

Week 8

Thursday, August 11, 2016 12:29



8.1

08.01 Question Answering

NATURAL LANGUAGE
PROCESSING



NLP

08.01 Question Answering

NATURAL LANGUAGE
PROCESSING



Introduction to NLP

Question Answering

Siri



Ask Jeeves (ask.com)

looks like
this is not
the frontier

Who played Linus in Lost?

Answers **Q&A Community**

Ben Linus Benjamin Ben Linus is a fictional character portrayed by Michael Emerson on the ABC television series *Lost*. Ben was the leader of a group of island natives called the Others and was initially known as Henry Gale, the survivor of Oceanic Flight 815. He began as the mysterious Ben Linus in the first two seasons, but as the series progressed, he becomes something of an uneasy ally to the main characters; even so, however, throughout the series, he is persistently characterized by spontaneous coldblended...

Related Questions:

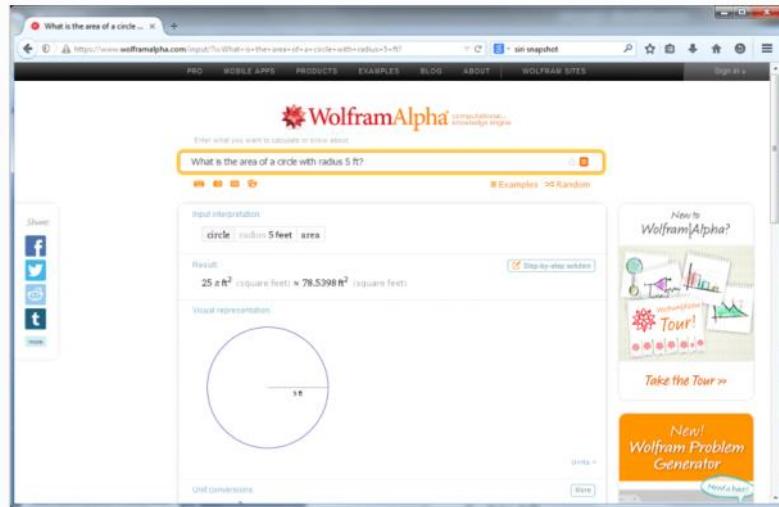
- Spatially speaking, who is better looking: Ben Linus, Henry Gale, or Michael Emerson?
- Order of importance of these Lost Characters? In my opinion it's 1. Jack Sheppard 2. Kate 3. Ben Linus 4. John Locke 5. James Sawyer Ford 6. Hurley 7. Desmond 8. Juliet 9. Sun 10. Claire

View more Q&A »

Michael Emerson - IMDb
www.imdb.com/title/tt0256237/
Michael Emerson, Actor: *Lost*, *Michael* ... He is an actor, known for *Lost* (2004), *Person of Interest* (2011) and *Saw* (2004). He has ... Narrator / Benjamin Linus.

Roger Linus - Lostpedia - The Lost Encyclopedia - Wikia
lostpedia.fandom.com/wiki/Roger_Linus
Roger Linus was the father of Benjamin Linus, whom he regularly abused, both verbally and ... Portrayed by ... Outside references to *Lost*/Video games.

WolframAlpha



IBM's Watson



<http://www.geekwire.com/2013/ibm-takes-watson-cloud/>

Watson on Jeopardy

- Sample questions

- On December 8, 2008 this national newspaper raised its newsstand price by 25 cents to \$1
 - *USA Today*
- In 2010 this former first lady published the memoir "Spoken From the Heart"
 - *Laura Bush*
- This person is appointed by a testator to carry out the directions & requests in his will
 - *Executor*
- Familiarity is said to breed this, from the Latin for "Despise"
 - *Contempt*
- As of 2010, Croatia & Macedonia are candidates but this is the only former Yugoslav republic in the EU
 - *Slovenia*
- The ancient "Lion of Nimrud" went missing from this city's national museum in 2003 (along with a lot of other stuff)
 - *Baghdad*
- It's just a bloody nose! You don't have this hereditary disorder once endemic to European royalty
 - *Haemophilia*
- It's Michelangelo's fresco on the wall of the Sistine Chapel, depicting the saved and the damned
 - *The Last Judgement*
- She "Died in the church and was buried along with her name. Nobody came"
 - *Eleanor Rigby*
- A camel is a horse designed by this
 - *Committee*

Watson Videos

- IBM's Watson Supercomputer Destroys Humans in Jeopardy | Engadget
 - https://www.youtube.com/watch?v=WFR3lOm_xhE
- "What is IBM Watson?" 7 Videos from the Jeopardy! Era
 - <http://mentalfloss.com/article/51543/what-ibm-watson-7-videos-jeopardy-era>

People Ask Questions Online

- Excite corpus
 - 2,477,283 queries (one day's worth)
- 8.4% of them are questions
 - 43.9% factual (what is the country code for Belgium)
 - 56.1% procedural (how do I get out of debt) or other

interesting data

The Excite Corpus

*Multiple
answer*

- In what year did baseball become an offical sport?
- How do I get out of debt?
- Where can I found out how to pass a drug test?
- When is the Super Bowl?
- Who is California's District State Senator? *location, time*
- Where can I buy extra nibs for a fountain pen?
- What time is it in west samoa?
- Where can I buy a little kitty cat?
- What are the symptoms of attention deficit disorder?
- Where can I get some information on Michael Jordan?
- How does the character Seyavash in Ferdowsi's Shahnameh exhibit characteristics of a hero?
- When did the Neanderthal man live?
- Which Frenchman declined the Nobel Prize for Literature for ideological reasons?



Murax Questions

*Answer questions
based on PDFs*

- What U.S. city is at the junction of the Allegheny and Monongahela rivers?
- Who wrote “Across the River and into the Trees” ?
- Who married actress Nancy Davis?
- What’s the capital of the Netherlands?
- Who was the last of the Apache warrior chiefs?
- What chief justice headed the commission that declared: “Lee Harvey Oswald . . . acted alone.”?
- What famed falls are split in two by Goat Island?
- What is November’s birthstone?
- Who’s won the most Oscars for costume design?
- What is the state flower of Alaska?

Julian Kupiec 1993. MURAX: A Robust Linguistic Approach For Question Answering

AOL Corpus Questions

- what does cerebral cortical atrophy mean
- what fraction is closest to pi
- what is the highest calories consumed by a person in a 24 hor period
- what happens to limewater in the presence of carbon dioxide
- what language is Numa Numa in
- who discovered Saturn
- who discovered the gene for Huntington's disease
- who is the daughter of indian sitar maestro Ravi Shankar
- who invented marshmallow sandwich cookies
- who invented lithograph
- who was the first Michigan football player to put on the cover of a video game
- who invented smallpox vaccine
- when was the berlin wall torn down
- when was yahoo launched
- when was the wedding march by Mendelssohn composed
- when was the declaration of independence signed
- when did France become a republic
- when was Alexander the Great born
- which fish contains low mercury
- when is the youngest a minor can move out of their parents house in Texas
- when should take a child to see an psychologist
- what hormone imbalance could mimic pregnancy symptoms

Question Type Taxonomy

- **Question Types**
 - Yes/no vs. wh-
 - Factual vs. procedural
 - Single answer vs. multiple answers
 - Objective vs. subjective
 - Context-specific (personalized) vs. generic
 - Known answer in the collection, y/n

About evaluation

State of the Art

- **Question types**
 - Mostly factual, short-answer questions
- **System architecture**
 - IR component
 - Statistical approaches using lots of data
 - Relatively little knowledge

Older Systems (pre-TREC)

References

- BASEBALL (Green et al., 1961)
 - Baseball
- ELIZA (Weizenbaum 1966)
 - Dialogue, Therapist
- SHRDLU (Winograd 1972)
 - Blocks world
- LUNAR (Woods, 1973)
 - Lunar rocks
- Murax (Kupiec 1993) *general domain.*
 - Open-domain, encyclopedia-based
- START (Katz 1997)
 - Open-domain, web-based
- Deep Red (Hirschman & al. 1999)
 - Reading comprehension
- Jupiter (Zue et al., 2000)
 - Spoken questions

Eliza (Weizenbaum 1966)

- Eliza plays a therapist
 - and doesn't really answer questions ☺
- Simple pattern matching
- Sample dialogue

E: I am the psychotherapist. Please, describe your problems.
 P: I am always tired.
 E: Why do you say you are always tired?
 P: Because I cannot sleep.
 E: Is it because you cannot sleep that you came to me?
 P: Yes, help me please.
 E: Earlier you said you cannot sleep?

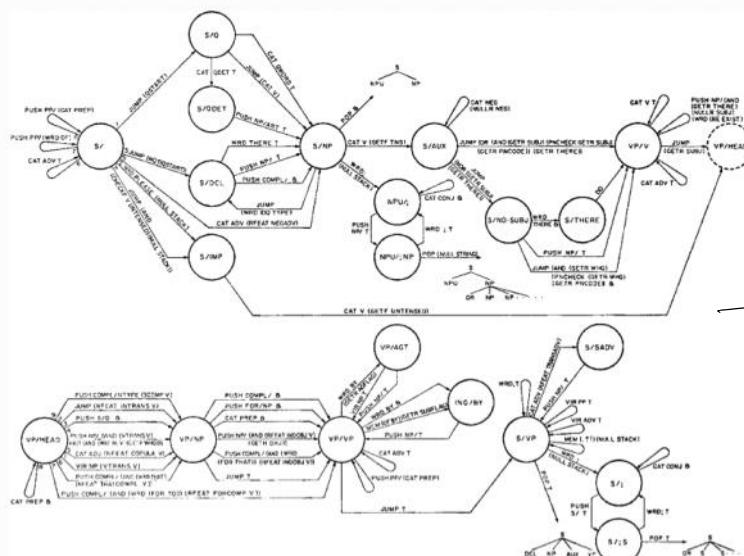
clarification

Lunar (Woods et al., 1972)

- Lunar Sciences Natural Language Information System
 - Apollo 11 moon rocks ← about very restricted domain
 - Questions in English
 - Components
 - a general-purpose ATN (augmented transition network) grammar for English
 - a procedural-semantics framework for semantic interpretation
 - mapping from the syntactic grammar to a logic representation
 - 3,500 word dictionary
 - 13,000 entries in the knowledge base
 - Lunar answered 78% of sentences asked by geologists at 1971 Lunar Rocks conference
 - How many breccias contain olivine?
 - What are they?Tsample.

J Sample.

Lunar Grammar Fragment



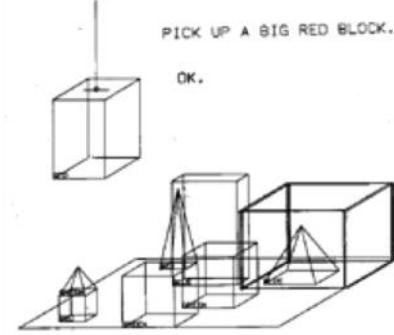
Woods et al
1972

→ automata
state
transition
initial state
final state
alphabet

SHRDLU (Winograd 1972)

- Simple questions about the blocks world

- Person: Pick up a big red block.
- Computer: OK.
- Person: Grasp the pyramid.
- Computer: I don't understand which pyramid you mean.
- Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
- Computer: By "it", I assume you mean the block which is taller than the one I am holding.
- Computer: OK.
- Person: What does the box contain?
- Computer: The blue pyramid and the blue block.
- Person: What is the pyramid supported by?
- Computer: The box.
- Person: How many blocks are not in the box?
- Computer: Four of them.
- Person: Is at least one of them narrower than the one which I told you to pick up?
- Computer: Yes, the red cube.



<http://hci.stanford.edu/~winograd/shrdlu/>

Start (Katz 1997)

- Open-domain
- Uses the Web
- Online since 1993

The START Natural Language Question Answering System has been on-line and continuously operating since December, 1993. It has been developed by Boris Katz and his associates of the InfoLab Group at the MIT Computer Science and Artificial Intelligence Laboratory. Unlike information retrieval systems (e.g., search engines), START aims to supply users with "just the right information," instead of merely providing a list of hits. Currently, the system can answer millions of English questions about places (e.g., cities, countries, lakes, coordinates, weather, maps, demographics, political and economic systems), movies (e.g., titles, actors, directors), people (e.g., birth dates, biographies), dictionary definitions, and much, much more. Below is a list of some of the things START knows about, with example questions. You can type your question above or select from the following examples. [less...](#)

Geography

- What South-American country has the largest population?
- What's the largest city in Florida?
- Give me the states that border Colorado.
- What cities are within 250 miles of the capital of Italy?
- How many people live in Israel?
- Show me a map of Denmark.
- Which is deeper, the Baltic Sea or the North Sea?
- How far is Mount Kilimanjaro from Mount Everest?
- List the three cities in Argentina.
- Show the capital of the 2nd largest country in Asia.
- How much does it cost to study at MIT?
- More examples.

Science and Reference

- What is Jupiter's atmosphere made of?
- Who first discovered radiocarbon dating?
- How far is Neptune from the sun?
- Why is the sky blue?
- What planet has the smallest surface area?
- How many feet are there in a kilometer?
- Convert 100 dollars into Euros.
- Show me a metro map of Moscow.
- How many languages are spoken in Afghanistan?
- Give me the GDP of Taiwan.
- How is the weather in Boston today?
- More examples.

[HOME](#) [People](#) [Publications](#) [How START works](#)

<http://start.csail.mit.edu>



NLP



8.2



NLP

Introduction to NLP

Evaluation of QA

The TREC Q&A evaluation

- Run by NIST
 - Voorhees and Tice 1999, 2000
 - 2GB of input
 - AQUAINT corpus
 - 200 questions in 1999
 - 693 questions in 2000
 - Essentially fact extraction
 - Who was Lincoln's secretary of state?
 - What does the Peugeot company manufacture?
 - Assumptions
 - Questions are based on text
 - Answers are assumed to be present
 - Systems return five passages of 50 or 250 bytes
 - After 2002, only a single passage of 50 bytes + confidence score + NIL
 - No inference needed
- confidence score*

TREC 1999

- What date in 1989 did East Germany open the Berlin Wall?
– Nov 9
- Who was Johnny Mathis' high school track coach?
– Lou Vasquez
- What is the shape of a porpoises' tooth?
– spade-shaped
- What is the number of buffaloes thought to have been living in North America when Columbus landed in 1492?
– 60 million
- The Faroes are a part of what northern European country?
– Denmark
- The symptoms of Parkinson's disease are linked to the demise of cells in what area of the brain?
– substantia nigra
- What was the date of the Chelsea flower show in 1992?
– May 18
- What hotel was used for a setting of the Agatha Christie novel, "And Then There Were None"?
– Burgh Island Hotel
- What year was the Magna Carta signed?
– 1215
- Who was Lincoln's Secretary of State?
– William Seward
- How long does it take to fly from Paris to New York in a Concorde?
– 3 1/2 hours

TREC 1999 – Test Questions

- Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?
- What was the monetary value of the Nobel Peace Prize in 1989?
- What does the Peugeot company manufacture?
- How much did Mercury spend on advertising in 1993?
- What is the name of the managing director of Apricot Computer?
- Why did David Koresh ask the FBI for a word processor?
- What debts did Qintex group leave?
- What is the name of the rare neurological disease with symptoms such as: involuntary movements (tics), swearing, and incoherent vocalizations (grunts, shouts, etc.)?
- How far is Yaroslavl from Moscow?
- Name the designer of the shoe that spawned millions of plastic imitations, known as "jellies".
- Who was President Cleveland's wife?
- How much did Manchester United spend on players in 1993?
- How much could you rent a Volkswagen bug for in 1966?
- What country is the biggest producer of tungsten?
- When was London's Docklands Light Railway constructed?
- What two US biochemists won the Nobel Prize in medicine in 1992?
- How long did the Charles Manson murder trial last?
- Who was the first Taiwanese President?
- Who was the leader of the Branch Davidian Cult confronted by the FBI in Waco, Texas in 1993?
- Where is Inoco based?
- Who was the first American in space?
- When did the Jurassic Period end?
- When did Spain and Korea start ambassadorial relations?
- When did Nixon visit China?
- Who was the lead actress in the movie "Sleepless in Seattle"?
- What is the name of the "female" counterpart to El Nino, which results in cooling temperatures and very dry weather?
- What is the brightest star visible from Earth?
- What are the Valdez Principles?



Evaluation

→ ranking
all q's. MRR ?

$$\text{MRR} = \frac{1}{n} \left(\sum_{i=1}^n \frac{1}{\underline{\text{rank}_i}} \right)$$

Average
over all
q's

highest rank
of next
answer

- **MRR**

- Mean Reciprocal Rank
- Introduced by TREC in 1999

- **Example**

- What is the capital of Canada?
- 1.Toronto, 2.Ottawa, 3.Albany, 4.Philadelphia, 5.Ottawa
- Correct answer ranks = 2, 5 (highest rank = 2)
- MRR = 1/2 = 0.5
- TRR = 1/2 + 1/5 = 0.7 = total reciprocal rank

- **Later years**

- Confidence-weighted score

Performance

- **1999**
 - Cymphony, SMY, AT&T, IBM, XRCE, Umd
- **2000**
 - SMU, ISI, Waterloo, IBM, CUNY-QC
- **2001**
 - InsightSoft, LCC, Oracle, ISI
- **2002**
 - LCC, InsightSoft, NUS
- **2003**
 - LCC, NUS, LexiClone, ISI
- **2004**
 - LCC, NUS, Uwales, IBM
- **2005**
 - LCC, NUS, IBM, Albany

Other Types of Questions

- **Definitional**
 - What is a boll weevil?
- **List**
 - Which states signed the US Declaration of Independence?
- **Crosslingual**
 - E.g., questions in Spanish, documents in English
- **Series**
 - Next slide

TREC Series Questions

- What are prions made of?
- Who discovered prions?
- What researchers have worked with prions?
- Who is the lead singer/musician in Nirvana?
- Who are the band members?
- When was the band formed?
- What is their biggest hit?
- What are their albums?
- What style of music do they play?
- What industry is Rohm and Haas in?
- Where is the company located?
- What is their annual revenue?
- How many employees does it have?
- What kind of insect is a boll weevil?
- What type of plant does it damage?
- What states have had problems with boll weevils?



NLP

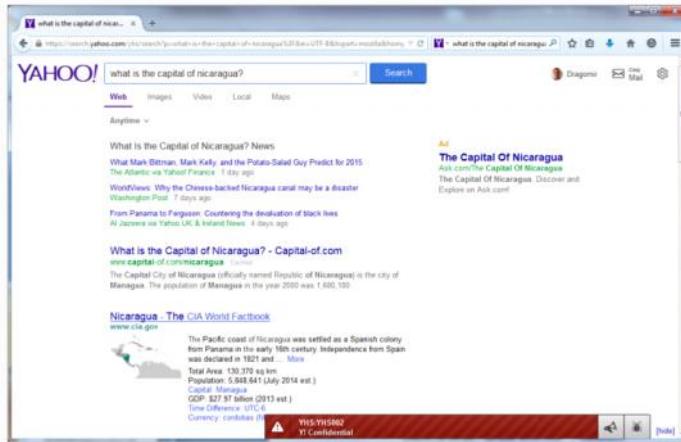


Introduction to NLP

System Architecture

System Architecture

- Many questions can be answered by traditional search engines



What is the Largest City in Northern Afghanistan?

... Afghanistan, **Kabul**, 2,450 ... Administrative capital and **largest city** (1997 est ... Undetermined. Panama, **Panama City**, 450,668. ... of the Gauteng, **Northern** Province, Mpumalanga ... www.infoplease.com/cgi-bin/id/A0855603

... died in **Kano**, northern Nigeria's **largest city**, during two days of anti-American riots led by Muslims protesting the US-led bombing of **Afghanistan**, according to ... www.washingtonpost.com/wp-dyn/print/world/

... air strikes on the **city** ... the Taliban militia in **northern Afghanistan** in a significant blow ... defection would be the **largest** since the United States ... www.afgha.com/index.php - 60k

... **Kabul** is the capital and **largest city of Afghanistan**. met. area pop. 2,029,889), is the **largest city** in Uttar Pradesh, a state in **northern** India. school.discovery.com/homeworkhelp/worldbook/atozgeography/k/k1menu.html

... **Gudermes**, Chechnya's **second largest** town. The attack ... location in **Afghanistan's** outlying regions ... in the **city of Mazar-i-Sharif**, a **Northern** Alliance-affiliated ... english.pravda.ru/hotspots/2001/09/17/

... Get Worse By RICK BRAGG Pakistan's **largest city** is getting a jump on the ... Region: Education Offers Women in **Northern Afghanistan** a Ray of Hope. ... www.nytimes.com/pages/world/asia/

... within three miles of the airport at **Mazar-e-Sharif**, the **largest city in northern Afghanistan**, held since 1998 by the Taliban. There was no immediate comment ... uk.fc.yahoo.com/photos/a/afghanistan.html

System Components

- Source identification
 - semi-structured vs. text sources
- Query modulation
 - best paraphrase of a NL question given the syntax of a search engine
 - Example: Who wrote Hamlet → author | wrote Hamlet
- Document retrieval
- Sentence ranking
- Answer extraction
- Answer ranking
 - question type classification
 - phrase chunking

What is the largest city in Northern Afghanistan?

Query modulation

(largest OR biggest) city "Northern Afghanistan"

Document retrieval

www.infoplease.com/cgi-bin/id/A0855603
www.washingtonpost.com/wp-dyn/print/world/

Sentence ranking

Gudermes, Chechnya's second largest town ... location in Afghanistan's outlying regions
 within three miles of the airport at Mazar-e-Sharif, the largest city in northern Afghanistan

Answer extraction

Gudermes
 Mazer-e-Sharif

Answer ranking

Mazer-e-Sharif
 Gudermes

Use ml to rank answers

Question Type Classification

- Can help find the right answers in the text
- Example
 - Who wrote Anna Karenina?
 - Looking for a PERSON/INDIVIDUAL/WRITER

SYN-classes (IBM AnSel)

QA-token	Question type	Example
PLACE	Where?	In the Rocky Mountains
COUNTRY	Where? What country?	United Kingdom
STATE	Where? What state?	Massachusetts
PERSON	Who?	Albert Einstein
ROLE	Who?	Doctor
NAME	Who? What? Which?	The Shakespeare Festival
ORG	Who? What?	The U.S. Post Office
DURATION	How long?	For 5 centuries
AGE	How old?	30 years old
YEAR	When? What year?	1999
TIME	When? What time?	In the afternoon
DATE	When? What date?	July 4 th , 1776
VOLUME	How big? How large?	3 gallons
AREA	How big? How large?	4 square inches
LENGTH	How long? How big?	3 miles
WEIGHT	How heavy? How big?	25 tons
NUMBER	How many?	1,134.5
RATE	How much? What percentage?	50 per cent
MONEY	How much?	4 million dollars

UIUC Question Types

- **ENTITY:** entities
 - animal: animals
 - body: organs of body
 - color: colors
 - creative: inventions, books and other creative pieces
 - currency: currency names
 - dis.med.: diseases and medicine
 - event: events
 - food: food
 - instrument: musical instrument
 - lang: languages
 - letter: letters like a-z
 - other: other entities
 - plant: plants
 - product: products
 - religion: religions
 - sport: sports
 - substance: elements and substances
 - symbol: symbols and signs
 - technique: techniques and methods
 - term: equivalent terms
 - vehicle: vehicles
 - word: words with a special property
- **ABBREVIATION:** abbreviation
 - abb abbreviation
 - exp expression abbreviated
- **DESCRIPTION:** description and abstract concepts
 - definition: definition of sth.
 - description: description of sth.
 - manner: manner of an action
 - reason: reasons
- **HUMAN:** human beings
 - group: a group or organization of persons
 - ind: an individual
 - title: title of a person
 - description: description of a person
- **LOCATION:** locations
 - city: cities
 - country: countries
 - mountain: mountains
 - other: other locations
 - state: states
- **NUMERIC:** numeric values
 - code: postcodes or other codes
 - count: number of sth.
 - date: dates
 - distance: linear measures
 - money: prices
 - order: ranks
 - other: other numbers
 - period: the lasting time of sth.
 - percent: fractions
 - speed: speed
 - temp: temperature
 - size: size, area and volume
 - weight: weight

UIUC Examples

- **NUM:date**
 - When did Rococo painting and architecture flourish ?
- **LOC:country**
 - What country 's national passenger rail system is called Via ?
- **HUM:ind**
 - Who invented Make-up ?
- **DESC:desc**
 - What is the origin of the word `` attic '' ?
- **DESC:desc**
 - What did Delilah do to Samson 's hair ?
- **ENTY:animal**
 - What kind of animals were in the Paleozoic era ?
- **HUM:ind**
 - Which of the following was Rhodes Scholar ?
- **HUM:ind**
 - Who comprised the now-defunct comic book team known as the Champions ?
- **DESC:**
 - manner How do you make a paintball ?
- **LOC:**
 - state What U.S. state is Fort Knox in ?
- **ENTY:**
 - animal What is a female rabbit called ?
- **LOC:mount**
 - Where is the highest point in Japan ?
- **DESC:desc**
 - Where do chihuahuas come from ?
- **LOC:other**
 - Where does Barney Rubble go to work after he drops Fred off in the `` Flintstones '' cartoon series?

UIUC Papers On Question Classification

question classification

- Xin Li and Dan Roth. Experimental Data for Question Classification
 - <http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/QC/>
- Xin Li, Dan Roth. Learning Question Classifiers: The Role of Semantic Information
 - <http://l2r.cs.uiuc.edu/~danr/Papers/LiRo05a.pdf>
 - <http://l2r.cs.uiuc.edu/~cogcomp/asoftware.php?skey=LBI>
- Data set
 - training (http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/QC/train_5500.label)
 - test (http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/QC/TREC_10.label)

NLP



Tech for question classification

NLP

Techniques for Question Classification

- Classification task
 - Use standard techniques
- Regular expressions
 - WHO is|was -> PERSON
 - WHICH STATE -> STATE

Query Formulation

- Aka Query Modulation
 - Determine which words to include in the IR query
- Natural language
information query
↓ step-by-step operation*

No.	Operator	Paraphrase	IRDR
0	IDENTITY	What country is the "biggest producer" of tungsten	0.837
1	DEL_WH	country is the "biggest producer" of tungsten	1.755
2	DEL_AUX	What country is the "biggest producer" of tungsten	1.322
3	DEL_ART	What country is "biggest producer" of tungsten	1.436
4	DEL_PREP	What country is the "biggest producer" tungsten	1.436
5	DEL_STOP	What country is the "biggest producer" of tungsten	0.837
6	REPL_IN	What (" administrative district " OR " political unit " " OR people OR region OR " geographical area " OR " rural area ") is the "biggest producer" of tungsten	1.181
7	REPL_2N	What country is the "biggest producer" of (" metallic element " OR wolfram OR w OR " atomic number 74 ")	1.419

Passage Retrieval

- Features
 - Proper nouns that match the query
 - Near each other
 - Entities that match the expected answer type

Answer Retrieval

- Use NER to identify the matching phrases
 - E.g., “January 1, 1951” as a DATE
- Features
 - Distance to query words
 - Answer type
 - Wordnet similarity
 - Redundancy

Redundancy

- *What is the capital of Spain?*
 - *Madrid* is the capital of Spain → redundant
 - En route to Spain's capital of Madrid
 - Madrid, Spain's capital city is situated almost at the geographical epicentre of the country
 - The capital of Spain is Madrid
 - Madrid, Spain's sunny capital
 - Madrid became Spain's capital
 - In 1561, it was elevated to status as Spain's capital city
 - Madrid has been the capital of Spain since 1562
 - Madrid, the physical and cultural capital of Spain

Can we
filtering
answers

Redundancy

- When did French revolutionaries storm the Bastille?
 - The storming of the Bastille occurred in Paris on the morning of 14 July 1789
 - The storming of the Bastille, 14 July 1789
 - The storming of the Bastille prison on July 14th 1789 was an event that paved the way to further civil disorder and upheaval in France.
 - French revolutionaries storm Bastille, 1789
 - The storming of the Bastille (Louis XVI's prison) by French revolutionaries took place on July 14th, 1789
- Who killed Mahathma Gandhi?
 - Mohandas Karamchand Gandhi (often called Mahatma Gandhi) was assassinated on 30 January 1948, shot at point-blank range by Nathuram Godse.
 - Nathuram Godse killed Gandhi
 - Godse killed Gandhi
 - Mahatma Gandhi was assassinated on 30 January 1948, shot at point-blank range by Nathuram Godse
 - Mohandas Gandhi was shot dead by Nathuram Godse, a Hindu extremist
 - Godse assassinated Mahatma Gandhi on January 30, 1948, approaching him during the evening prayer, bowing, and shooting him three times at close range with a Beretta semi-automatic pistol.

On the Web

- Significantly larger corpus
- No pre-annotation is possible
- Search engines only partially helpful
 - Stop words
 - Question types
 - Restrictions on queries
 - Issues with reliability, timeliness, inaccurate answers

*← Distinction
↑
corpus based sys*

good answer valid now



NLP



8.4



NLP

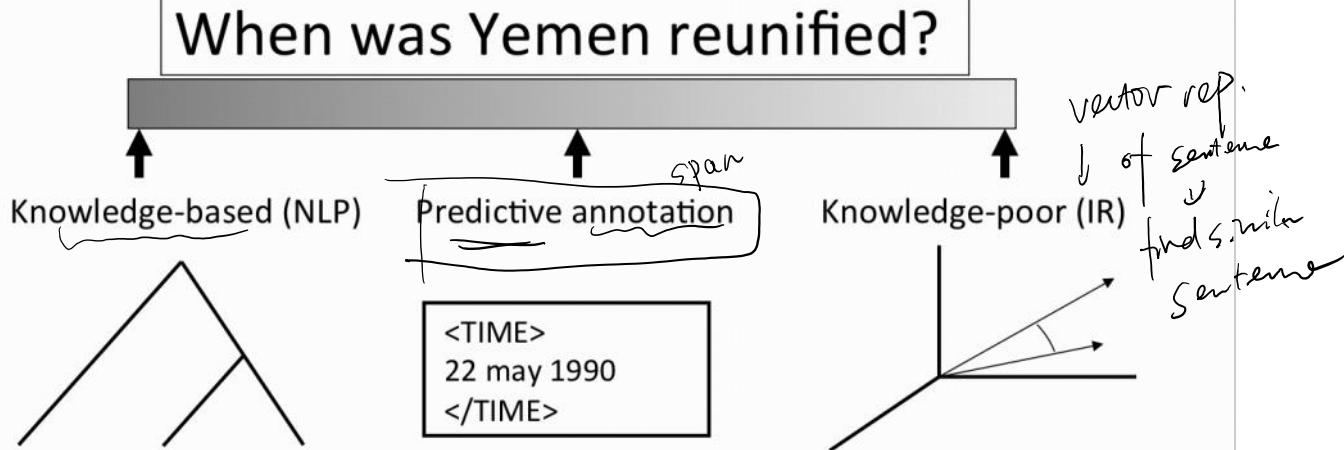
Introduction to NLP

Question Answering Systems

AnSel (Prager et al. 1999)

- IBM System
- Built for TREC
- Components
 - Predictive Annotation
 - Logistic Regression

Predictive Annotation



Predictive Annotation

```

<p><NUMBER>1</NUMBER></p>

<p><QUERY>Who is the author of the book, "The Iron Lady: A Biography of Margaret Thatcher"?</QUERY></p>

<p><PROCESSED_QUERY>@excwin(*dynamic* @weight(2.00001001 *Iron_Lady) @weight(200
Biography_of_Margaret_Thatcher) @weight(200 Margaret) @weight(100 author) @weight(100 book)
@weight(100 iron) @weight(100 lady) @weight(100 :) @weight(100 biography) @weight(100
thatcher) @weight(400 @syn(PERSON$ ORG$ NAME$ ROLE$) ) )</PROCESSED_QUERY></p>

<p><DOC>LA090290-0118</DOC></p>

<p><SCORE>1020.8114</SCORE></p>

<TEXT><p>THE IRON LADY; A <span class="NAME"> Biography of Margaret Thatcher </span> by
<span class="PERSON"> Hugo Young </span> (<span class="ORG"> Farrar , Straus & Giroux </span> )
The central riddle revealed here is why, as a woman <span class="PLACEDEF"> in a
man </span> 's world, <span class="PERSON"> Margaret Thatcher </span> evinces such an
exclusionary attitude toward women.</p></TEXT>

```

Annotat^{ed}

Some Observations

- In documents that contain the answers, the query terms tend to occur in close proximity to each other
- The answers to fact-seeking questions are usually phrases
- These phrases can be categorized by question type
- The phrases can be identified in text by pattern matching techniques

Feature Selection

Average
↓ distances

Avgdst: the average distance in words between the beginning of the span and the words in the query that also appear in the passage. Example: given the question "Who was Johnny Mathis' high school track coach?" and the passage **Tim O'Donohue**, **Woodbridge High School's varsity baseball coach, resigned Monday and will be replaced by assistant Johnny Ceballos, Athletic Director Dave Cowen said.**" and the span **Tim O'Donohue**, the value of avgdst is equal to 8.

Notinq: the number of words in the span that do not appear in the query. Example: Notinq ("Woodbridge high school") = 1, because both "high" and "school" appear in the query while "Woodbridge" does not. It is set to -100 when the actual value is 0.

Frequency: number of times a given span appears in the hit list.

Sscore: passage relevance as computed by the search engine.

search engine rel. Score

Number: position of the span among all spans returned. Example: "Lou Vasquez" was the first span returned by GuruQA on the sample question.

Rspanno: position of the span among all spans returned within the current passage.

relative span #

Count: number of spans of any span class retrieved within the current passage.

Type: the position of the span type in the list of potential span types. Example: Type ("Lou Vasquez") = 1, because the span type of "Lou Vasquez", namely "PERSON" appears first in the list of potential span types, "PERSON ORG NAME ROLE".

there should be
another list
of types

Span	Type	Number	Rspanno	Count	Notinq	Type	Avgdst	Sscore	TOTAL
Lou Vasquez	PERSON	1	1	6	2	1	16	0.02507	-9.93
Tim O'Donohue	PERSON	17	1	4	2	1	8	0.02257	-12.57
Athletic Director Dave Cowen	PERSON	23	6	4	4	1	11	0.02257	-15.87
Johnny Ceballos	PERSON	22	5	4	1	1	9	0.02257	-19.07
Civic Center Director Martin Durham	PERSON	13	1	2	5	1	16	0.02505	-19.36
Johnny Hodges	PERSON	25	2	4	1	1	15	0.02256	-25.22
Derrick Evans	PERSON	33	4	4	2	1	14	0.02256	-25.37
NEWSWIRE Johnny Majors	PERSON	30	1	4	2	1	17	0.02256	-25.47
Woodbridge High School	ORG	18	2	4	1	2	6	0.02257	-28.37
Evan	PERSON	37	6	4	1	1	14	0.02256	-29.57
Gary Edwards	PERSON	38	7	4	2	1	17	0.02256	-30.87
O I Simpson	NAME	3	3	2	2	2	13	0.02507	-27.40

Span	Type	Number	Rspanno	Count	Notinq	Type	Avgdst	Sscore	TOTAL
<i>Lou Vasquez</i>	PERSON	1	1	6	2	1	16	0.02507	-9.93
Tim O'Donohue	PERSON	17	1	4	2	1	8	0.02257	-12.57
Athletic Director Dave Cowen	PERSON	23	6	4	4	1	11	0.02257	-15.87
Johnny Ceballos	PERSON	22	5	4	1	1	9	0.02257	-19.07
Civic Center Director Martin Durham	PERSON	13	1	2	5	1	16	0.02505	-19.36
Johnny Hodges	PERSON	25	2	4	1	1	15	0.02256	-25.22
Derrick Evans	PERSON	33	4	4	2	1	14	0.02256	-25.37
NEWSWIRE Johnny Majors	PERSON	30	1	4	2	1	17	0.02256	-25.47
Woodbridge High School	ORG	18	2	4	1	2	6	0.02257	-28.37
Evan	PERSON	37	6	4	1	1	14	0.02256	-29.57
Gary Edwards	PERSON	38	7	4	2	1	17	0.02256	-30.87
O.J. Simpson	NAME	2	2	6	2	3	12	0.02507	-37.40
South Lake Tahoe	NAME	7	5	6	3	3	14	0.02507	-40.06
Washington High	NAME	10	6	6	1	3	18	0.02507	-49.80
Morgan	NAME	26	3	4	1	3	12	0.02256	-52.52
Tennesseefootball	NAME	31	2	4	1	3	15	0.02256	-56.27
Ellington	NAME	24	1	4	1	3	20	0.02256	-59.42
assistant	ROLE	21	4	4	1	4	8	0.02257	-62.77
the Volunteers	ROLE	34	5	4	2	4	14	0.02256	-71.17
Johnny Mathis	PERSON	4	4	6	-100	1	11	0.02507	-211.33
Mathis	NAME	14	2	2	-100	3	10	0.02505	-254.16
coach	ROLE	19	3	4	-100	4	4	0.02257	-259.67

IONAUT (Abney et al. 2000)

- Passage retrieval
 - Uses START (Salton, Buckley)
- Entity recognition
- Uses Cass (Abney) – partial parser
- Entity classification
- Simple patterns for 8 question types

Mulder (Kwok et al. 2001)

- First large-scale Web QA system
- Components
 - Maximum entropy parser (Charniak)
 - PC-Kimmo for unknown words
 - Link parser (Sleator and Temperley)

- maximum entropy parser (Charniak)
- PC-Kimmo for unknown words
- Link parser (Sleator and Temperley)
- Google 

- **Tokenization**
 - phrases in quotes
- **Query transformations**

– “When did Nixon visit China” -> “Nixon visited China”

transform



Summarization

NSIR (Radev et al. 2002)

- **Probabilistic phrase reranking**
 - $P(qtype|signature)$
 - Signature = POS sequence (e.g., “NNP NNP” for “Bill Gates”)
- **Search engines**
 - AlltheWeb, NorthernLight, Altavista, Google

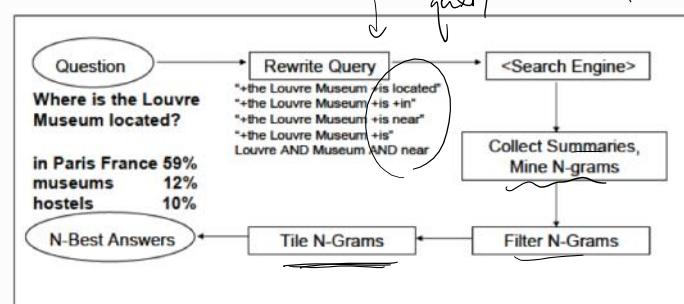
AskMSR (Banko et al. 2002)

Microsoft Research

- **Assumption**
 - Someone has already answered this question on the Web
- **Components**
 - Query rewriting
 - Snippet retrieval
 - N-gram ranking
- **Tiling matches**

Combining A B C and B C D into A B C D

- E.g., “Mr. Charles” and “Charles Dickens” into “Mr. Charles Dickens”



tiling

Echihabi and Marcu

- Based on the noisy-channel model
- Find the sentence S that maximizes $p(q|S)$ *max probability given the question sentence (q)*
- Requires simplifying the sentences

NLP



NLP

LCC (e.g. Moldovan et al. 2003)

- Uses logic form transformations
- Uses axioms for inference
 - e.g. Lexical chains

Example:

Heavy selling of Standard & Poor's 500-stock index futures in Chicago relentlessly beat stocks downward.

LF:

```
heavyJJ(x1) & selling_NN(x1) & of_IN(x1,x6) &
Standard_NN(x2) & &_CC(x13,x2,x3) & Poor_NN(x3)
& s_POS(x6,x13) & 500-stockJJ(x6) & index_NN(x4)
& future_NN(x5) & nn_NNC(x6,x4,x5) & in_IN(x1,x8)
& Chicago_NN(x8) & relentlessly_RB(e12) &
beat_VB(e12,x1,x9) & stocks_NN(x9) & downward_RB(e12)
```

Q1394: What country did the game of croquet originate in?

Answer: Croquet is a 15th-century French sport that has largely been dominated by older, wealthier people who play at exclusive clubs.

Lexical chains:

- (1) game:n#3 → HYPERNYM → recreation:n#1 → HYPONYM → sport:n#1
- (2) originate.in:v#1 → HYPONYM → stem:v#1 → GLOSS → origin:n#1 → GLOSS → be:v#1

QASM (Radev & al. 2001)

Noisy channel model

- Noisy channel model
- Convert natural language question into query
 - What country is the biggest producer of tungsten?
 - (biggest OR largest) producer tungsten
- Channel operators
 - DELETE, e.g., delete prepositions, stop words
 - REPLACE, e.g., replace a noun phrase with a WordNet expansion
 - DISJUNCT, e.g., replace a noun phrase with a disjunction

Ravinchandran and Hovy 2002

- Characteristics
 - Automatically learn surface patterns
 - Starts with a seed
 - Query Web
 - Find patterns that contain both the question and the answer terms
- Example
 - Mozart was born in 1756
 - <NAME> was born on <BIRTHDATE>

Watson (Ferrucci et al. 2010)

- Jeopardy winner (2011)
- Architecture
 - Uses "DeepQA": a technology that enables computer systems to directly and precisely answer natural language questions over an open and broad range of knowledge
 - 10 racks of IBM Power 750 servers running Linux
 - 16 terabytes of RAM
 - 2,880 processor cores
 - Capable of operating at 80 teraflops.
 - Mostly in Java but also some C++ and Prolog
 - Integrated using UIMA

Ferrucci et al. 2010. Building Watson: An Overview of the DeepQA Project. AI Magazine. Fall 2010. 59-79.

<http://www.pcmag.com/article2/0,2817,2380351,00.asp>

<http://blog.reddit.com/2011/02/ibm-watson-research-team-answers-your.html>

Watson (Ferrucci et al. 2000)

- Knowledge sources
 - 200 million pages of structured and unstructured content consuming four terabytes of disk storage
 - Including Wikipedia, WordNet, Yago
- Betting strategy
 - Buzz if at least 50% certain
- Performance
 - Watson answers 66 correct and 9 incorrect.
 - Watson's two day winning streak was \$77,147. Ken Jennings ended with \$24,000 and Brad Rutter with \$21,600.

Jeopardy winner (2011)
Deep QA
written in Java and
some C++

F

<http://www.quora.com/What-questions-were-asked-in-the-Jeopardy-episode-involving-Watson>

<http://theswimmingsubmarine.blogspot.com/2011/02/how-ibms-deep-question-answering.html>

<http://www.theatlantic.com/technology/archive/2011/02/is-it-time-to-welcome-our-new-computer-overlords/71388/>

Watson (Ferrucci et al. 2010)

- Question types
 - 2,500 of them
 - 200 of them are very common

*The Watson system
 has good strategies
 ... can wait for others*

- 2,500 of them
- 200 of them are very common

- **Jeopardy Question Archive**

- <http://j-archive.com/>

I can wait for others
to answer!
If not go on
get information from
others answer



Problems that
wait to be resolved

QA Challenges

- Word Sense Disambiguation
- Co-reference Resolution
- Semantic Role Labeling → to be discussed
- Temporal questions
- Categories on Jeopardy

e.g. $\frac{\text{Subj}}{\text{buy}} / \frac{\text{Obj}}$

answers the change over time

e.g. asking about a US → answer a Canadian city



NLP

The 5th slide is updated (about Watson Sys)