

CSE 842 Natural Language Processing

Lecture 19: Information Retrieval and Question Answering

3/26/2009

CSE842, Spring 2009, MSU

1

Do people ask questions to a search engine?

- Examples from AltaVista query log
 - Who invented surf music?
 - How to make stink bomb?
 - How tall is the sears tower?
 - Which english translation of the bible is used in official catholic liturgies?
- Examples from Excite query log
 - How can I find someone in texas
 - What are the seven wonders of the world?
 - What vacuum cleaner does Consumers Guide recommend?
- Percentage of query logs that are questions: 10-15%

3/26/2009

CSE842, Spring 2009, MSU

2

Open Domain Question Answering

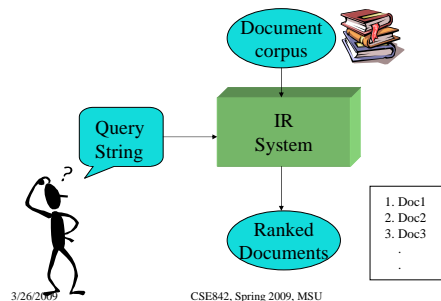
- Find answers to open domain natural language questions by searching a large collection of documents
- Documents: texts, web pages, digital libraries, maps, databases, images, etc.
- Today's topics:
 - Information retrieval – vector space model
 - Factoid question answering

3/26/2009

CSE842, Spring 2009, MSU

3

Information Retrieval

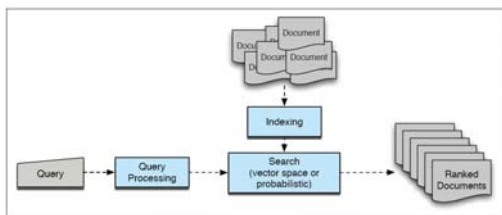


3/26/2009

CSE842, Spring 2009, MSU

4

IR Architecture



3/26/2009

CSE842, Spring 2009, MSU

5

Typical IR Task

- Given:
 - A corpus of textual natural-language documents.
 - A user query in the form of a textual string.
- Find:
 - A ranked set of documents that are relevant to the query.
- Concerned
 - firstly with retrieving *relevant* documents to a query.
 - secondly with retrieving from *large* sets of documents *efficiently*.
- CSE484 offers a comprehensive coverage on IR.

3/26/2009

CSE842, Spring 2009, MSU

6

IR System Components

- Text Operations forms index words (tokens).
 - Stopword removal
 - Stemming
- Indexing constructs an *inverted index* of word to document pointers.
- Searching retrieves documents that contain a given query token from the inverted index.
- Ranking scores all retrieved documents according to a relevance metric.

3/26/2009

CSE842, Spring 2009, MSU

7

IR System Components (continued)

- User Interface manages interaction with the user:
 - Query input and document output.
 - Relevance feedback.
 - Visualization of results.
- Query Operations transform the query to improve retrieval:
 - Query expansion using a thesaurus.
 - Query transformation using relevance feedback.

3/26/2009

CSE842, Spring 2009, MSU

8

Retrieval Models

- A retrieval model specifies the details of:
 - Document representation
 - Query representation
 - Retrieval function
- Determines a notion of relevance.
- Notion of relevance can be binary or continuous (i.e. *ranked retrieval*).

3/26/2009

CSE842, Spring 2009, MSU

9

Classes of Retrieval Models

- Boolean models (set theoretic)
 - Extended Boolean
- Vector space models (statistical/algebraic)
 - Generalized VS
 - Latent Semantic Indexing
- Probabilistic models

3/26/2009

CSE842, Spring 2009, MSU

10

Common Preprocessing Steps

- Strip unwanted characters/markup (e.g. HTML tags, punctuation, numbers, etc.).
- Break into tokens (keywords) on whitespace.
- Stem tokens to “root” words
 - computational → compute
- Remove common stopwords (e.g. a, the, it, etc.).
- Detect common phrases (possibly using a domain specific dictionary).
- Build inverted index (keyword → list of docs containing it).

3/26/2009

CSE842, Spring 2009, MSU

11

Boolean Model

- A document is represented as a **set** of keywords.
- Queries are Boolean expressions of keywords, connected by AND, OR, and NOT, including the use of brackets to indicate scope.
 - [[Rio & Brazil] | [Hilo & Hawaii]] & hotel & !Hilton]
- Output: Document is relevant or not. No partial matches or ranking.

3/26/2009

CSE842, Spring 2009, MSU

12

Boolean Retrieval Model

- Popular retrieval model:
 - Easy to understand for simple queries.
 - Clean formalism.
 - Reasonably efficient implementations possible for normal queries.
- Problems:
 - Very rigid: AND means all; OR means any.
 - Difficult to control the number of documents retrieved.
 - All matched documents will be returned.
 - Difficult to rank output.
 - All matched documents logically satisfy the query

3/26/2009

CSE842, Spring 2009, MSU

13

Statistical Models

- A document is typically represented by a *bag of words* (unordered words with frequencies).
- Bag = set that allows multiple occurrences of the same element.
- User specifies a set of desired terms with optional weights:
 - Weighted query terms:
Q = < database 0.5; text 0.8; information 0.2 >
 - Unweighted query terms:
Q = < database; text; information >
 - No Boolean conditions specified in the query.

3/26/2009

CSE842, Spring 2009, MSU

14

Statistical Retrieval

- Retrieval based on *similarity* between query and documents.
- Output documents are ranked according to similarity to query.
- Similarity based on occurrence *frequencies* of keywords in query and document.

3/26/2009

CSE842, Spring 2009, MSU

15

The Vector-Space Model

- Assume t distinct terms remain after preprocessing; call them index terms or the vocabulary.
- These “orthogonal” terms form a vector space.
Dimension = $t = |\text{vocabulary}|$
- Each term, i , in a document or query, j , is given a real valued weight, w_{ij} .
- Both documents and queries are expressed as t dimensional vectors:

$$d_j = (w_{1j}, w_{2j}, \dots, w_{tj})$$

3/26/2009

CSE842, Spring 2009, MSU

16

Document Collection

- A collection of n documents can be represented in the vector space model by a term document matrix.
- An entry in the matrix corresponds to the “weight” of a term in the document; zero means the term has no significance in the document or it simply doesn’t exist in the document

$$\begin{pmatrix} & d_1 & d_2 & \dots & d_n \\ t_1 & w_{11} & w_{12} & \dots & w_{1n} \\ t_2 & w_{21} & w_{22} & \dots & w_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ t_m & w_{m1} & w_{m2} & \dots & w_{mn} \end{pmatrix}$$

3/26/2009

CSE842, Spring 2009, MSU

17

Term Weights: Term Frequency

- More frequent terms in a document are more important, i.e. more indicative of the topic.

$$tf_{ij} = \text{frequency of term } i \text{ in document } j$$

- Term frequency is usually dampened by a function like:

$$f(tf) = \sqrt{tf}$$

$$f(tf) = 1 + \log(tf)$$

3/26/2009

CSE842, Spring 2009, MSU

18

Term Weights: Inverse Document Frequency

- Terms that appear in many *different* documents are *less* indicative of overall topic.
 df_i = document frequency of term i
= number of documents containing term i
 idf_i = inverse document frequency of term i ,
= $\log_2 (N/df_i)$
(N : total number of documents)
- An indication of a term's *discrimination* power.
- Log used to dampen the effect.

3/26/2009

CSE842, Spring 2009, MSU

19

TF-IDF Weighting

- A typical combined term importance indicator is *tf-idf weighting*:
$$W_{ij} = (1 + \log_2(tf_{ij})) \log_2(N/df_i)$$
- A term occurring frequently in the document but rarely in the rest of the collection is given high weight.
- Many other ways of determining term weights have been proposed.
- Experimentally, *tf-idf* has been found to work well.

3/26/2009

CSE842, Spring 2009, MSU

20

Computing TF-IDF -- An Example

Given a document containing terms with given frequencies:

A(3), B(2), C(1)

Assume collection contains 10,000 documents and document frequencies of these terms are:

A(50), B(1300), C(250)

Then:

A: $tf = 1 + \log 3 = 2.6$; $idf = \log(10000/50) = 5.3$; $tf-idf = 13.8$

B: $tf = 1 + \log 2 = 2$; $idf = \log(10000/1300) = 2.0$; $tf-idf = 4.0$

C: $tf = 1$; $idf = \log(10000/250) = 3.7$; $tf-idf = 3.7$

3/26/2009

CSE842, Spring 2009, MSU

21

Query Vector

- Query vector is typically treated as a document and also tf-idf weighted.
- Alternative is for the user to supply weights for the given query terms.

3/26/2009

CSE842, Spring 2009, MSU

22

Similarity Measure

- A **similarity measure** is a function that computes the *degree of similarity* between two vectors.
- Using a similarity measure between the query and each document:
 - It is possible to rank the retrieved documents in the order of presumed relevance.
 - It is possible to enforce a certain threshold so that the size of the retrieved set can be controlled.

3/26/2009

CSE842, Spring 2009, MSU

23

Similarity Measure - Inner Product

- Similarity between vectors for the document d_j and query q can be computed as the vector inner product:

$$\text{sim}(d_j, q) = d_j \cdot q = \sum_{i=1}^t w_{ij} \cdot w_{iq}$$

where w_{ij} is the weight of term i in document j and w_{iq} is the weight of term i in the query

- For binary vectors, the inner product is the number of matched query terms in the document (size of intersection).
 - For weighted term vectors, it is the sum of the products of the weights of the matched terms.
- Inner Product needs to be normalized

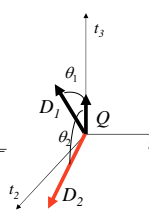
3/26/2009

CSE842, Spring 2009, MSU

24

Cosine Similarity Measure

- Cosine similarity measures the cosine of the angle between two vectors.
- Inner product normalized by the vector lengths.

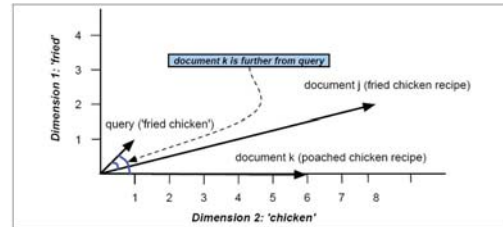
$$\text{CosSim}(d_j, q) = \frac{\vec{d}_j \cdot \vec{q}}{|\vec{d}_j| \cdot |\vec{q}|} = \frac{\sum_{i=1}^n (w_{ij} \cdot w_{iq})}{\sqrt{\sum_{i=1}^n w_{ij}^2} \cdot \sqrt{\sum_{i=1}^n w_{iq}^2}}$$


3/26/2009

CSE842, Spring 2009, MSU

25

How to Measure Relevancy?



3/26/2009

CSE842, Spring 2009, MSU

26

Summary on Vector Space Models

- Simple, mathematically based approach.
- Considers both local (*tf*) and global (*idf*) word occurrence frequencies.
- Provides partial matching and ranked results.
- Tends to work quite well in practice despite obvious weaknesses.
- Allows efficient implementation for large document collections.

3/26/2009

CSE842, Spring 2009, MSU

27

Problems with Vector Space Model

- Missing semantic information (e.g. word sense).
- Missing syntactic information (e.g. phrase structure, word order, proximity information).
- Assumption of term independence (e.g. ignores synonymy).
- Lacks the control of a Boolean model (e.g., *requiring* a term to appear in a document).
 - Given a two-term query “A B”, may prefer a document containing A frequently but not B, over a document that contains both A and B, but both less frequently.

3/26/2009

CSE842, Spring 2009, MSU

28

Evaluation

- There are many retrieval models/ algorithms/ systems, which one is the best?
- What is the best component for:
 - Ranking function (dot product, cosine, ...)
 - Term selection (stopword removal, stemming...)
 - Term weighting (TF, TF-IDF, ...)
- How far down the ranked list will a user need to look to find some/all relevant documents?

3/26/2009

CSE842, Spring 2009, MSU

29

Difficulties in Evaluating IR Systems

- Effectiveness is related to the *relevancy* of retrieved items.
- Relevancy is not typically binary but continuous.
- Even if relevancy is binary, it can be a difficult judgment to make.
- Relevancy, from a human point of view, is:
 - Subjective: Depends upon a specific user's judgment.
 - Situational: Relates to user's current needs.
 - Dynamic: Changes over time
- Relevancy may include:
 - Being on the proper subject.
 - Being timely (recent information).
 - Being authoritative (from a trusted source).

3/26/2009

CSE842, Spring 2009, MSU

30

Human Labeled Corpora (Gold Standard)

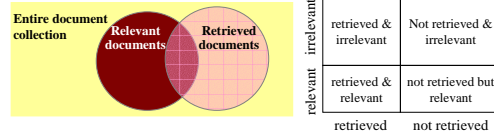
- Start with a corpus of documents.
- Collect a set of queries for this corpus.
- Have one or more human experts exhaustively label the relevant documents for each query.
- Typically assumes binary relevance judgments.
- Requires considerable human effort for large document/query corpora.

3/26/2009

CSE842, Spring 2009, MSU

31

Precision and Recall



$$\text{recall} = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of relevant documents}}$$

$$\text{precision} = \frac{\text{Number of relevant documents retrieved}}{\text{Total number of documents retrieved}}$$

3/26/2009

CSE842, Spring 2009, MSU

32

Precision and Recall

- Precision
 - The ability to retrieve top ranked documents that are mostly relevant.
- Recall
 - The ability of the search to find *all* of the relevant items in the corpus.

3/26/2009

CSE842, Spring 2009, MSU

33

Determining Recall is Difficult

- Total number of relevant items is sometimes not available:
 - Sample across the database and perform relevance judgment on these items.
 - Apply different retrieval algorithms to the same database for the same query. The aggregate of relevant items is taken as the total relevant set.

3/26/2009

CSE842, Spring 2009, MSU

34

Computing Recall/Precision Points

- For a given query, produce the ranked list of retrievals.
- Adjusting a threshold on this ranked list produces different sets of retrieved documents, and therefore different recall/precision measures.
- Mark each document in the ranked list that is relevant according to the gold standard.
- Compute a recall/precision pair for each position in the ranked list that contains a relevant document.

3/26/2009

CSE842, Spring 2009, MSU

35

Computing Recall/Precision Points:

An Example

n	doc #	relevant
1	588	x
2	589	x
3	576	
4	590	x
5	986	
6	592	x
7	984	
8	988	
9	578	
10	985	
11	103	
12	591	
13	772	x
14	990	

Let total # of relevant docs = 6
Check each new recall point:

R=1/6=0.167; P=1/1=1

R=2/6=0.333; P=2/2=1

R=3/6=0.5; P=3/4=0.75

R=4/6=0.667; P=4/6=0.667

R=5/6=0.833; P=5/13=0.38

Missing one relevant document.
Never reach 100% recall

3/26/2009

CSE842, Spring 2009, MSU

36

Interpolating a Recall/Precision Curve

- Interpolate a precision value for each *standard recall level*:
 - $r_j \in \{0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$
 - $r_0 = 0.0, r_1 = 0.1, \dots, r_{10} = 1.0$
- The interpolated precision at the j -th standard recall level is the maximum known precision at any recall level greater than j .

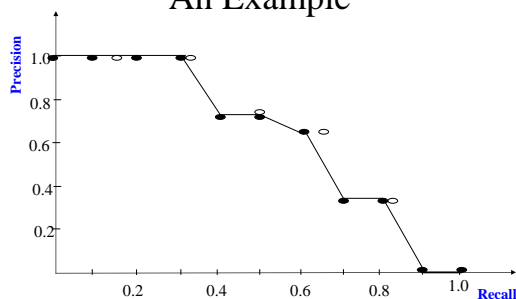
$$P(r_j) = \max_{r \geq r_j} P(r)$$

3/26/2009

CSE842, Spring 2009, MSU

37

Interpolating a Recall/Precision: An Example



3/26/2009

CSE842, Spring 2009, MSU

38

Average Recall/Precision Curve

- Typically average performance over a large *set* of queries.
- Compute average precision at each standard recall level across all queries.
- Plot average precision/recall curves to evaluate overall system performance on a document/query corpus.

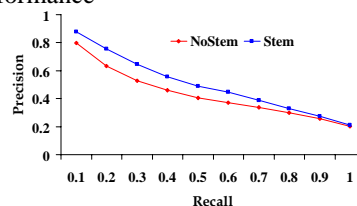
3/26/2009

CSE842, Spring 2009, MSU

39

Compare Two or More Systems

- The curve closest to the upper right-hand corner of the graph indicates the best performance



3/26/2009

CSE842, Spring 2009, MSU

40

R- Precision

- Precision at the R -th position in the ranking of results for a query that has R relevant documents.

n	doc #	relevant
1	588	x
2	589	x
3	576	
4	590	x
5	986	
6	592	x
7	984	
8	988	
9	578	
10	985	
11	103	
12	591	
13	772	x
14	990	

$R = \# \text{ of relevant docs} = 6$

$R\text{-Precision} = 4/6 = 0.67$

3/26/2009

CSE842, Spring 2009, MSU

41

The TREC Benchmark

- TREC: Text REtrieval Conference (<http://trec.nist.gov/>)
 - Originated from the TIPSTER program sponsored by Defense Advanced Research Projects Agency (DARPA).
- Became an annual conference in 1992, co-sponsored by the National Institute of Standards and Technology (NIST) and DARPA.
- Participants are given parts of a standard set of documents and TOPICS (from which queries have to be derived) in different stages for training and testing.
- Participants submit the P/R values for the final document and query corpus and present their results at the conference.

3/26/2009

CSE842, Spring 2009, MSU

42

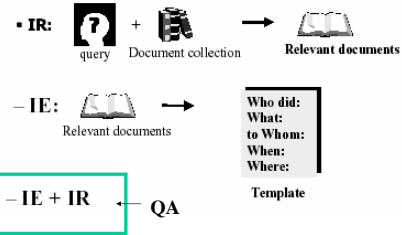
Question Answering

- It is not a new research area
- QA systems can be found in many areas of NLP research
 - Natural language queries to database systems
 - Spoken dialog systems (e.g., ATIS)
- Open domain QA is new
 - MURAX (Kupiec 1993): Encyclopedia answers
 - Hirschman: reading comprehension tests
 - TREC QA competition: 1999-2007

3/26/2009

CSE842, Spring 2009, MSU

43



3/26/2009

CSE842, Spring 2009, MSU

44

Factoid Question Answering

Question	Answer
Where is the Louvre Museum located?	in Paris, France
What's the abbreviation for limited partnership?	L.P.
What are the names of Odin's ravens?	Huginn and Muninn
What currency is used in China?	the yuan
What kind of nuts are used in marzipan?	almonds
What instrument does Max Roach play?	drums
What's the official language of Algeria?	Arabic
What is the telephone number for the University of Colorado, Boulder?	(303)492-1411
How many pounds are there in a stone?	14

3/26/2009

CSE842, Spring 2009, MSU

45

Example

- Q: what is the fastest car in the world?
- Correct answer:
 - ... the Jaguar XJ220 is the dearest (pounds 415,000), fastest (217mph/350 kmh) and most sought-after car in the world
- Wrong answer:
 - ...will stretch Volkswagen's lead in the world's fastest-growing vehicle market. Demand is expected to soar in the next few years as more Chinese are able to afford their own cars.

3/26/2009

CSE842, Spring 2009, MSU

46

Dimensions of the Problem

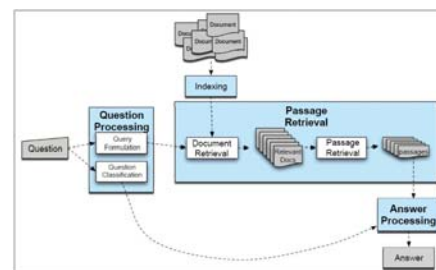
- To answer a question, a QA system must
 - Analyze the question in context of some ongoing interaction
 - Find one or more answers by consulting online resources
 - Present the answer to the user in appropriate form
 - Associate answer with justification or supporting material

3/26/2009

CSE842, Spring 2009, MSU

47

Anatomy of a QA System



3/26/2009

CSE842, Spring 2009, MSU

48

Question Analysis

- Simplest question analysis attempts to identify the expected semantic type of the answer
 - Some systems have just a few answer types
 - Some have about 100
- Many TREC systems use simple (but robust) regular expression approaches
 - Who -> PERSON
 - Where -> LOCATION
 - When -> DATE

3/26/2009

CSE842, Spring 2009, MSU

49

TREC QA Track: Overview

- TREC (Text REtrieval Conference) supports a QA track. (1999-2007)
- Each year the QA task is a bit more difficult
 - Factoid QA (focused in early years)
 - Context QA (a sequence of questions, start around 2002)
 - Complex QA (later years)
 - Interactive QA (later years)
- For each TREC QA track, NIST provides
 - A document collection, and
 - A test set of questions.
- Participants build a QA system and return a ranked list of five *[document-id, answer-string]* pairs per question.

3/26/2009

CSE842, Spring 2009, MSU

50

TREC Text Collection

- The mostly used collection (AQUAINT)
 - Size: 1,033,461 documents, 3GB of text
 - Sources:
 - AP newswire, 1998-2000
 - New York Times newswire, 1998-2000
 - Xinhua News Agency newswire, 1996-2000
- Too much of text to process entirely using advanced NLP techniques: IR + more advanced processing
- Many supplement this text with use of the web, and other knowledge bases

3/26/2009

CSE842, Spring 2009, MSU

51

Answer Assessment

- Human assessors read each answer string and decide whether the string actually contains an answer to the question in the context provided by the document
- An individual question receives a score equal to the reciprocal rank:
 - 1 for answer at rank 1,
 - 1/5 for answer at rank 5,
 - 0 if none of the five responses contained the correct answer
- The score for a submission for factoid QA is the mean of the individual questions' reciprocal ranks (MRR)

3/26/2009

CSE842, Spring 2009, MSU

52

Scoring Metric

- MRR Advantages:
 - It is closely related to the average precision measure
 - It is bounded between 0 and 1
 - Is penalized for not retrieving any correct answers
- Disadvantages:
 - Only take six values (0, 0.2, 0.25, 0.33, 0.5, 1)
 - No credit is given for different correct answers
 - No credit is given for realizing that no answer exists

3/26/2009

CSE842, Spring 2009, MSU

53

SpartansQA



SpartansQA

Question Answering

Enter your question here:

Answer:

Question Classification LingPipe NER Parse Tree Semantics Query Paragraphs Answer Presentation

```

*html version="1" id="encoding=UTF-8"
*QUESTION_ID="4"
*EXPECTED_ANSWER=
*TYPE=NUMBER+ITYPE=
*SUBTYPE=GROUP+SUBTYPE=
*EXPECTED_ANSWER=
*QUESTION=
    
```

Finished processing question. Waiting for next question.

SpartansQA

Question Answering

Enter your question here:

Answer:

Question Classification LingPipe NER Parse Tree Semantics Query Paragraphs Answer Presentation

```

*html version="1" id="encoding=UTF-8"
*QUESTION_ID="4"
*EXPECTED_ANSWER=
*TYPE=NUMBER+ITYPE=
*SUBTYPE=GROUP+SUBTYPE=
*EXPECTED_ANSWER=
*QUESTION=
    
```

Finished processing question. Waiting for next question.

Classify a question according to the type of answer as suggested by its question word

- Who: a person or an organization
- When: a time designation

SpartansQA

Question Answering

Enter your question here:

Answer:

Question Classification LingPipe NER Parse Tree Semantics Query Paragraphs Answer Presentation

Rank 1:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 2:
 The resignation "is a failure for Lionel Jospin," said Michèle Allot-Marie the head of President Jacques Chirac's **Conservative Party for the Republic** party. "In his own camp we see his closest people denounce the ambiguities and risks of the accord just as we have done."

Rank 3:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 4:
 PARIS - "When the French want to change policy they will turn to you," President Francois Mitterand told Lionel Jospin his fellow **Socialist** and long-time party aide in 1995 after Jospin unsuccessfully sought the presidency.

Rank 5:
 PARIS - "When the French want to change policy they will turn to you," President Francois Mitterand told Lionel Jospin his fellow **Socialist** and long-time party aide in 1995 after Jospin unsuccessfully sought the presidency.

Rank 6:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 7:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 8:
 PARIS - "When the French want to change policy they will turn to you," President Francois Mitterand told Lionel Jospin his fellow **Socialist** and long-time party aide in 1995 after Jospin unsuccessfully sought the presidency.

Rank 9:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 10:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Finished processing question. Waiting for next question.

SpartansQA

Question Answering

Enter your question here:

Answer:

Question Classification LingPipe NER Parse Tree Semantics Query Paragraphs Answer Presentation

Rank 1:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 2:
 The resignation "is a failure for Lionel Jospin," said Michèle Allot-Marie the head of President Jacques Chirac's **Conservative Party for the Republic** party. "In his own camp we see his closest people denounce the ambiguities and risks of the accord just as we have done."

Rank 3:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 4:
 PARIS - "When the French want to change policy they will turn to you," President Francois Mitterand told Lionel Jospin his fellow **Socialist** and long-time party aide in 1995 after Jospin unsuccessfully sought the presidency.

Rank 5:
 PARIS - "When the French want to change policy they will turn to you," President Francois Mitterand told Lionel Jospin his fellow **Socialist** and long-time party aide in 1995 after Jospin unsuccessfully sought the presidency.

Rank 6:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 7:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 8:
 PARIS - "When the French want to change policy they will turn to you," President Francois Mitterand told Lionel Jospin his fellow **Socialist** and long-time party aide in 1995 after Jospin unsuccessfully sought the presidency.

Rank 9:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Rank 10:
 -> Prime Minister Alain Juppé today formally handed over the prime minister's seal to **Socialist Party** first secretary Lionel Jospin.

Finished processing question. Waiting for next question.

Retrieve a small portion of the collection using standard IR technologies.

Shallow parse to detect entities of the same type as the answer

SpartansQA

Question Answering

Enter your question here:

Answer:

Question Classification LingPipe NER Parse Tree Semantics Query Paragraphs Answer Presentation

Waiting for a question.

SpartansQA

Question Answering

Enter your question here:

Answer:

Question Classification LingPipe NER Parse Tree Semantics Query Paragraphs Answer Presentation

```

*html version="1" id="encoding=UTF-8"
*QUESTION_ID="1"
*EXPECTED_ANSWER=
*TYPE=NUMBER+ITYPE=
*SUBTYPE=CATE+SUBTYPE=
*EXPECTED_ANSWER=
*QUESTION=
    
```

Finished processing question. Waiting for next question.

SpartansQA



Desiderata of a QA System

- **Timeliness**
 - The answer must be provided in real-time
 - New data sources must be incorporated as soon as they become available
- **Accuracy**
 - Incorrect answers are worse than no answers
- **Usability**
 - The system must be able to mine answers regardless of the data source format
 - The answer must be delivered in any format desired by the user

3/26/2009

CSE842, Spring 2009, MSU

62

Desiderata of a QA System

- **Completeness**
 - Complete answers are desirable
 - Answer fusion in a coherent information is required
 - Must incorporate capabilities of reasoning and using high performing knowledge bases
- **Relevance**
 - The answer must be relevant within a specific context
 - Interactive QA in which a sequence of questions helps clarify an information need may be necessary

3/26/2009

CSE842, Spring 2009, MSU

63

The Issues

- **Question Classes**
 - Need for question taxonomies
 - Lehnert : 13: Causal Antecedent (e.g., why did John go to New York?), etc.
 - Graesser: 18: Comparison (e.g., how is X different from Y)
 - Determine the question focus
 - E.g., why did John roller-skate to McDonald's last night?
 - People are largely unconscious of the cognitive processes involved in answering a question, and are consequently insensitive to the complexities of these processes

3/26/2009

CSE842, Spring 2009, MSU

64

The Issues

- **Question Processing**
 - Understanding, ambiguities, implicatures and reformulations
 - *What French cities are larger than Bordeaux?*
 - *Do you have a light?*
 - *How rich is Bill Gates?*
 - A semantic model of question understanding and processing is needed
 - A high-performance QA system should be able to extract the same answer for any equivalent question posed

3/26/2009

CSE842, Spring 2009, MSU

65

The Issues

- **Context**
 - Questions are asked within a context and answers are provided within that specific context
- **Data Sources**
 - Before a question can be answered, it must be known what knowledge sources are available
- **Answer extraction**
 - Extraction of simple and distributed answers
 - Answer justification and evaluation of answer correctness

3/26/2009

CSE842, Spring 2009, MSU

66

The Issues

- Answer Formulation
 - The result of a QA system should be presented in a way as natural as possible
 - Fusion of information from different sources
- Real Time Question Answering
 - There is need for QA systems that are capable of extracting answers from large data sets in several seconds
- Multi lingual Question Answering
 - Develop QA systems for other languages than English

3/26/2009

CSE842, Spring 2009, MSU

67

The Issues

- Interactive Question Answering
 - The questioner might want not only to reformulate the question, but (s)he might want to have a dialogue with the system
- Advanced Reasoning for Question Answering
 - More sophisticated questioners expect answers which are outside the scope of written texts or structured databases
 - A knowledge-based component of a QA system should be viewed as an augmentation, rather than a rival to the retrieval-based approaches

3/26/2009

CSE842, Spring 2009, MSU

68

The Issues

- User Profiling for Question Answering
 - Data about the questioner
 - Context data
 - Domain of interest
 - Reasoning schemes frequently used by the questioner
 - Common ground established within different dialogues between the system and the user
- Collaborative Question Answering
 - Detect users operating in the same context
 - Mine for unrelated questions that have related answers and vice versa.

3/26/2009

CSE842, Spring 2009, MSU

69