

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? 2013–2017

- 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? 2013–2017

- “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, clickbait, fake? sites

Introduction to Information Retrieval

The role of knowledge bases

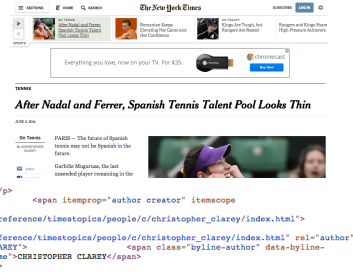
- Google Knowledge Graph
- Facebook Graph Search
- Bing’s Satori
- Things like Wolfram Alpha

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   by De Zevalle
278 </p>
279 <p class="byline">By Christopher Clarey
280 </p>
281 </p>
282 </p>
283 </p>
284 </p>
285 </p>
286 </p>
287 </p>
288 </p>
289 </p>
290 </p>
291 </p>
292 </p>
293 </p>
294 </p>
295 </p>
296 </p>
297 </p>
298 </p>
299 </p>
300 </p>
301 </p>
302 </p>
303 </p>
304 </p>
305 </p>
306 </p>
307 </p>
308 </p>
309 </p>
310 </p>
311 </p>
312 </p>
313 </p>
314 </p>
315 </p>
316 </p>
317 </p>
318 </p>
319 </p>
320 </p>
321 </p>
322 </p>
323 </p>
324 </p>
325 </p>
326 </p>
327 </p>
328 </p>
329 </p>
330 </p>
331 </p>
332 </p>
333 </p>
334 </p>
335 </p>
336 </p>
337 </p>
338 </p>
339 </p>
340 </p>
341 </p>
342 </p>
343 </p>
344 </p>
345 </p>
346 </p>
347 </p>
348 </p>
349 </p>
350 </p>
351 </p>
352 </p>
353 </p>
354 </p>
355 </p>
356 </p>
357 </p>
358 </p>
359 </p>
360 </p>
361 </p>
362 </p>
363 </p>
364 </p>
365 </p>
366 </p>
367 </p>
368 </p>
369 </p>
370 </p>
371 </p>
372 </p>
373 </p>
374 </p>
375 </p>
376 </p>
377 </p>
378 </p>
379 </p>
380 </p>
381 </p>
382 </p>
383 </p>
384 </p>
385 </p>
386 </p>
387 </p>
388 </p>
389 </p>
390 </p>
391 </p>
392 </p>
393 </p>
394 </p>
395 </p>
396 </p>
397 </p>
398 </p>
399 </p>
400 </p>
401 </p>
402 </p>
403 </p>
404 </p>
405 </p>
406 </p>
407 </p>
408 </p>
409 </p>
410 </p>
411 </p>
412 </p>
413 </p>
414 </p>
415 </p>
416 </p>
417 </p>
418 </p>
419 </p>
420 </p>
421 </p>
422 </p>
423 </p>
424 </p>
425 </p>
426 </p>
427 </p>
428 </p>
429 </p>
430 </p>
431 </p>
432 </p>
433 </p>
434 </p>
435 </p>
436 </p>
437 </p>
438 </p>
439 </p>
440 </p>
441 </p>
442 </p>
443 </p>
444 </p>
445 </p>
446 </p>
447 </p>
448 </p>
449 </p>
450 </p>
451 </p>
452 </p>
453 </p>
454 </p>
455 </p>
456 </p>
457 </p>
458 </p>
459 </p>
460 </p>
461 </p>
462 </p>
463 </p>
464 </p>
465 </p>
466 </p>
467 </p>
468 </p>
469 </p>
470 </p>
471 </p>
472 </p>
473 </p>
474 </p>
475 </p>
476 </p>
477 </p>
478 </p>
479 </p>
480 </p>
481 </p>
482 </p>
483 </p>
484 </p>
485 </p>
486 </p>
487 </p>
488 </p>
489 </p>
490 </p>
491 </p>
492 </p>
493 </p>
494 </p>
495 </p>
496 </p>
497 </p>
498 </p>
499 </p>
500 </p>
501 </p>
502 </p>
503 </p>
504 </p>
505 </p>
506 </p>
507 </p>
508 </p>
509 </p>
510 </p>
511 </p>
512 </p>
513 </p>
514 </p>
515 </p>
516 </p>
517 </p>
518 </p>
519 </p>
520 </p>
521 </p>
522 </p>
523 </p>
524 </p>
525 </p>
526 </p>
527 </p>
528 </p>
529 </p>
530 </p>
531 </p>
532 </p>
533 </p>
534 </p>
535 </p>
536 </p>
537 </p>
538 </p>
539 </p>
540 </p>
541 </p>
542 </p>
543 </p>
544 </p>
545 </p>
546 </p>
547 </p>
548 </p>
549 </p>
550 </p>
551 </p>
552 </p>
553 </p>
554 </p>
555 </p>
556 </p>
557 </p>
558 </p>
559 </p>
560 </p>
561 </p>
562 </p>
563 </p>
564 </p>
565 </p>
566 </p>
567 </p>
568 </p>
569 </p>
570 </p>
571 </p>
572 </p>
573 </p>
574 </p>
575 </p>
576 </p>
577 </p>
578 </p>
579 </p>
580 </p>
581 </p>
582 </p>
583 </p>
584 </p>
585 </p>
586 </p>
587 </p>
588 </p>
589 </p>
590 </p>
591 </p>
592 </p>
593 </p>
594 </p>
595 </p>
596 </p>
597 </p>
598 </p>
599 </p>
600 </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 ...

Introduction to Information Retrieval

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)

Search results for "painkillers that don't thin blood":

- painkillers that don't thin blood
- 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 69,300 results (0.14 seconds)

Search results for "painkillers that don't upset stomach":

About 29,700 results (0.15 seconds)

Symptoms of Crohn's - 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 ...

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 ...

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. — bonemr ... Perhaps part of the problem is that you don't always have the option to rest your injury if ...

Search results for "class negation synonyms painkillers that don't upset stomach":

About 29,700 results (0.15 seconds)

Symptoms of Crohn's - 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 ...

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 ...

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. — bonemr ... Perhaps part of the problem is that you don't always have the option to rest your injury if ...

Search results for "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 ...
[www.disabled-world.com/.../painkiller-chart.shtml](#) - Cached - Similar

List of Painkillers and what they are used for: Over the Counter...
 File Format: PDF/Adobe Acrobat - Quick View
 Common **adverse effects include**: nausea, dyspepsia, gastrointestinal ulcers/rate/bleeding, ... **upset stomach**, not for last pregnancy trimester. Individuals ...
[healingwithofflaxers.com/.../Painkiller%20side%20effects%20chart%20.pdf](#)

upset stomach ⇒ adverse effect (on stomach)

Introduction to Information Retrieval

Search: tylenol stomach

About 1,390,000 results (0.28 seconds)

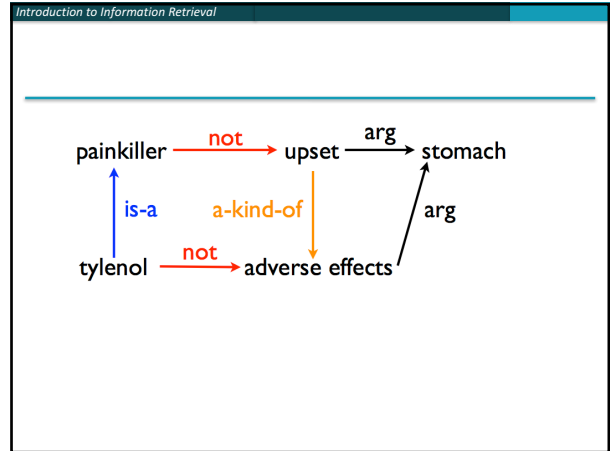
Pain Relievers And Upset Stomach - Tylenol Extra Strength Pain...
 Tylenol Extra Strength Pain Reliever Fever Reducer 100 Caplets - User Rating: 5 stars.
 Review Summary: Taking the wrong pain reliever can cause you **stomach** ...
[www.epinions.com > Wellness & Beauty > Health Aids](#) - Cached - Similar

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/ch078/ch078d.html](#) - Cached - Similar

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

Introduction to Information Retrieval

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)

Introduction to Information Retrieval

bing flights to maui

Flights from Seattle, WA to Maui, HI

Seattle, WA (SEA) - Seatl Kahului, HI (OGG) - Kahul

LEAVE: 03/23/2012 RETURN: 03/25/2012 Find flights

WEEKEND ESTIMATES

SE39 Depart March 09

5748 Depart March 16

5649 Depart March 23

Price prediction

Task completion

bing royal lahaina resort

Royal Lahaina Resort - Bing Travel

2780 Kikasa Dr - Lahaina (808) 661-9469

Details - Photos - Amenities - Directions - Website

Sample rate: \$230

Aggregate ratings

Compare rates from major travel sites

03/15/2012 03/17/2012 Find rate

Task completion

Introduction to Information Retrieval

bing mel gibson movies

Mei Gibson - movies

Mei Gibson - IMDb

Mei Gibson Movies | Movies.com

Videos of mei gibson movies

Mei Gibson - Wikipedia, the free encyclopedia

Mei Gibson filmography - Wikipedia, the free encyclopedia

Introduction to Information Retrieval

Google How old is mariah carey

About 152,000,000 results (0.55 seconds)

45 years (March 27, 1970)

Direct Answer

Structured Data

Mariah Carey

Mariah Carey - Wikipedia, the free encyclopedia

Nick Cannon & Mariah Carey Fighting Split Caused By ...

Mariah Carey - Biography - Music Producer, Film Actress ...

Mariah Carey Divorce Explained: Nick Cannon Cheating ...

Introduction to Information Retrieval

Google how old is christopher manning

About 18,500,000 results (0.64 seconds)

Christopher D. Manning / Age

50 years

September 18, 1965

Christopher Manning, Stanford NLP

Christopher Manning's ex home page

Christopher Manning - Stanford University | Coursera

Introduction to Information Retrieval

Google how old is pandu nayak

About 207,000 results (0.80 seconds)

Pandu Nayak

Pandu Nayak | LinkedIn

Google Panda - Was it named after Pandu Nayak? - Quora

Introduction to Information Retrieval

Patrick Pantel talk (Then) Current experience

bing jetbeam rrt-0

RELATED SEARCHES

ALL RESULTS

JETBeam RRT-0 Review

JETBeam Flashlights

JETBeam RRT-0 R5

JETBeam Raptor

JETBeam MIX

JETBeam

JETBeam RRT-0

JETBeam RRT-0 R5

240 Lumens

SEARCH HISTORY

SEARCH MORE TO SEE YOUR HISTORY

Clear all Turn off

JETBeam RRT-0 Overview With Lots of Pictures

JETBeam Flashlights

Introduction to Information Retrieval

Desired experience: Towards actions

Recognize entity in query

Actions easily accessible

Introduction to Information Retrieval

Politician

Recognize entity in query

Actions easily accessible

Introduction to Information Retrieval

Actions vs. Intents

User Intents and Goals

Query

Query Intent

Finer-grained Intents

Actions on Entities

Introduction to Information Retrieval

Learning actions from web usage logs

- Three months of user web logs
- Annotate with Freebase
- 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 |

Introduction to Information Retrieval

hannah montana

A hodgepodge of related strings. Only actionable through search

Introduction to Information Retrieval

hannah montana

Introduction to Information Retrieval

Big Wins

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

Introduction to Information Retrieval

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
- 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,_and_the_Beautiful
http://.../wiki/Clint_Eastwood
http://.../wiki/Honorary_Academy_of_Art_and_Letters

http://.../wiki/The_Good,_the_Bad,_and_the_Beautiful
<http://.../wiki/Metallica>
[http://.../wiki/Bellagio_\(casino\)](http://.../wiki/Bellagio_(casino))
http://.../wiki/Ennio_Morricone

http://.../wiki/Sergio_Leone
http://.../wiki/The_Good,_the_Bad,_and_the_Beautiful
http://.../wiki/For_a_Few_Dollars_You're_in_the_Gang
http://.../wiki/Ennio_Morricone

Introduction to Information Retrieval

Understanding questions

Chris Watson
The 3rd Prime Minister of Australia

Chris Watson - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Chris_Watson - Wikipedia - 3rd Prime Minister of Australia ... He was the first prime minister from the Australian Labour Party, and the first prime minister from the labour movement in the ... Early life - Union activities and colonial ... - Federation - Prime Minister in 1904

Introduction to Information Retrieval

Search results for "australia's third prime minister":

- Chris Watson - Wikipedia, the free encyclopedia
- List of Prime Ministers of Australia by time in office - Wikipedia
- List of Prime Ministers of Australia - Wikipedia, the free encyclopedia

Introduction to Information Retrieval

Search results for "australia's 3rd prime minister":

| | | | | |
|-----|-----|--------------------------|-------|------------------------|
| 24. | 3rd | Chris Watson (1867-1941) | Labor | Blond, NSW (1861-1926) |
|-----|-----|--------------------------|-------|------------------------|

Additional results include a link to "List of Prime Ministers of Australia by time in office - Wikipedia".

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

45

Introduction to Information Retrieval

Text-based (mainly factoid) QA

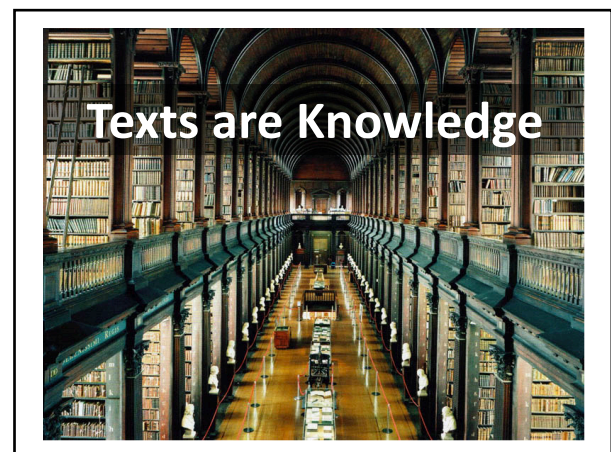
- QUESTION PROCESSING
 - Detect **question type**, **answer type**, focus, relations
 - Formulate queries to send to a search engine
- PASSAGE RETRIEVAL
 - Retrieve ranked documents
 - Break into suitable passages and rerank
- ANSWER PROCESSING
 - Extract candidate answers (as named entities)
 - Rank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

Introduction to Information Retrieval


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

47



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
Microsoft Research

bing 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 - Wikipedia, the free encyclopedia

Who won the best actor Oscar in 1973?

Jack Lemmon - IMDb

Images of Jack Lemmon

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]

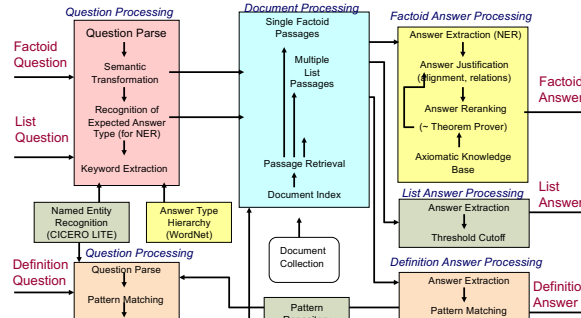
- Assume that there is an underlying alignment
 - Describes which words in and can be associated

See if the (syntactic/semantic) relations support the answer

Introduction to Information Retrieval

Full NLP QA: LCC (Harabagiu/Moldovan)

[below is the architecture of LCC's QA system circa 2003]



```

    graph TD
        subgraph Question_Processing [Question Processing]
            QP[Question Parse] --> ST[Semantic Transformation]
            ST --> R[Recognition of Expected Answer Type for NER]
            R --> KE[Keyword Extraction]
            KE --> NER[Named Entity Recognition CICERO LITE]
            KE --> ATH[Answer Type Hierarchy WordNet]
        end

        subgraph Document_Processing [Document Processing]
            DP[Document Index] --> PR[Passage Retrieval]
            PR --> MLP[Multiple List Passages]
            MLP --> SFP[Single Factoid Passages]
        end

        subgraph Factoid_Answer_Processing [Factoid Answer Processing]
            FAP[Answer Extraction NER] --> AJ[Answer Justification Alignment, relations]
            AJ --> AR[Answer Reranking Theorem Prover]
            AR --> AKB[Axiomatic Knowledge Base]
            AKB --> LAP[List Answer Processing]
            LAP --> AE[Answer Extraction]
            AE --> TC[Threshold Cutoff]
            TC --> FA[Factoid Answer]
        end

        subgraph Definition_Answer_Processing [Definition Answer Processing]
            DAP[Question Parse] --> PM[Pattern Matching]
            PM --> KE2[Keyword Extraction]
            KE2 --> DR[Document Collection]
            DR --> PR2[Passage Retrieval]
            PR2 --> P[Pattern Repository]
            P --> PM2[Pattern Matching]
            PM2 --> AE2[Answer Extraction]
            AE2 --> FA2[Definitior Answer]
        end

        QP --> PR
        R --> PR
        KE --> PR
        MLP --> FAP
        MLP --> DAP
    
```

DrQA: Open-domain Question Answering

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

WIKIPEDIA → ask me anything

55

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) ^{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

56

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

Document Retriever → Document Reader → 833,500

57

(Chen et al, 2017)

Document Retriever

| 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 |

Traditional tf.idf inverted index + efficient bigram hash

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

58

Document Reader: Stanford Attentive Reader (Chen et al, 2016)

Q characters in " @placeholder " movies have gradually become more diverse

Bidirectional LSTMs

characters in " @placeholder " more diverse

59

Stanford Attentive Reader (Chen et al, 2016)

Q characters in " @placeholder " movies have gradually become more diverse

Bidirectional LSTMs

P → p_i

Attention

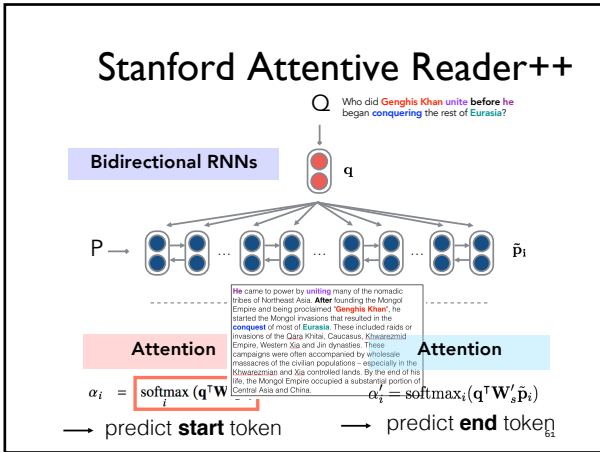
$$\alpha_i = \text{softmax}_i(q^T W_s p_i)$$

$$o = \sum_i \alpha_i p_i$$

$$a = \arg \max_{a \in P \cap E} W_a^T o$$

A entity6

60

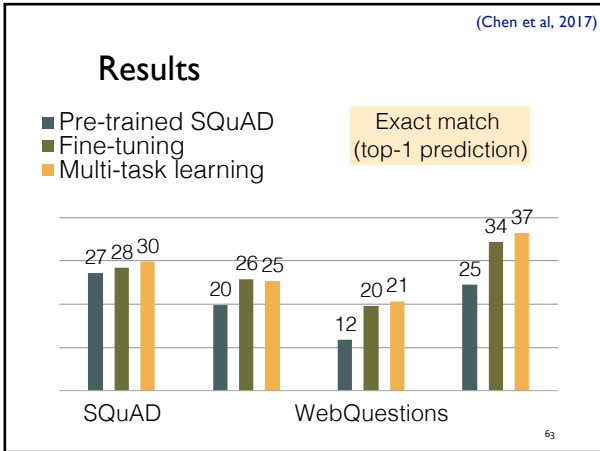


(Chen et al, 2017)

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) | 80.8 |
| Human performance | 91.2 |

62



Demo

Hi!

Hello! Please ask a question.

Who invented LSTM?

Hmmm. Tough one. I would say... Sepp Hochreiter

Where is Stanford University located?

I'm quite confident the answer is... California

64

Demo

What is the highest mountain on Earth?

I'm quite confident the answer is... Mount Everest

What year was the American Declaration of Independence?

I'm quite confident the answer is... 1776

65

Demo

What is the meaning of life?

I'm quite confident the answer is... What is the meaning in asking?

Who will solve AI?

Hmmm. Tough one. I would say... Computers

66

Introduction to Information Retrieval

Thanks everyone! 😊

