Statistical Natural Language Processing

Lecture 14: Question Answering

Dr. Momtazi
Amirkabir University of Technology

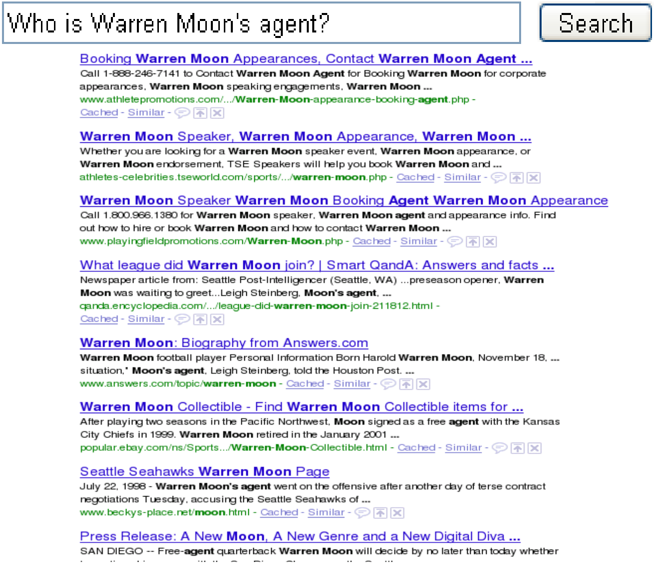
Outline

Outline

Motivation

- Finding small segments of text which answer users' questions

Motivation


Who is Warren Moon's agent?
Search
Booking Warren Moon Appearances, Contact Warren Moon Agent ...
Call 1-888-246-7141 to Contact Warren Moon Agent for Booking Warren Moon for corporate appearances, Warren Moon speaking engagements, Warren Moon ...
www.athletepromotions.com/.../Warren-Moon-appearance-booking-agent.php -
Cached - Similar -
Warren Moon Speaker, Warren Moon Appearance, Warren Moon ...
Whether you are looking for a Warren Moon speaker event, Warren Moon appearance, or Warren Moon endorsement, TSE Speakers will help you book Warren Moon and ...
athletes-celebrities.tseworld.com/sports/.../warren-moon.php - Cached - Similar -
Warren Moon Speaker Warren Moon Booking Agent Warren Moon Appearance
Call 1.800.966.1380 for Warren Moon speaker, Warren Moon agent and appearance info. Find out how to hire or book Warren Moon and how to contact Warren Moon ...
www.playingfieldpromotions.com/Warren-Moon.php - Cached - Similar -
What league did Warren Moon join? | Smart QandA: Answers and facts ...
Newspaper article from: Seattle Post-Intelligencer (Seattle, WA) ...preseason opener, Warren Moon was waiting to greet...Leigh Steinberg, Moon's agent, ...
qanda.encyclopedia.com/.../league-did-warren-moon-join-211812.html -
Cached - Similar -
Warren Moon: Biography from Answers.com
Warren Moon football player Personal Information Born Harold Warren Moon, November 18, ... situation," Moon's agent, Leigh Steinberg, told the Houston Post. ...
www.answers.com/topic/warren-moon - Cached - Similar -
Warren Moon Collectible - Find Warren Moon Collectible items for ...
After playing two seasons in the Pacific Northwest, Moon signed as a free agent with the Kansas City Chiefs in 1999. Warren Moon retired in the January 2001 ...
popular.ebay.com/ns/Sports.../Warren-Moon-Collectible.html - Cached - Similar -
Seattle Seahawks Warren Moon Page
July 22, 1998 - Warren Moon's agent went on the offensive after another day of terse contract negotiations Tuesday, accusing the Seattle Seahawks of ...
www.beckys-place.net/moon.html - Cached - Similar -
Press Release: A New Moon, A New Genre and a New Digital Diva ...
SAN DIEGO -- Free-agent quarterback Warren Moon will decide by no later than today whether

Motivation

Who is Warren Moon's agent? | Answer

Short Answers

Answers 1-5

- AGENT LEIGH STEINBERG
- MANNY RAMIREZ WILL CLARK STEVE
- QUARTERBACK WARREN
- CLARK STEVE YOUNG
- YOUNG WARREN

Search Engine vs. Question Answering

longer input →

keywords | **natural language questions**

documents | **short answer strings**

shorter output →

QA Types

Closed-domain

Answering questions from a specific domain

Open-domain

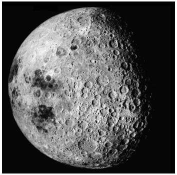
Answering any domain independent question

Outline

History

- BASEBALL [Green et al., 1963]
  - One of the earliest question answering systems
  - Developed to answer users' questions about dates, locations, and the results of baseball matches

- LUNAR [Woods, 1977]
  - Developed to answer natural language questions about the geological analysis of rocks returned by the Apollo moon missions
  - Able to answer 90% of questions in its domain posed by people not trained on the system



History

- STUDENT
  - Built to answer high-school students' questions about algebraic exercises
- PHLIQA
  - Developed to answer the user's questions about European computer systems
- UC (Unix Consultant)
  - Answered questions about the Unix operating system
- LILOG
  - Was able to answer questions about tourism information of cities in Germany

Closed-domain QA

- Closed-domain systems

- Extracting answers from structured data (database)

  Labor intensive to build

- Converting natural language questions to database queries

  Easy to implement

Open-domain QA

Closed-domain QA ⇒ Open-domain QA

Using a large collection of unstructured data (e.g., the Web) instead of databases

- Many subjects are covered
- Information is constantly added and updated
- No manual work is required to build databases

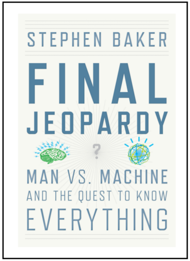
More complex systems are required

- Information is not always up-to-date
- Wrong information is not avoidable
- Much irrelevant information is found

Open-domain QA

IBM Watson

- Playing against two greatest champions of Jeopardy
- Challenges
  - Knowledge
  - Speed
  - Confidence



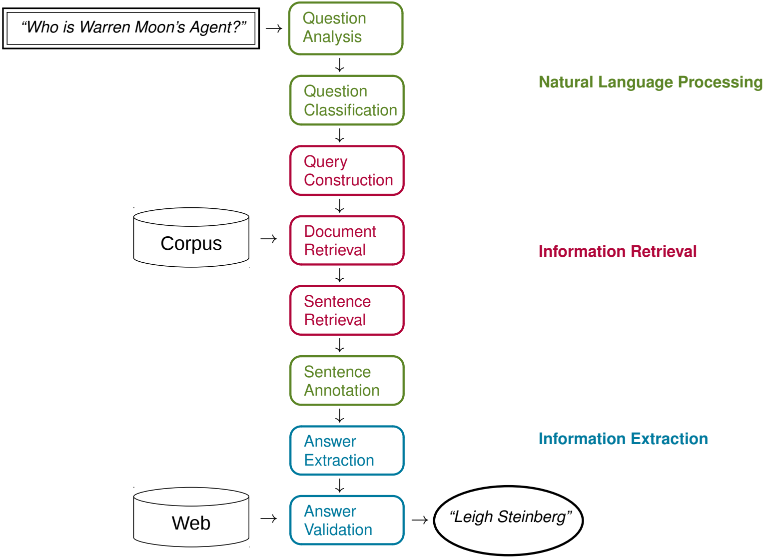
Building Watson: An Overview of the DeepQA Project

David Ferrucci, Eric Brown, Jennifer Chu-Carroll, James Fan, David Gondek, Aditya A. Kalyanpur, Adam Lally, J. William Murdock, Eric Nyberg, John Prager, Nico Schlaefer, and Chris Welty

AI Magazine, 2010

Outline

Architecture



Question Analysis

- Named Entity Recognition
- Surface Text Pattern Learning
- Syntactic Parsing
- Semantic Role Labeling

Q Analysis: Named Entity Recognition

- Recognizing the named entities in the text to extract the target of the question
- Using the question's target in the query construction step

Example:

Question: *"In what country was Albert Einstein born?"*

Target: *"Albert Einstein"*

Q Analysis: Pattern Learning

- Extracting a pattern from the question
- Matching the pattern with a list of pre-defined question patterns
- Finding the corresponding answer pattern
- Realizing the position of the answer in the sentence in the answer extraction step

Example:

Question: *“In what country was Albert Einstein born?”*

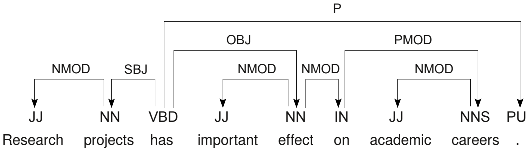
Question Pattern: *“In what country was X born?”*

Answer Pattern: *“X was born in Y.”*

Q Analysis: Syntactic Parsing

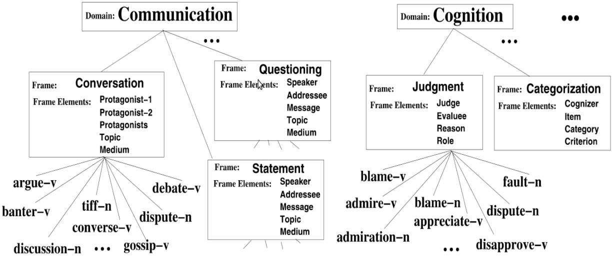
- Using a dependency parser to extract the syntactic relations between question terms

- Using the dependency relation paths between question terms to extract the correct answer in the answer extraction step



Q Analysis:
Semantic Role Labeling

- FrameNet: a lexical database for English
- More than 170,000 manually annotated sentences
- Frame Semantics: describes the type of event, relation, or entity and the participants in it.



Q Analysis: Semantic Role Labeling

- FrameNet: a lexical database for English
- More than 170,000 manually annotated sentences
- Frame Semantics: describes the type of event, relation, or entity and the participants in it.

Example:

“ $\underbrace{\textit{John}}_{\textit{Cook}}$ *grills* $\underbrace{\textit{a fish}}_{\textit{Food}}$ $\underbrace{\textit{on an open fire}}_{\textit{Heating-Instrument}}$.”

Q Analysis: Semantic Role Labeling

- Frame assignment
- Role labeling

Example:

“ $\underbrace{\textit{Jim}}_{\textbf{Driver}}$ *flew* $\underbrace{\textit{his plane}}_{\textbf{Vehicle}}$ $\underbrace{\textit{to Texas}}_{\textbf{Goal}}$.”

OPERATE–VEHICLE

Example:

“ $\underbrace{\textit{Alice}}_{\textbf{Destroyer}}$ *destroys* $\underbrace{\textit{the item}}_{\textbf{Undergoer}}$ $\underbrace{\textit{with a plane}}_{\textbf{Instrument}}$.”

DESTROYING

Q Analysis: Semantic Role Labeling

- Finding the question's head verb

Example:

" *Who* *purchased* *YouTube* ?"

$\underbrace{\text{Who}}_{\text{Buyer}}$ purchased $\underbrace{\text{YouTube}}_{\text{Goods}}$

COMMERCE−BUY

- Buyer [Subj,NP] verb Goods [Obj,NP]
- Buyer [Subj,NP] verb Goods [Obj,NP] Seller [Dep,PP-from]
- Goods [Subj,NP] verb Buyer [Dep,PP-by]
- ...

Example:

" *In 2006, YouTube was purchased by Google for $1.65 billion.*"

In 2006, $\underbrace{\text{YouTube}}_{\text{Goods}}$ was purchased $\underbrace{\text{by Google}}_{\text{Buyer}}$ for $1.65 billion.

Question Classification

- Classifying the input question into a set of question types
- Mapping question types to the available named entity labels
- Finding strings that have the same type as the input question in the answer extraction step

Example:

Question: *“In what country was Albert Einstein born?”*

Type: LOCATION - Country

Question Classification

- Classifying the input question into a set of question types
- Mapping question types to the available named entity labels
- Finding strings that have the same type as the input question in the answer extraction step

Example (NER):

S1: “ $\underbrace{\textit{Albert Einstein}}_{\textbf{Person}}$ *was born in* $\underbrace{\textit{14 March 1879}}_{\textbf{Date}}$ *.”*

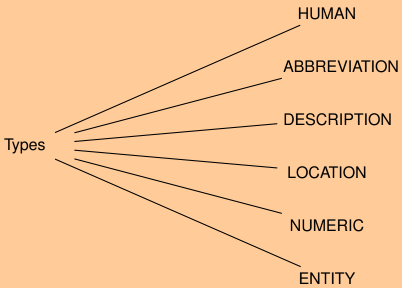
S2: “ $\underbrace{\textit{Albert Einstein}}_{\textbf{Person}}$ *was born in* $\underbrace{\textit{Germany}}_{\textbf{Country}}$ *.”*

S3: “ $\underbrace{\textit{Albert Einstein}}_{\textbf{Person}}$ *was born in a* $\underbrace{\textit{Jewish}}_{\textbf{Religion}}$ *family.”*

Question Classification

- Classification taxonomies
  - BBN
  - Pasca & Harabagiu
  - Li & Roth

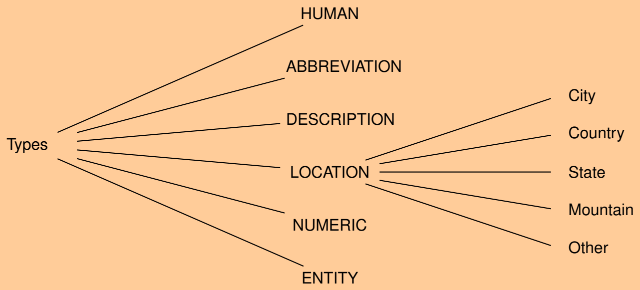
6 coarse- and 50 fine-grained classes



Question Classification

- Classification taxonomies
  - BBN
  - Pasca & Harabagiu
  - Li & Roth

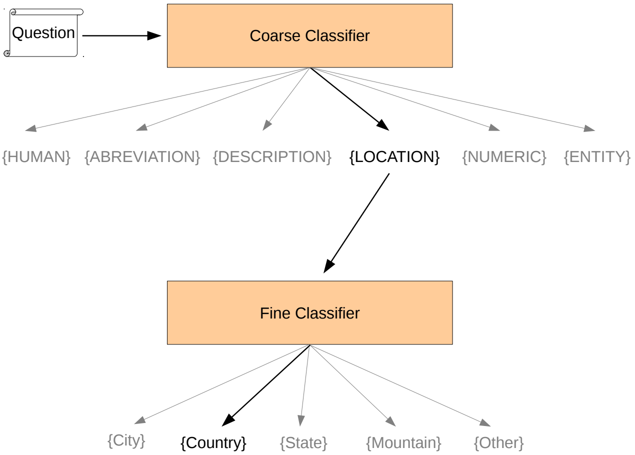
6 coarse- and 50 fine-grained classes



Question Classification

Question	Type	Sub-type
"Who killed Gandhi?"	HUMAN	Individual
"Who has won the most Super Bowls?"	HUMAN	Group
"What city did Duke Ellington live in?"	LOCATION	City
"Where is the highest point in Japan?"	LOCATION	Mountain
"What do sailors use to measure time?"	ENTITY	Technique
"Who is Desmond Tutu?"	DESCRIPTION	human

Question Classification



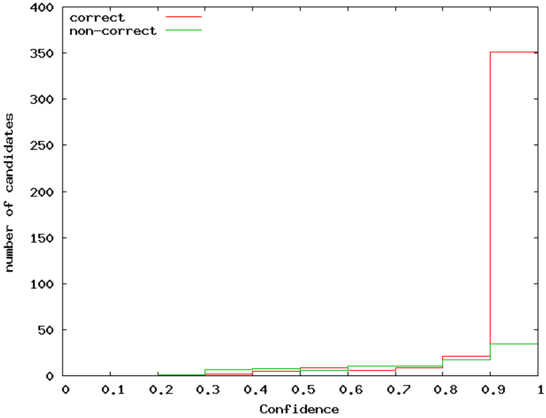
Question Classification

- Using any kinds of supervised classifiers
  - *K* Nearest Neighbor
  - Support Vector Machines
  - Naïve Bayes
  - Maximum Entropy
  - Logistic Regression
  - ...

- Benefiting from available toolkits
  - Support Vector Machine: SVM-light
  - Maximum Entropy: Maxent, Yasmet

Question Classification

- Considering the confidence measure of the classification to filter the result



Query Construction

- Goal:
  - Formulating a query with a high chance of retrieving relevant documents

- Task:
  - Assigning a higher weight to the question's target
  - Using query expansion techniques to expand the query

Document Retrieval

- Importance:
  - QA components use computationally intensive algorithms
  - Time complexity of the system strongly depends on the size of the to be processed corpus

- Task:
  - Reducing the search space for the subsequent components
  - Retrieving relevant documents from a large corpus
  - Selecting top *n* retrieved document for the next steps

Document Retrieval

- Using available information retrieval models
  - Vector Space Model
  - Probabilistic Model
  - Language Model

- Using available information retrieval toolkits

Sentence Retrieval

- Task:
  - Finding small segments of text that contain the answer

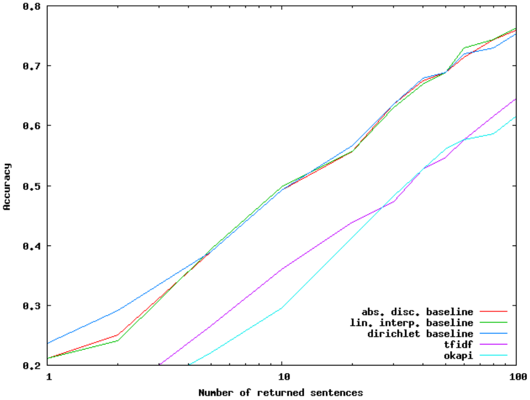
- Benefits beyond document retrieval:
  - Documents are very large
  - Documents span different subject areas
  - The relevant information is expressed locally
  - Retrieving sentences simplifies the answer extraction step

Sentence Retrieval

- Information retrieval models for sentence retrieval
  - Vector Space Model
  - Probabilistic Model
  - Language Model
    - Jelinek-Mercer Linear Interpolation
    - Bayesian Smoothing with Dirichlet Prior
    - Absolute Discounting

Sentence Retrieval

- Comparing language modeling with traditional methods



Sentence Retrieval

- Comparison of the effects of text length on information retrieval

	MAP	P @ 10	P @ 20
Documents	0.191	0.232	0.200
750 bytes	0.064	0.186	0.149
500 bytes	0.055	0.166	0.142
250 bytes	0.036	0.136	0.117
Sentences	0.030	0.098	0.081

based on work by Vanessa Murdock

- Main problem of sentence retrieval: sentence brevity

Sentence Retrieval

- Challenges:
  - The brevity of sentences
    - The term mismatch problem in sentence retrieval is more critical than document retrieval
    - Even if the query words appear in a sentence, the frequency of those words is not higher than one or two
  - The notion of relevance
    - In addition to aboutness, the sentence must contain the answer

Example:

Question: *“Who invented the automobile?”*

Sentence: *“The first car was built by Karl Benz.”*

Sentence Retrieval

- Approaches to overcome the sentence brevity problem:
  - Query expansion, (pseudo-)relevance feedback

	P @ 10	P @ 20
Query Likelihood	0.111	0.096
Query Expansion	0.044	0.040
Pseudo-Relevance Feedback	0.113	0.094

based on work by Vanessa Murdock

Sentence Retrieval

- Approaches to overcome the sentence brevity problem:
  - Query expansion, (pseudo-)relevance feedback

    Do not work at sentence-level retrieval

  - Term relationship models
    - Translation model
    - Term clustering model

Sentence Retrieval

- Translation Model
  - Considering the relationship between sentence and query words
  - Estimating the probability of generating a query as a translation of a sentence

Word model:

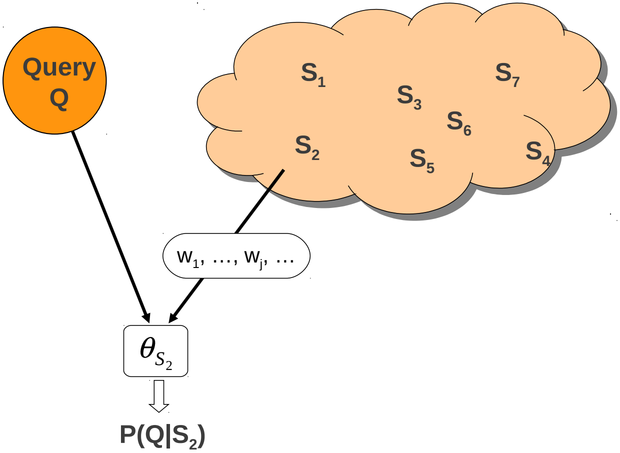
$$P(Q|S) = \prod_{i=1}^{M} P(q_i|S)$$

Translation model:

$$P(Q|S) = \prod_{i=1}^{M} \sum_{t \in S} P(q_i|t) \cdot P(t|S)$$

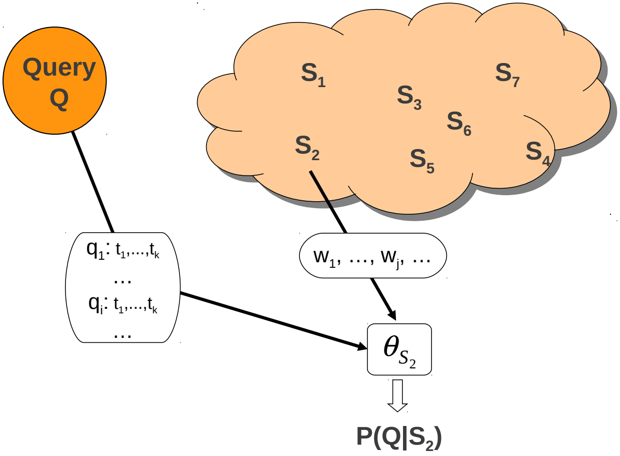
Sentence Retrieval

- Word Model



Sentence Retrieval

- Translation Model



Sentence Retrieval

- Class Model
  - Using a word clustering algorithm to cluster lexical items
  - Assigning similar words to the same cluster
  - Estimating the probability of a query term given a sentence based on the cluster which the query term belongs to

Word model:

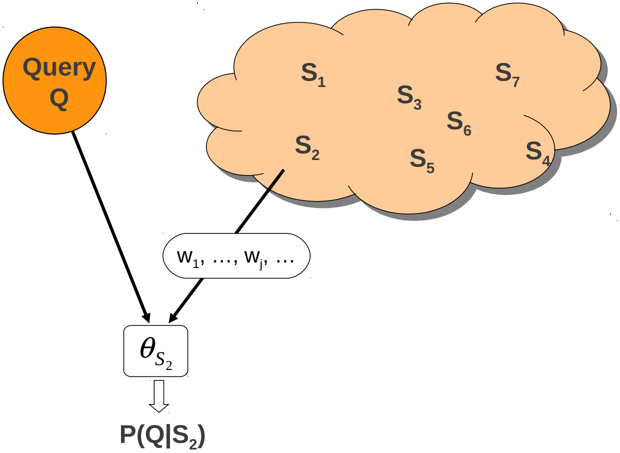
$$P(Q|S) = \prod_{i=1}^{M} P(q_i|S)$$

Class model:

$$P(Q|S) = \prod_{i=1}^{M} P(q_i|C_{q_i}, S) \cdot P(C_{q_i}|S)$$

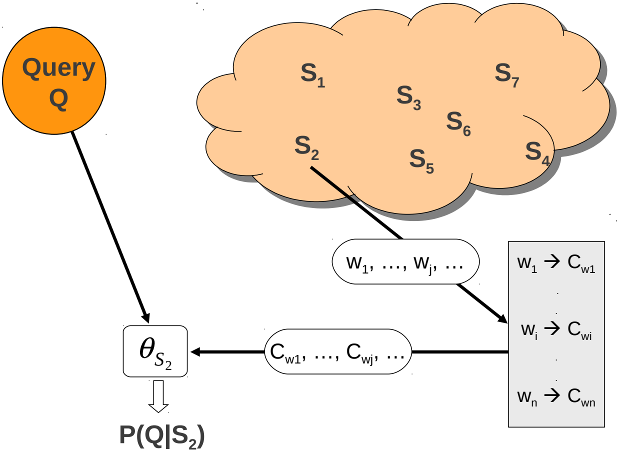
Sentence Retrieval

- Word Model



Sentence Retrieval

- Class Model



Sentence Retrieval

- Approaches to overcome the sentence relevance problem:
  - Modeling the expected answer type into language model
  - Finding the relation between query words and named entities

Example:

Question: *"When was Albert Einstein born?"*

Sent1: *"$\underbrace{\text{Albert Einstein}}_{\textit{Person}}$ was born in $\underbrace{\text{14 March 1879}}_{\textit{Date}}$."*

Sent2: *"$\underbrace{\text{Albert Einstein}}_{\textit{Person}}$ was born in $\underbrace{\text{Germany}}_{\textit{Location}}$."*

Sentence Retrieval

- Approaches to overcome the sentence relevance problem:
  - Modeling the expected answer type into language model
    - Expanding the query with the expected answer type
    - Expanding each sentence with the expected answer type, if it exists in the sentence

Example:

Question: *"When was Albert Einstein born?"* Date

Sent1: *"Albert Einstein was born in 14 March 1879."* Date

Sent2: *"Albert Einstein was born in Germany."*

Sentence Retrieval

- Approaches to overcome the sentence relevance problem:
  - Finding the relation between query words and named entities
    - Training a model on a corpus of question and answer pairs
    - Using the named-entity of the sentences besides the sentence words

$$P(Q|S) = \prod_{i=1}^{M} \sum_{t \in S} P(q_i|t) \cdot P(t|S)$$

$\boxed{t\text{: named entity label}}$

Example:

Question: *"When was Albert Einstein born?"*

Sent1: *"Albert Einstein was born in 14 March 1879."* Person Date

Sent2: *"Albert Einstein was born in Germany."* Person Location

Sentence Annotation

- Annotating relevant sentences using linguistic analyses
  - Named entity recognition
  - Dependency parsing
  - Noun phrase chunking
  - Semantic role labeling

Similar to Question Analysis

Example:

Question: *"In what country was Albert Einstein born?"*

Sentence Annotation

- Annotating relevant sentences using linguistic analyses
  - Named entity recognition
  - Dependency parsing
  - Noun phrase chunking
  - Semantic role labeling

Similar to Question Analysis

Example (NER):

Sentence1: “ $\underbrace{\textit{Albert Einstein}}_{\text{Person}}$ *was born in* $\underbrace{\textit{14 March 1879}}_{\text{Date}}$.”

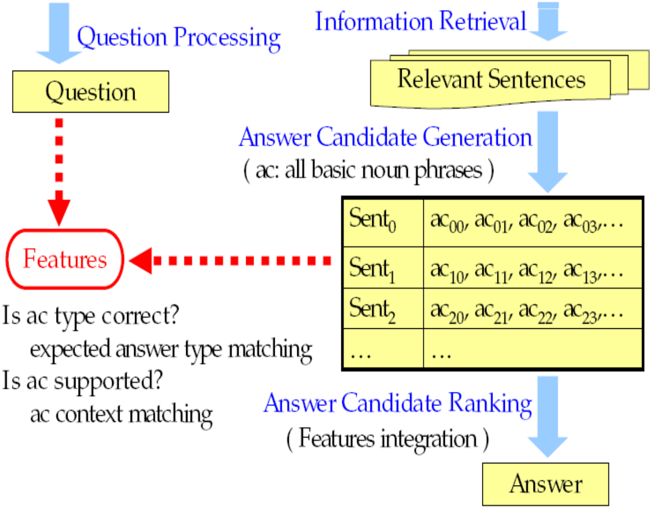
Sentence2: “ $\underbrace{\textit{Albert Einstein}}_{\text{Person}}$ *was born in* $\underbrace{\textit{Germany}}_{\text{Country}}$.”

Sentence3: “ $\underbrace{\textit{Albert Einstein}}_{\text{Person}}$ *was born in a* $\underbrace{\textit{Jewish}}_{\text{Religion}}$ *family.*”

Answer Extraction

- Extracting candidate answers based on various information
  - Question
    - Question Analysis: patterns
    - Question Analysis: dependency parse
    - Question Analysis: semantic roles
    - Question Classification: question type
  - Sentence
    - Sentence Annotation: all annotated data

Answer Extraction



Answer Extraction

- Extracting candidate answers based on various information

Example (Pattern):

Sentence1: *"Albert Einstein was born in 14 March 1879."*

Sentence2: *"Albert Einstein was born in Germany."*

Sentence3: *"Albert Einstein was born in a Jewish family."*

Answer Extraction

- Extracting candidate answers based on question type and entity type

Example:

Question: *"In what country was Albert Einstein born?"*

Question Type: LOCATION - Country

Answer Extraction

- Extracting candidate answers based on question type and entity type

Example (NER):

Sentence1: “$\underbrace{\textit{Albert Einstein}}_{\textbf{Person Name}}$ *was born in* $\underbrace{\textit{14 March 1879}}_{\textbf{Date}}$.”

Sentence2: “$\underbrace{\textit{Albert Einstein}}_{\textbf{Person Name}}$ *was born in* $\underbrace{\textit{Germany}}_{\textbf{Country}}$.”

Sentence3: “$\underbrace{\textit{Albert Einstein}}_{\textbf{Person Name}}$ *was born in a* $\underbrace{\textit{Jewish}}_{\textbf{Religion}}$ *family*.”

Answer Extraction: Pattern based Approach

- Pattern based approach

Example:

Question: *"In what country was Albert Einstein born?"*

Question Pattern: In what country was X born?
Answer Pattern: X was born in Y.

Answer Extraction

- Different wordings possible, but similar syntactic structure

Q: Who founded the Black Panthers organization ?

subj obj

S1: Bobby Seale, a student at Merritt College, founded the Black Panther Party for self-defense .

subj obj

S2: The Black Panther Party, co-founded by Seale and Newton, flourished...

obj subj

based on work by Dan Shen

Answer Extraction

- Many syntactic variations → need robust matching approach

Q: Who founded the Black Panthers organization ?

subj obj

S1: Bobby Seale, a student at Merritt College, founded the Black Panther Party for self-defense .

subj obj

S2: The Black Panther Party, co-founded by Seale and Newton, flourished...

obj subj

S3: Hilliard introduced Bobby Seale, who co-founded the Black Panther Party here .

rel, i obj

S4: Black Panthers Co-founder Bobby Seale visits UMM.

nn appo

Answer Extraction

- Using semantic roles

Example:

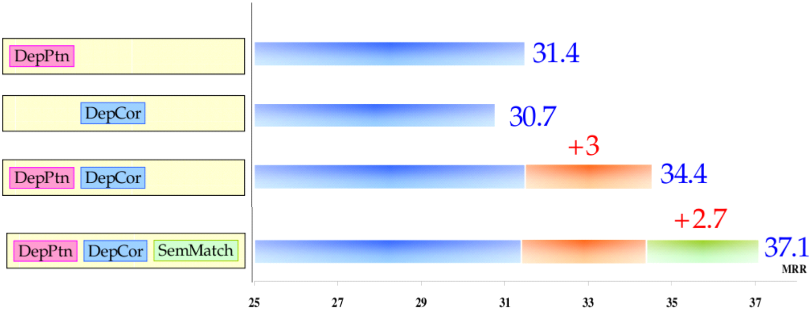
“$\underbrace{\textit{Who}}_{\textbf{Buyer}}$ *purchased* $\underbrace{\textit{YouTube}}_{\textbf{Goods}}$?”

Example:

“ *In 2006,* $\underbrace{\textit{YouTube}}_{\textbf{Goods}}$ *was purchased* $\underbrace{\textit{by Google}}_{\textbf{Buyer}}$ *for \$1.65 billion.*”

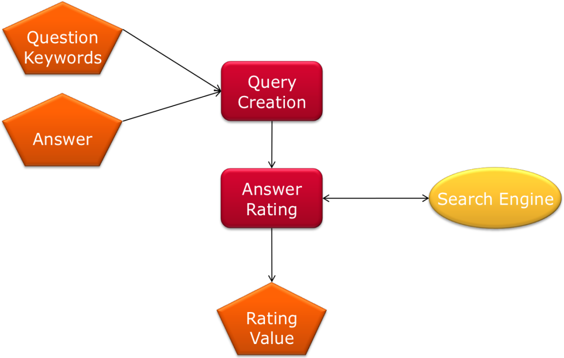
Answer Extraction

- Comparing answer extraction features



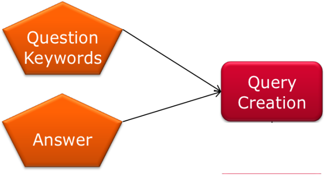
Answer Validation

- Using Web as a knowledge resource for validating answers



Answer Validation

- Combining the answer with a subset of the question keywords
- Using sequences of keywords, if available
- Choosing different combinations of subsets
  - Bag-of-Word
  - Noun-Phrase-Chunks
  - Declarative-Form



Answer Validation

- Query model:
  - Bag-of-Word
  - Noun-Phrase-Chunks
  - Declarative-Form

Example:

Question: *"In what country was Albert Einstein born?"*

Answer Candidate: *Germany*

Answer Validation

- Query model:
  - Bag-of-Word
  - Noun-Phrase-Chunks
  - Declarative-Form

Bag-of-Word:

Albert Einstein born Germany

Noun-Phrase-Chunks:

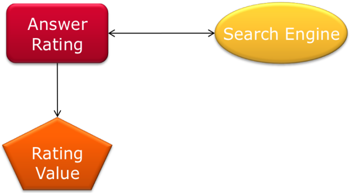
"Albert Einstein" born Germany

Declarative-Form:

"Albert Einstein born Germany"

Answer Validation

- Passing the query to a search engine
- Analyzing the result of the search engine
  - Counting the results
  - Parsing the result snippets
- Other possibilities:
  - Using knowledge bases to find relations between the question keywords and the answer
    - More complex
    - Good for special question types
    - Quality depends on the quality of the knowledge bases

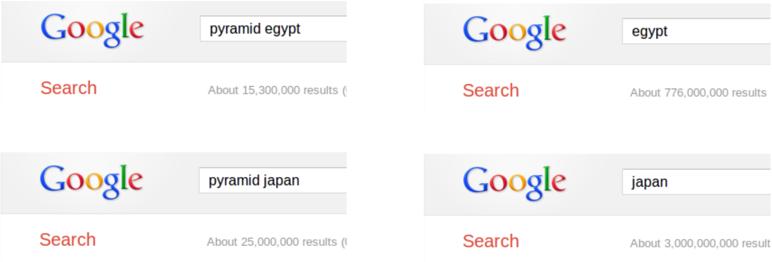


Answer Validation

Example:

Question: *“Where are Pyramids?”*

Answer Candidate 1: *“Egypt”*
Answer Candidate 2: *“Japan”*

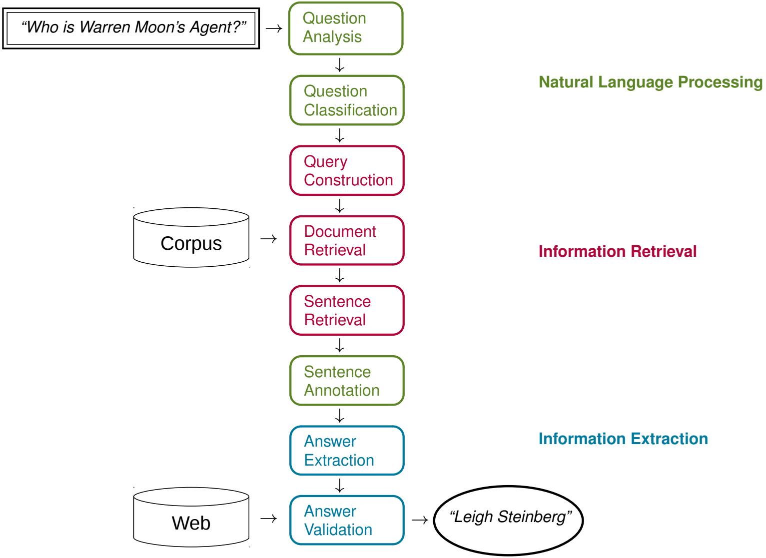


Answer Validation

- Modeling the keywords' associations

Query	Hits (question+answer)	Hits (answer)	Association
pyramid egypt	15.3 M	766 M	0.019
pyramid japan	25 M	3000 M	0.008

Architecture



Fact vs. Opinion

- Factual questions

"Who is Warren Moon's Agent?"

"When was Mozart born?"

"In what country was Albert Einstein born?"

"Who is the director of the Hermitage Museum?"

Fact vs. Opinion

- Opinionated questions

"What do people like about Wikipedia?"

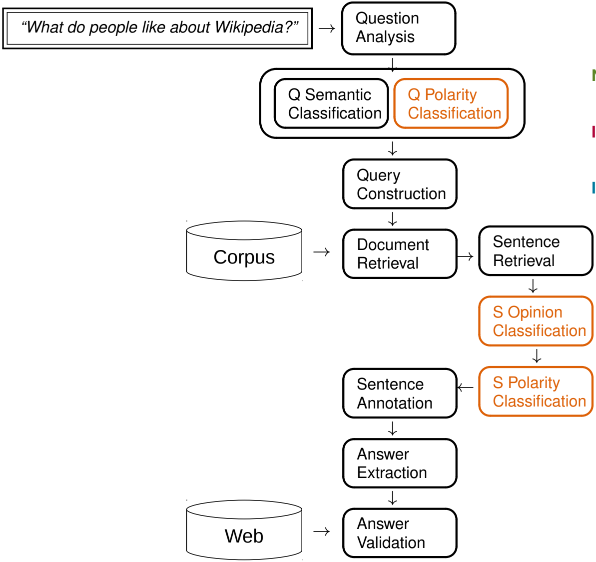
"What organizations are against universal health care?"

"What are the public opinions on human cloning?"

"What were the most popular movies in 2010?"

Outline

Architecture



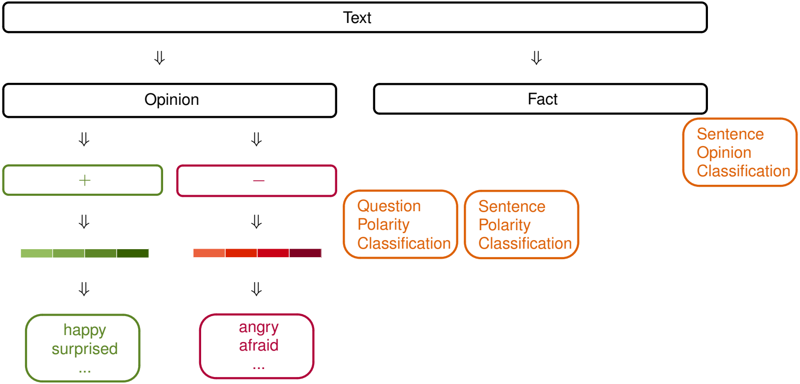
Natural Language Processing

Information Retrieval

Information Extraction

Opinion Mining

Opinion Mining



Question Polarity Classification

- Running in parallel with question classification

- Task:
  - Deciding whether the input question has a positive or negative polarity

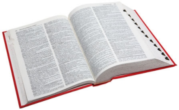
Example:

 *"What do people like about Wikipedia?"*

 *"Why people hate reading Wikipedia articles?"*

Question Polarity Classification

- Approaches:
  - Machine learning
    - Running a classifier trained on an annotated corpus
    - Using the subjectivity lexicon for feature selection
  - Rule-based
    - Identifying opinion words in the text using a subjectivity lexicon



good love brave intelligent kind ...	bad angry lie hate poor ...

Question Polarity Classification

Corpus	Method	Feature Set	Accuracy
MPQA	data-driven	in-domain vocabulary	57.58
MPQA	data-driven	Subjectivity Lexicon	60.61
NTCIR	data-driven	in-domain vocabulary	54.55
NTCIR	data-driven	Subjectivity Lexicon	63.64
—	rule-based	Subjectivity Lexicon	**93.94**

- lack of matching training data $\rightarrow$ rule based best

based on work by Michael Wiegand

Sentence Opinion Classification

- Importance:
  - Sentence retrieval output is mixed (factual & opinionated)
  - Opinion question answering systems are looking for opinionated sentences

- Goal:
  - Classifying retrieved sentences as opinionated or factual

Example:

Question: *"What do people like about Wikipedia?"*

Sentence Opinion Classification

- Importance:
  - Sentence retrieval output is mixed (factual & opinionated)
  - Opinion question answering systems are looking for opinionated sentences

Example:

S1: *"I agree Wikipedia is very much in handy when your online; however, I can not use it when I am not online."*

S2: *"Wikipedia began as a complementary project for Nupedia, a free online English-language encyclopedia project."*

S3: *"Jimmy Wales and Larry Sanger co-founded Wikipedia in January 2001."*

S4: *"Wikipedia is a great way to access lots of information."*

Sentence Opinion Classification

- Importance:
  - Sentence retrieval output is mixed (factual & opinionated)
  - Opinion question answering systems are looking for opinionated sentences

Opinion

S1: *"I agree Wikipedia is very much in handy when your online; however, I can not use it when I am not online."*

S4: *"Wikipedia is a great way to access lots of information."*

Fact

S2: *"Wikipedia began as a complementary project for Nupedia, a free online English-language encyclopedia project."*

S3: *"Jimmy Wales and Larry Sanger co-founded Wikipedia in January 2001."*

Sentence Opinion Classification

- Importance:
  - Sentence retrieval output is mixed (factual & opinionated)
  - Opinion question answering systems are looking for opinionated sentences

- Goal:
  - Classifying retrieved sentences as opinionated or factual

- Approaches:
  - Running a classifier trained on an annotated corpus
    - SVM
    - Maximum Entropy
    - Naive Bayes

Sentence Polarity Classification

- Task:
  - Distinguishing positive and negative sentences
  - Returning sentences which have the same polarity as the input question

- Approach:
  - Using a classifier to classify opinionated sentences as positive or negative
    - Using a small set of lightweight linguistic polarity features
    - Considering the distance between polarity features and the topic in the sentence
    - Using a dependency parser to consider syntactic features

Outline

Text REtrieval Conference (TREC)

- Goal:
  - Preparing an infrastructure necessary for the large-scale evaluation of text retrieval technologies

- Contributions:
  - Introducing standard test sets
  - Providing large data collections
  - Using standard evaluation metrics

Text REtrieval Conference (TREC)

- Co-sponsors:
  - National Institute for Science and Technology (NIST)
  - Defense Advanced Research Projects Agency (DARPA)

- Start date: 1992

QA at TREC

- 1999-2007:
  - Question answering as one of the TREC subtasks

- 2008:
  - Opinion question answering as a subtask in the newly established Text Analysis Conference (TAC)

QA at TREC

- Participation
  - Releasing a data collection and a set of input queries by TREC annually
  - Running question answering systems on the provided data by participants
  - Submitting the retrieved data within one week of the evaluation period
  - Accessing the results manually and evaluating them by NIST

QA at TREC

- Data collection at TREC
  - 1999-2001: 528,000 news articles
  - 2002-2007: more than 1,000,000 news articles from
    - Xinhua News Service (XIN)
    - New York Times News Service (NYT)
    - Associated Press Worldstream News Service (APW)

Question Classes

Factoid questions

Q: *“In what year was Albert Einstein born?”*

List questions

Q: *“List the names of books written by Herman Hesse.”*

Definition questions

Q: *“What is the Lambda calculus?”*

Question Classes

Question	Class
"How far is it from Earth to Mars?"	factoid
"What book did Rachel Carson write in 1962?"	factoid
"How many time zones are there in the world?"	factoid
"When was Prince Charles born?"	factoid
"Who are professional female boxers?"	list
"List brands of pianos."	list
"What is a golden parachute?"	definition
"Who is Allen Iverson?"	definition

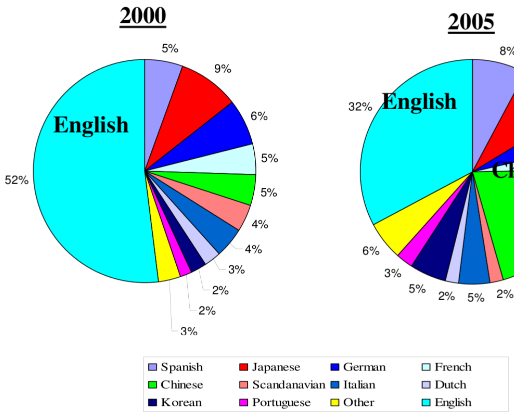
Cross Language Evaluation Forum (CLEF)

- Goal:
  - Promoting R&D in multilingual information access

- Contributions:
  - Developing an infrastructure for testing, tuning, and evaluating of information retrieval systems
  - Focusing on European languages in both monolingual and cross-language contexts
  - Creating test suites of reusable data to be employed by system developers for benchmarking purposes

Cross Language Evaluation Forum (CLEF)

- Language distribution on the Web



Source: Global Reach

QA at CLEF

- Question answering as one of the CLEF subtasks (2003-2008)

- Similar to the TREC QA task, but with questions and documents in different languages
  - Using queries expressed in a single language
  - Finding answers from documents written in any language

- Covering various languages
  - 2003: 3 languages
  - ...
  - 2008: 10 languages

QA at CLEF

- Data collection at CLEF
  - 2003-2004: News articles (1994)
  - 2005-2006: + News articles (1995)
  - 2007-2008: + Wikipedia articles (2006)

Outline

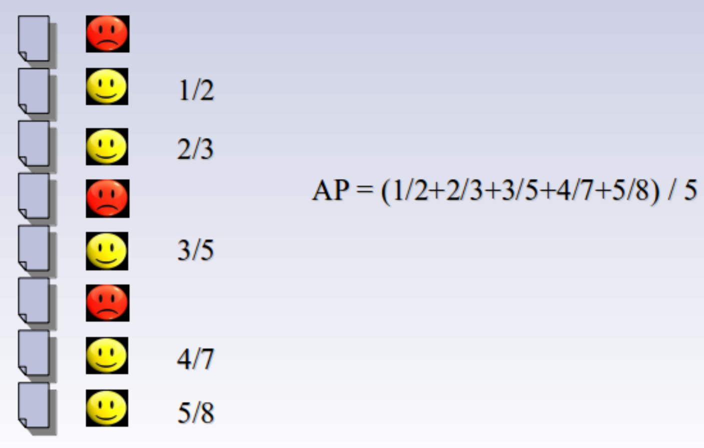
Evaluation

- Evaluate ranking, not just Boolean classification
- Idea: Calculate precision and recall at every rank position

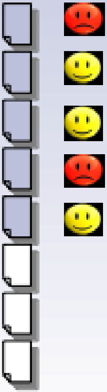
Evaluation

- Problem: Long lists are unwieldy and difficult to compare
- Ideas:
  - Calculating precision at standard recall levels, from 0.0 to 1.0 in increments of 0.1 => “Precision-Recall Curve”
  - Averaging the precision values from the rank positions where a relevant document was retrieved => “Average Precision”
  - Calculating precision at small number of fixed rank positions = > “Precision at rank k”
    - Ignores ranking after p; ignores ranking within 1 to p
  - Reciprocal of the rank at which the first relevant document is retrieved = > “Reciprocal Rank”
    - Very sensitive to rank position, regards only first relevant document

Average Precision

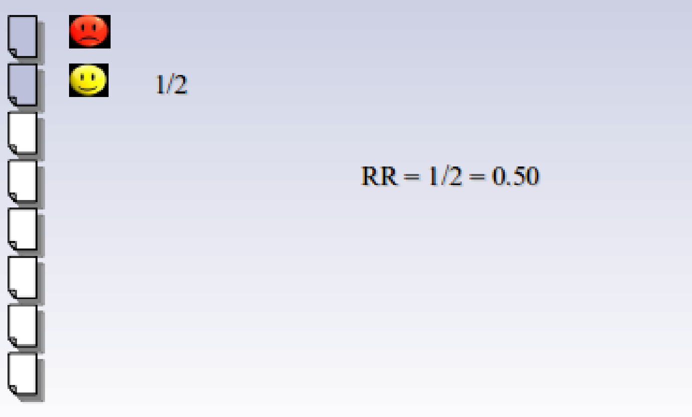


Precision at rank k



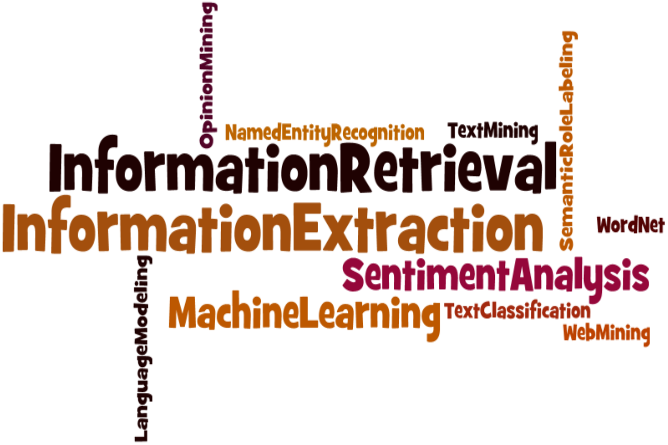
Precision@5 = 3/5 = 0.60

Reciprocal Rank



Outline

Summary


OpinionMining
NamedEntityRecognition
TextMining
SemanticRoleLabeling
InformationRetrieval
InformationExtraction
WordNet
SentimentAnalysis
LanguageModeling
MachineLearning
TextClassification
WebMining