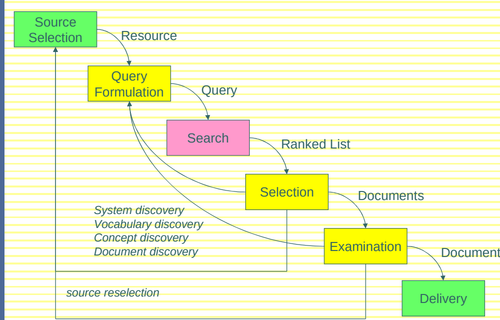
LBSC 796/INFM 718R: Week 12

Question Answering

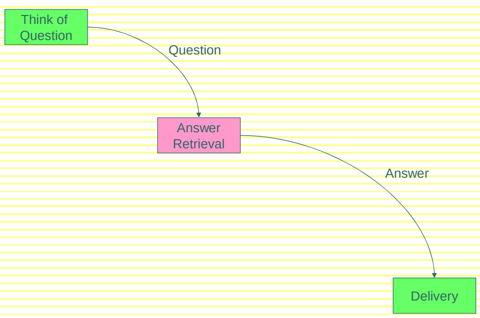
Jimmy Lin
College of Information Studies
University of Maryland

Monday, April 24, 2006

The Information Retrieval Cycle



Question Answering



Information Seeking Behavior

- Potentially difficult or time-consuming steps of the information seeking process:
  - Query formulation
  - Query refinement
  - Document examination and selection
- What if a system can directly satisfy information needs phrased in natural language?
  - Question asking is intuitive for humans
  - Compromised query = formalized query

This is a goal of question answering…

When is QA a good idea?

- Question asking is effective when…
  - The user knows exactly what he or she wants
  - The desired information is short, fact-based, and (generally) context-free

    Who discovered Oxygen?
    When did Hawaii become a state?
    Where is Ayer's Rock located?
    What team won the World Series in 1992?
- Question asking is less effective when…
  - The information need is vague or broad
  - The information request is exploratory in nature

Contrasting Information Needs

- *Ad hoc* retrieval: find me documents "like this"

  Identify positive accomplishments of the Hubble telescope since it was launched in 1991.

  Compile a list of mammals that are considered to be endangered, identify their habitat and, if possible, specify what threatens them.
- Question answering

"Factoid"	Who discovered Oxygen? When did Hawaii become a state? Where is Ayer's Rock located? What team won the World Series in 1992?
"List"	What countries export oil? Name U.S. cities that have a "Shubert" theater.
"Definition"	Who is Aaron Copland? What is a quasar?

From this…

To this…

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

Why is this better than Google?

- Keywords cannot capture semantic constraints between query terms:
  - **"home run records"**
    - Who holds the record for most home runs hit in a season?
    - Where is the home page for Run Records?
  - **"Boston sublet"**
    - I am looking for a room to sublet in Boston.
    - I have a room to sublet in Boston.
  - **"Russia invade"**
    - What countries have invaded Russia in the past?
    - What countries has Russia invaded in the past?
- Document retrieval systems cannot fuse together information from multiple documents

Who would benefit?

- Sample target users of a QA system
  - **Journalists checking facts:**
    - When did Mount Vesuvius last erupt?
    - Who was the president of Vichy France?
  - **Analysts seeking specific information:**
    - What is the maximum diving depth of a Kilo sub?
    - What's the range of China's newest ballistic missile?
  - **School children doing homework:**
    - What is the capital of Zimbabwe?
    - Where was John Wilkes Booth captured?
- Question answering fills an important niche in the broader information seeking environment

Roots of Question Answering

- Information Retrieval (IR)
- Information Extraction (IE)

Information Retrieval (IR)

- Can substitute "document" for "information"
- IR systems
  - Use statistical methods
  - Rely on frequency of words in query, document, collection
  - Retrieve complete documents
  - Return ranked lists of "hits" based on relevance
- Limitations
  - Answers questions indirectly
  - Does not attempt to understand the "meaning" of user's query or documents in the collection

Information Extraction (IE)

- IE systems
  - Identify documents of a specific type
  - Extract information according to pre-defined templates
  - Place the information into frame-like database records

Weather disaster:	Type	Damage
	Date	Deaths
	Location	...

- Templates = pre-defined questions
- Extracted information = answers
- Limitations
  - Templates are domain dependent and not easily portable
  - One size does not fit all!

Basic Strategy for Factoid QA

- Determine the semantic type of the expected answer

  "Who won the Nobel Peace Prize in 1991?" is looking for a PERSON

- Retrieve documents that have keywords from the question

  Retrieve documents that have the keywords "won", "Nobel Peace Prize", and "1991"

- Look for named-entities of the proper type near keywords

  Look for a PERSON near the keywords "won", "Nobel Peace Prize", and "1991"

An Example

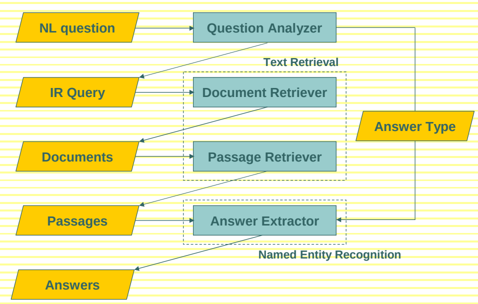
Who won the Nobel Peace Prize in 1991?

But many foreign investors remain sceptical, and western governments are withholding aid because of the Slorc's dismal human rights record and the continued detention of Ms Aung San Suu Kyi, the opposition leader who won the Nobel Peace Prize in 1991.

The military junta took power in 1988 as pro-democracy demonstrations were sweeping the country. It held elections in 1990, but has ignored their result. It has kept the 1991 Nobel peace prize winner, Aung San Suu Kyi - leader of the opposition party which won a landslide victory in the poll - under house arrest since July 1989.

The regime, which is also engaged in a battle with insurgents near its eastern border with Thailand, ignored a 1990 election victory by an opposition party and is detaining its leader, Ms Aung San Suu Kyi, who was awarded the 1991 Nobel Peace Prize. According to the British Red Cross, 5,000 or more refugees, mainly the elderly and women and children, are crossing into Bangladesh each day.

Generic QA Architecture

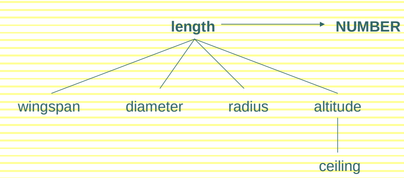


Question analysis

- Question word cues
  - Who → person, organization, location (e.g., city)
  - When → date
  - Where → location
  - What/Why/How → ??
- Head noun cues
  - What city, which country, what year...
  - Which astronaut, what blues band, ...
- Scalar adjective cues
  - How long, how fast, how far, how old, ...
- Cues from verbs
  - For example, *win* implies person or organization

Using WordNet

What is the service ceiling of an U-2?



Extracting Named Entities

Person: Mr. Hubert J. Smith, Adm. McInnes, Grace Chan

Title: Chairman, Vice President of Technology, Secretary of State

Country: USSR, France, Haiti, Haitian Republic

City: New York, Rome, Paris, Birmingham, Seneca Falls

Province: Kansas, Yorkshire, Uttar Pradesh

Business: GTE Corporation, FreeMarkets Inc., Acme

University: Bryn Mawr College, University of Iowa

Organization: Red Cross, Boys and Girls Club

More Named Entities

Currency: 400 yen, $100, DM 450,000

Linear: 10 feet, 100 miles, 15 centimeters

Area: a square foot, 15 acres

Volume: 6 cubic feet, 100 gallons

Weight: 10 pounds, half a ton, 100 kilos

Duration: 10 day, five minutes, 3 years, a millennium

Frequency: daily, biannually, 5 times, 3 times a day

Speed: 6 miles per hour, 15 feet per second, 5 kph

Age: 3 weeks old, 10-year-old, 50 years of age

How do we extract NEs?

- Heuristics and patterns
- Fixed-lists (gazetteers)
- Machine learning approaches

Indexing Named Entities

- Why would we want to index named entities?
- Index named entities as special tokens

  In reality, at the time of Edison's 1879 patent, the light bulb

  PERSON **DATE**

  had been in existence for some five decades ….

- And treat special tokens like query terms

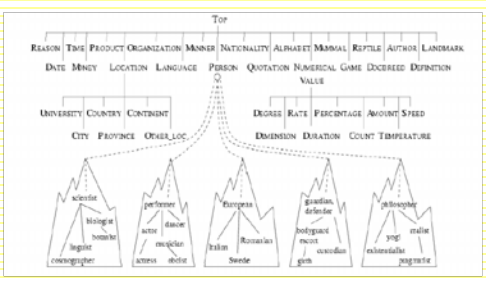
  Who patented the light bulb? → patent light bulb **PERSON**

  When was the light bulb patented? → patent light bulb **DATE**

- Works pretty well for question answering

John Prager, Eric Brown, and Anni Coden. (2000) Question-Answering by Predictive Annotation. Proceedings of SIGIR 2000.

Answer Type Hierarchy



When things go awry…

- Where do lobsters like to live?
  - on a Canadian airline
- Where do hyenas live?
  - in Saudi Arabia
  - in the back of pick-up trucks
- Where are zebras most likely found?
  - near dumps
  - in the dictionary
- Why can't ostriches fly?
  - Because of American economic sanctions
- What's the population of Maryland?
  - three

Question Answering… 2001

- TREC QA Track: TREC-8 (1999), TREC-9 (2000)
  - Formal evaluation of QA sponsored by NIST
  - Answer unseen questions using a newspaper corpus
- Question answering systems consisted of
  - A named-entity detector…
  - tacked on to a traditional document retrieval system
- General architecture:
  - Identify question type: person, location, date, etc.
  - Get candidate documents from off-the-shelf IR engines
  - Find named-entities of the correct type

Elaborate Ontologies

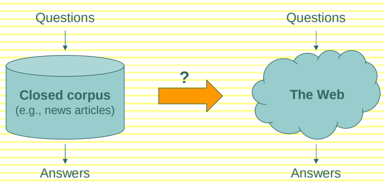
- Falcon: SMU's 2000 system in TREC:
  - 27 named entity categories
  - 15 top level nodes in answer type hierarchy
  - Complex many-to-many mapping between entity types and answer type hierarchy
- Webclopedia: ISI's 2000 system TREC:
  - Manually analyzed 17,384 questions
  - QA typology with 94 total nodes, 47 leaf nodes

As conceived, question answering was an incredibly labor-intensive endeavor…

Is there a way to shortcut the knowledge engineering effort?

Just Another Corpus?

- Is the Web just another corpus?
- Can we simply apply traditional IR+NE-based question answering techniques on the Web?



Not Just Another Corpus…

- The Web is qualitatively different from a closed corpus
- Many IR+NE-based question answering techniques are still effective
- But we need a different set of techniques to capitalize on the Web as a document collection

Using the Web for QA

- How big is the Web?
  - Tens of terabytes? No agreed upon methodology on how to measure it
  - Google indexes over 8 billion Web pages
- How do we access the Web?
  - Leverage existing search engines
- **Size gives rise to data redundancy**
  - Knowledge stated multiple times…
    - in multiple documents
    - in multiple formulations

Other Considerations

- Poor quality of many individual pages
  - Documents contain misspellings, incorrect grammar, wrong information, etc.
  - Some Web pages aren't even "documents" (tables, lists of items, etc.): not amenable to named-entity extraction or parsing
- Heterogeneity
  - Range in genre: encyclopedia articles vs. weblogs
  - Range in objectivity: CNN articles vs. cult websites
  - Range in document complexity: research journal papers vs. elementary school book reports

Leveraging Data Redundancy

- Take advantage of different reformulations
  - The expressiveness of natural language allows us to say the same thing in multiple ways
  - This poses a problem for question answering

Question asked in one way ← How do we bridge these two? → **Answer stated in another way**

"When did Colorado become a state?"

"Colorado was admitted to the Union on August 1, 1876."

"Colorado became a state on August 1, 1876"

  - With data redundancy, it is likely that answers will be stated in the same way the question was asked
- Cope with poor document quality
  - When many documents are analyzed, wrong answers become "noise"

Shortcutting NLP Challenges

Data Redundancy = Surrogate for sophisticated NLP
Obvious reformulations of questions can be easily found

Who killed Abraham Lincoln?

(1) John Wilkes Booth killed Abraham Lincoln.
(2) John Wilkes Booth altered history with a bullet. He will forever be known as the man who ended Abraham Lincoln's life.

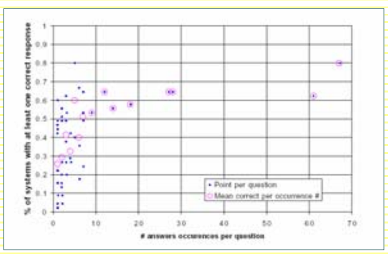
When did Wilt Chamberlain score 100 points?

(1) Wilt Chamberlain scored 100 points on March 2, 1962 against the New York Knicks.
(2) On December 8, 1961, Wilt Chamberlain scored 78 points in a triple overtime game. It was a new NBA record, but Warriors coach Frank McGuire didn't expect it to last long, saying, "He'll get 100 points someday." McGuire's prediction came true just a few months later in a game against the New York Knicks on March 2.

Effects of Data Redundancy

[Breck *et al.* 2001; Light *et al.* 2001]

Are questions with more answer occurrences "easier"?
Examined the effect of answer occurrences on question answering performance (on TREC-8 results)

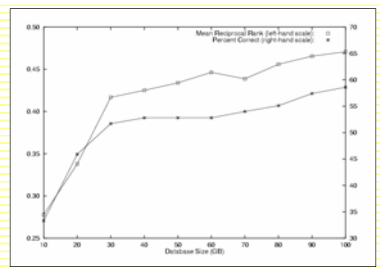


~27% of systems produced a correct answer for questions with 1 answer occurrence.
~50% of systems produced a correct answer for questions with 7 answer occurrences.

Effects of Data Redundancy

[Clarke *et al.* 2001a]

How does corpus size affect performance?
Selected 87 "people" questions from TREC-9; Tested effect of corpus size on passage retrieval algorithm (using 100GB TREC Web Corpus)



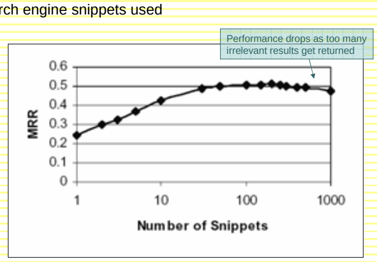
Conclusion: having more data improves performance

Effects of Data Redundancy

[Dumais *et al.* 2002]

How many search engine results should be used?
Plotted performance of a question answering system against the number of search engine snippets used

# Snippets	MRR
1	0.243
5	0.370
10	0.423
50	0.501
200	0.514

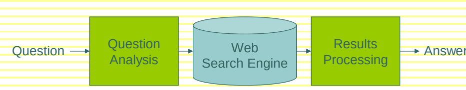


MRR as a function of number of snippets returned from the search engine. (TREC-9, q201-700)

Capitalizing on Search Engines

Data redundancy would be useless unless we could easily access all that data…

- Leverage existing information retrieval infrastructure [Brin and Page 1998]
  - The engineering task of indexing and retrieving terabyte-sized document collections has been solved
- Existing search engines are "good enough"
  - Build systems on top of commercial search engines, e.g., Google, FAST, AltaVista, Teoma, etc.



Redundancy-Based QA

- Reformulate questions into surface patterns likely to contain the answer
- Harvest "snippets" from Google
- Generate *n*-grams from snippets
- Compute candidate scores
- "Compact" duplicate candidates
- Apply appropriate type filters

Question Reformulation

- Anticipate common ways of answering questions
- Translate questions into surface patterns

  When did the Mesozoic period end?
  → The Mesozoic period ended *?x*

  - Apply simple pattern matching rules

    ***wh*-word did … verb → … verb+ed**
- Default to "bag of words" query if no reformulation can be found

N-Gram Mining

- Apply reformulated patterns to harvest snippets

Question: Who is Bill Gates married to?

→ Bill Gates is married to ?x

Default:
?x can be up to 5 tokens

… It is now the largest software company in the world. Today, Bill Gates is married to co-worker Melinda French. They live together in a house in the Redmond …

… I also found out that Bill Gates is married to Melinda French Gates and they have a daughter named Jennifer Katharine Gates and a son named Rory John Gates. I …

… of Microsoft, and they both developed Microsoft. * Presently Bill Gates is married to Melinda French Gates. They have two children: a daughter, Jennifer, and a …

Generate N-Grams from Google summary snippets (bypassing original Web pages)

co-worker, co-worker Melinda, co-worker Melinda French, Melinda, Melinda French, Melinda French they, French, French they, French they live…

Refining Answer Candidates

- Score candidates by frequency of occurrence and *idf*-values
- Eliminate candidates that are substrings of longer candidates
- Filter candidates by known question type

  What state … → answer must be a state
  What language … → answer must be a state
  How {fast, tall, far, …} → answer must be a number

What IS the answer?

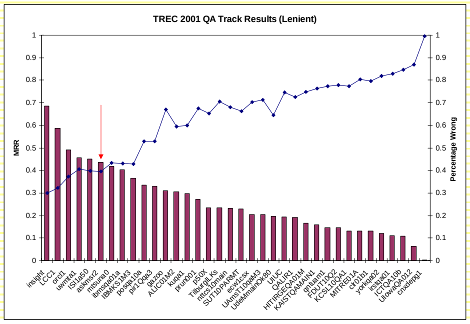
- Who is Bill Gates married to?
  - Melinda French
  - Microsoft
  - Mary Maxwell

Evaluation Metrics

- Mean Reciprocal Rank (MRR)
  - Reciprocal rank = inverse of rank at which first correct answer was found: {1, 0.5, 0.33, 0.25, 0.2, 0}
  - MRR = average over all questions
  - Judgments: correct, unsupported, incorrect

  Correct: answer string answers the question in a "responsive" fashion and is supported by the document
  Unsupported: answer string is correct but the document does not support the answer
  Incorrect: answer string does not answer the question
- Percentage wrong
  - Fraction of questions judged incorrect

How well did it work?



Pattern Learning [Ravichandran and Hovy 2002]

Automatically learn surface patterns for answering questions from the World Wide Web

BIRTHYEAR questions: When was <NAME> born?

<NAME> was born on <BIRTHYEAR>
<NAME> (<BIRTHYEAR>-
born in <BIRTHYEAR>, <NAME>
…

1. Start with a "seed", e.g. (Mozart, 1756)
2. Download Web documents using a search engine
3. Retain sentences that contain both question and answer terms
4. Extract the longest matching substring that spans <QUESTION> and <ANSWER>
5. Calculate precision of patterns
   - Precision for each pattern = # of patterns with correct answer / # of total patterns

Pattern Learning

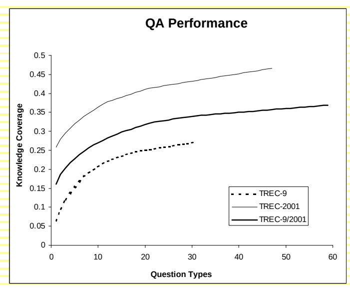
Example: DISCOVERER questions (Who discovered X?)

1.0	when <ANSWER> discovered <NAME>
1.0	<ANSWER>'s discovery of <NAME>
1.0	<ANSWER>, the discoverer of <NAME>
1.0	<ANSWER> discovers <NAME>
1.0	<ANSWER> discover <NAME>
1.0	<ANSWER> discovered <NAME>, the
1.0	discovery of <NAME> by <ANSWER>
0.95	<NAME> was discovered by <ANSWER>
0.91	of <ANSWER>'s <NAME>
0.9	<NAME> was discovered by <ANSWER> in

- Observations
  - Surface patterns perform better on the Web than on the TREC corpus
  - Surface patterns could benefit from notion of constituency, e.g., match not words but NPs, VPs, etc.

Zipf's Law in TREC

Cumulative distribution of question types in the TREC test collections



Ten question types alone account for ~20% of questions from TREC-9 and ~35% of questions from TREC-2001

The Big Picture

- Start with structured or semistructured resources on the Web
- Organize them to provide convenient methods for access
- Connected these resources to a natural language front end

Sample Resources

- Internet Movie Database
  - Content: cast, crew, and other movie-related information
  - Size: hundreds of thousands of movies; tens of thousands of actors/actresses
- CIA World Factbook
  - Content: geographic, political, demographic, and economic information
  - Size: approximately two hundred countries/territories in the world
- Biography.com
  - Content: short biographies of famous people
  - Size: tens of thousands of entries

"Zipf's Law of QA"

Observation: a few "question types" account for a large portion of all question instances

Similar questions can be parameterized and grouped into question classes, e.g.,

When was { Mozart / Einstein / Gandhi / … } born?

What is the { state bird / state capital / state flower / … } of { Alabama / Alaska / Arizona / … } ?

Where is { the Eiffel Tower / the Statue of Liberty / Taj Mahal / … } located?

Applying Zipf's Law of QA

- Observation: frequently occurring questions translate naturally into database queries

  What is the population of x? x ∈ {country}
  → get **population** of **x** from **World Factbook**

  When was x born? x ∈ {famous-person}
  → get **birthdate** of **x** from **Biography.com**

- How can we organize Web data so that such "database queries" can be easily executed?

Slurp or Wrap?

- Two general ways for accessing structured and semistructured Web resources
- **Wrap**
  - Also called "screen scraping"
  - Provide programmatic access to Web resources (in essence, an API)
  - Retrieve results dynamically by
    - Imitating a CGI script
    - Fetching a live HTML page
- **Slurp**
  - "Vacuum" out information from Web sources
  - Restructure information in a local database

Tradeoffs: Wrapping

- Advantages:
  - Information is always up-to-date (even when the content of the original source changes)
  - Dynamic information (e.g., stock quotes and weather reports) is easy to access
- Disadvantages:
  - Queries are limited in expressiveness

    Queries limited by the CGI facilities offered by the website
    Aggregate operations (e.g., max) are often impractical
  - Reliability issues: what if source goes down?
  - Wrapper maintenance: what if source changes layout/format?

Tradeoffs: Slurping

- Advantages:
  - Queries can be arbitrarily expressive

    Allows retrieval of records based on different keys
    Aggregate operations (e.g., max) are easy
  - Information is always available (high reliability)
- Disadvantages:
  - Stale data problem: what if the original source changes or is updated?
  - Dynamic data problem: what if the information changes frequently? (e.g., stock quotes and weather reports)
  - Resource limitations: what if there is simply too much data to store locally?

Putting it together

Connecting natural language questions to structured and semistructured data

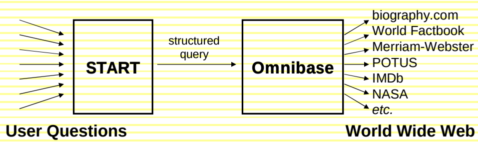
Natural Language System → structured query → **Semistructured Database** (slurp or wrap)

What is the population of x? x ∈ {country}
→ get **population** of **x** from **CIA Factbook**

When was x born? x ∈ {famous-person}
→ get **birthdate** of **x** from **Biography.com**

START and Omnibase

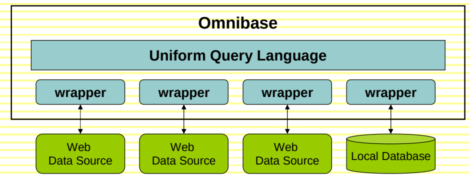
The first question answering system for the World Wide Web — online since 1993



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

Omnibase: Overview

- A "virtual" database that integrates structured and semistructured data sources
- An abstraction layer over heterogeneous sources



Omnibase: OPV Model

- The Object-Property-Value (OPV) data model
  - Relational data model adopted for natural language
  - Simple, yet pervasive

Sources contain **objects**
Objects have **properties**
Properties have **values**

Many natural language questions can be analyzed as requests for the value of a property of an object

- The "get" command:

(get **source object property**) → **value**

Omnibase: OPV Examples

- "What is the population of Taiwan?"
  - **Source:** CIA World Factbook
  - **Object:** Taiwan
  - **Property:** Population
  - **Value:** 22 million
- "When was Andrew Johnson president?"
  - **Source:** Internet Public Library
  - **Object:** Andrew Johnson
  - **Property:** Presidential term
  - **Value:** April 15, 1865 to March 3, 1869

Omnibase: OPV Coverage

10 Web sources mapped into the Object-Property-Value data model cover 27% of the TREC-9 and 47% of the TREC-2001 QA Track questions

Question	Object	Property	Value
Who wrote the music for the Titanic?	Titanic	composer	John Williams
Who invented dynamite?	dynamite	inventor	Alfred Nobel
What languages are spoken in Guernsey?	Guernsey	languages	English, French
Show me paintings by Monet.	Monet	works	

Omnibase: Wrappers

Omnibase Query
(get IPL "Abraham Lincoln" spouse)

Mary Todd (1818-1882), on November 4, 1842

Omnibase: Wrapper Operation

1. Generate URL
   - Map symbols onto URL

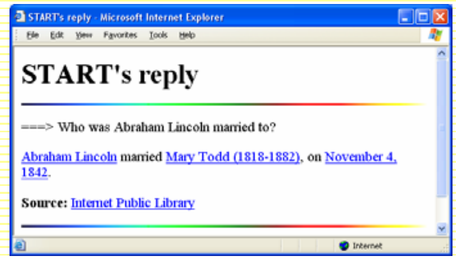
     Sometimes URLs can be computed directly from symbol
     Sometimes the mapping must be stored locally

     "Abraham Lincoln"
     "Abe Lincoln"
     "Lincoln"
     http://www.ipl.org/div/potus/alincoln.html
2. Fetch Web page
3. Extract relevant information
   - Search for textual landmarks that delimit desired information (usually with regular expressions)

     Married: (.*)

     Relevant information

Connecting the Pieces



Advanced Question Answering

- Move form "factoid" questions to more realistic question answering environments
- My current direction of research: clinical question answering

Beyond Counting Words…

- Information retrieval is based on counting words
- Different ways of "bookkeeping":
  - Vector space
  - Probabilistic
  - Language modeling
- Words alone aren't enough to capture meaning
- Retrieval of information:
  - Should be performed at the conceptual level
  - Should leverage knowledge about the information seeking process

What type of knowledge?

- Knowledge about user tasks
  - User intentions
  - *Why* does the user need the information?
- Knowledge about the problem structure
  - The method for approaching a problem
  - *What* information is needed?
- Knowledge about the domain
  - Background knowledge needed to understand the question and answers
  - Relationships such as hypernymy and synonymy

Jimmy Lin and Dina Demner-Fushman. The Role of Knowledge in Conceptual Retrieval: A Study in the Domain of Clinical Medicine. Proceedings of SIGIR 2006.

Knowledge: User Tasks

- Clinical tasks
  - Therapy: Selecting effective treatments, taking into account other factors such as risk and cost
  - Diagnosis: Selecting and interpreting diagnostic tests, while considering their precision, safety, cost, etc
  - Prognosis: Estimating the patient's likely course over time and anticipating likely complications
  - Etiology: Identifying the causes for diseases
- Considerations for strength of evidence
  - Strength of Recommendations Taxonomy (SORT): three evidence grades

Knowledge about user tasks
Knowledge about problem structure
Knowledge about the domain

Knowledge: Problem Structure

- EBM identifies four components of a question:
  - **P**roblem/**P**opulation
    What is the primary problem or disease? What are the characteristics of the patient (age, gender, etc.)?
  - **I**ntervention
    What is the main intervention (e.g., a medication, a diagnostic test, or therapeutic procedure)?
  - **C**omparison
    What is the main intervention compared to (e.g., another drug, a placebo)?
  - **O**utcome
    What is the effect of the intervention? Were the patient's symptoms relieved or eliminated?
- **= PICO frame**

Knowledge about user tasks
Knowledge about problem structure
Knowledge about the domain

Knowledge: Problem Structure

"In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?"

- **Population/Problem:** children/acute febrile illness
- **Intervention:** acetaminophen
- **Comparison:** ibuprofen
- **Outcome:** reducing fever

Knowledge about user tasks
Knowledge about problem structure
Knowledge about the domain

Knowledge: Domain

- The Unified Medical Language System (UMLS)
  - 2004 version: > 1 million biomedical concepts, > 5 million concept names
  - Semantic groups provide higher-level generalizations
- Software for leveraging this resource:
  - MetaMap for identifying concepts
  - SemRep for identifying relations

Knowledge about user tasks
Knowledge about problem structure
Knowledge about the domain

Matching PICO frames

Question:
In children with an acute febrile illness, what is the efficacy of single-medication therapy with acetaminophen or ibuprofen in reducing fever?

Task therapy
P children/acute febrile illness
I acetaminophen
C ibuprofen
O reducing fever

Answer:
Ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses.

P children/acute febrile illness
I acetaminophen
C ibuprofen
O reducing fever

MEDLINE

Knowledge Extraction Example

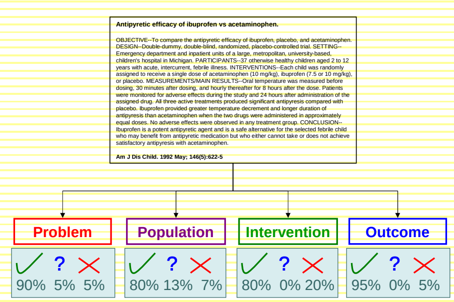
Antipyretic efficacy of ibuprofen vs acetaminophen.

OBJECTIVE--To compare the antipyretic efficacy of ibuprofen, placebo, and acetaminophen. DESIGN--Double-dummy, double-blind, randomized, placebo-controlled trial. SETTING--Emergency department and inpatient units of a large, metropolitan, university-based, children's hospital in Michigan. PARTICIPANTS--37 otherwise healthy children aged 2 to 12 years with acute, intercurrent, febrile illness. INTERVENTIONS--Each child was randomly assigned to receive a single dose of acetaminophen (10 mg/kg), ibuprofen (7.5 or 10 mg/kg), or placebo. MEASUREMENTS/MAIN RESULTS--Oral temperature was measured before dosing, 30 minutes after dosing, and hourly thereafter for 8 hours after the dose. Patients were monitored for adverse effects during the study and 24 hours after administration of the assigned drug. All three active treatments produced significant antipyresis compared with placebo. Ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses. No adverse effects were observed in any treatment group. CONCLUSION--Ibuprofen is a potent antipyretic agent and is a safe alternative for the selected febrile child who may benefit from antipyretic medication but who either cannot take or does not achieve satisfactory antipyresis with acetaminophen.

Am J Dis Child. 1992 May; 146(5):622-5

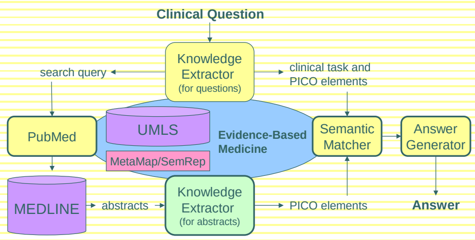
Population Problem Interventions Outcome

Knowledge Extractors



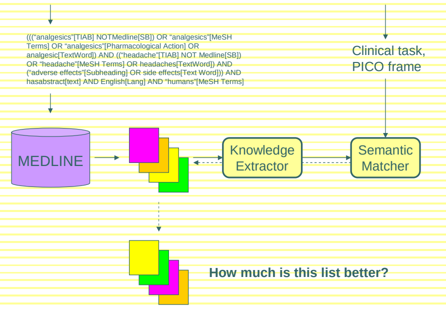
Dina Demner-Fushman and Jimmy Lin. Knowledge Extraction for Clinical Question Answering: Preliminary Results. Proceedings of the AAAI-05 Workshop on Question Answering in Restricted Domains, 2005.

Overall Architecture



Citation Reranking Experiment

Question: What is the best treatment for analgesic rebound headaches?



Experimental Results

- We created a test collection comprised of 50 realistic clinical questions
- Performance on held-out blind test set:

	Therapy	Diagnosis	Prognosis	Etiology	All
Precision at 10 (P@10)					
PubMed	.350	.150	.200	.320	.281
EBM	.792 (+126%)	.567 (+278%)	.433 (+117%)	.640 (+100%)	**.669 (+97%)**
Mean Average Precision (MAP)					
PubMed	.421	.279	.235	.364	.356
EBM	.775 (+84%)	.657 (+135%)	.715 (+204%)	.681 (+87%)	**.723 (+103%)**

Jimmy Lin and Dina Demner-Fushman. The Role of Knowledge in Conceptual Retrieval: A Study in the Domain of Clinical Medicine. Proceedings of SIGIR 2006.

Multiple Approaches to QA

- Employ answer type ontologies (IR+IE)
- Leverage Web redundancy
- Leverage semi-structured data sources
- Semantically model restricted domains for "conceptual retrieval"