutron

WikiMovies (Miller et al, 2016) **OMDb**
Q: Who wrote the film Gigli? A: Martin Brest

WIKIPEDIA

59



Document Retriever

(Chen et al, 2017)

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

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

61

Traditional
tf.idf
inverted
index +
efficient
bigram
hash

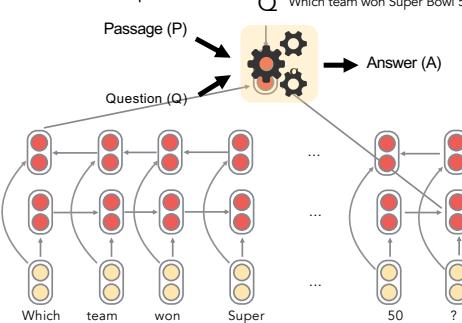
Stanford Attentive Reader

Input Q Which team won Super Bowl 50?

Passage (P)

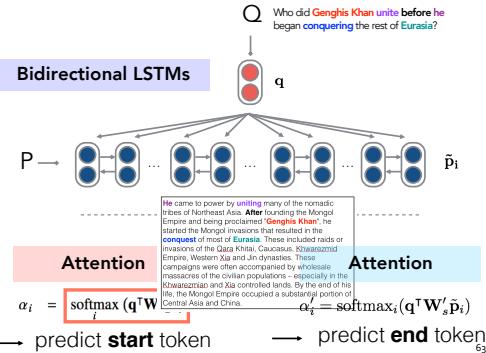
Question (Q)

Output Answer (A)



62

Stanford Attentive Reader



$\alpha_i = \text{softmax}_i(q^T W_i p_i)$

→ predict start token

→ predict end token

63

SQuAD Results (single 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