Question Answering Systems

An Introduction

Potsdam, Germany, 14 July 2011

Saeedeh Momtazi
Information Systems Group
Outline

1 Introduction
Outline

1 Introduction

2 History
Outline

1. Introduction

2. History

3. QA Architecture
   - Factoid QA
   - Opinion QA
Outline

1. Introduction
2. History
3. QA Architecture
   - Factoid QA
   - Opinion QA
4. Summary
Outline

1. Introduction

2. History

3. QA Architecture
   - Factoid QA
   - Opinion QA

4. Summary
Why QA
Why QA
Why QA

Who is Warren Moon's agent?
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...

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.teeworld.com/sports/.../warren-moon.php

Warren Moon Speaker Warren Moon Booking Agent Warren Moon Appearance
Call 1.800.996.1380 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...

What league did Warren Moon join? | Smart QandA: Answers and facts...
Newspaper article from: Seattle Post-Intelligencer (Seattle, WA) preseason opener, Warren Moon was waiving to greet...Leigh Steinberg, Moon's agent, ...
qanda.encyclopedia.com/.../league-did-warren-moon-join-211812.html

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

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

Seattle Seahawks Warren Moon Page
July 22, 1990 - 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

Press Release: A New Moon, A New Genre and a New Digital Diva...
Why QA

Warren Moon

Sports Celebrity Category: Football Celebrities, Hall of Fame Athletes

Warren Moon Available For: Corporate Appearances, Endorsements, Speaking & Autograph Signings

Warren Moon Appearance Booking Fee Range: Call 1.888.246.7141 for Booking Fees

Contact Booking Agent to Hire Warren Moon for an Appearance or Endorsement

At AthletePromotions.com, we provide information about Warren Moon’s accomplishments, achievements, statistics, accolades and appearance booking fees. We assist corporations in finding Warren Moon’s agent, contacting Warren Moon’s management company and business manager, booking Warren Moon appearances, hiring Warren Moon for endorsements, Warren Moon autograph signings, and Warren Moon speaking engagements. We are an athlete booking agency and sports marketing agency that also hires Warren Moon for corporate event appearances, celebrity golf tournaments, trade shows, conventions, store grand openings, licensing deals, print advertising and television commercials.

Warren Moon is now available for corporate event appearances, personal appearances, casino appearances, corporate appearances, corporate golf tournaments, sports camps, autograph signings, endorsement deals, television commercials, radio commercials, store grand openings, new product launch campaigns, spokesperson campaigns and speaking appearances. Book Warren Moon to meet and mingle with your best corporate clients, friends and business associates.

If your company is interested in finding out booking fees and availability for hiring Warren Moon for an appearance, endorsement or speaking engagement, call us at 1.888.246.7141 or fill out the form below.

>>> Submit Warren Moon Appearance Booking Request Here

WARREN MOON VIDEO HIGHLIGHTS:

City Chiefs in 1999. Warren Moon retired in the January 2001...

Seattle Seahawks Warren Moon Page

July 22, 1999 - Warren Moon’s agent went on the offensive after another day of tense contract negotiations Tuesday, accusing the Seattle Seahawks of...

Press Release: A New Moon, A New Genre and a New Digital Diva...
Why QA

Saeedeh Momtazi | QA | 14 July 2011

Press Release: A New Moon, A New Genre and a New Digital Diva...
Why QA

SPORTS & CELEBRITY TALENT AGENCY

CALL US TOLL FREE
800.966.1380

Recent Bookings:
Ron Jaworski for Glaxo Smith Kline
Clintont Portis for Pizza Hut
Konnio Lott for Hewlett-Packard
Dusty Baker for Pepsi

You are here: Speakers Bureau > Sports Speakers > Football > Warren Moon

Warren Moon

Category:
Football Hall of Famers, Keynote Speakers, Motivational Speakers, Popular Speakers, Sports Figures, Sports Motivation, Sports Speakers

Title:
Hall Of Fame Quarterback, Houston Oilers & Minnesota Vikings

Available For:
Appearances, Endorsements and Speaking Engagements

Travels From:
Washington

Fee Range:
$10,001 - $20,000

Book or Hire Warren Moon

Biography

PPF is a speakers bureau and booking agency providing Warren Moon speaker, personal appearance and agent information. PPF is the booking agent coordinator for Warren Moon. Here you can find Warren Moon agent information including how to hire Warren Moon for a speaking engagement appearance, autograph signing, endorsement deal or corporate event. Contact Warren Moon booking agent for more information on speaker fee costs, scheduling and availability for your next event.

Football great, Warren Moon, was born on November 18, 1955 in Los Angeles, California. Warren Moon played quarterback for the Canadian Football League's Edmonton Eskimos and the NFL Houston Oilers, Minnesota Vikings, Seattle Seahawks and Kansas City Chiefs. He is currently a broadcaster for the Seattle Seahawks and is also a business partner with his longtime friend and agent, Leigh Steinberg of Leigh Steinberg Sports & Entertainment out of Newport Beach, California.

Prior to the 1979 NFL Draft, some NFL scouts suggested that since the lack of interest in black quarterback Warren Moon had played in a

Press Release: A New Moon, A New Genre and a New Digital Diva ...

Saeedeh Momtazi | QA | 14 July 2011
Why QA

Saeedeh Momtazi | QA | 14 July 2011

FPF is a speakers bureau and booking agency providing Warren Moon speaker, personal appearance and agent information. PFP is the booking agent coordinator for Warren Moon. Here you can find Warren Moon agent information including how to hire Warren Moon for a speaking engagement, appearance, autograph signing, endorsement deal or corporate event. Contact Warren Moon booking agent for more information on speaker fees, scheduling and availability for your next event.

Football great, Warren Moon, was born on November 16, 1955 in Los Angeles, California. Warren Moon played quarterback for the Canadian Football League’s Edmonton Eskimos and the NFL Houston Oilers, Minnesota Vikings, Seattle Seahawks and Kansas City Chiefs. He is currently a broadcaster for the Seattle Seahawks and also a business partner with his lifelong friend and agent, Leigh Steinberg at Leigh Steinberg Sports & Entertainment out of Newport Beach, California.

Prior to the 1978 NFL Draft, some NFL scouts suggested that since University of Washington quarterback Warren Moon had played in a

Press Release: A New Moon, A New Genre and a New Digital Diva ...

Saeedeh Momtazi | QA | 14 July 2011
Why QA

Saeedeh Momtazi | QA | 14 July 2011

Press Release: A New Moon, A New Genre and a New Digital Diva ...
Why QA

It's never easy to put your best on the line. Showing great restraint, Moon refused to acknowledge the heckling, and when the boos turned to cheers he accepted the praise without bitterness.

His tenacity was rewarded in 1977, when the Huskies won their conference championship and met the University of Michigan in the 1978 Rose Bowl game. The underdog Huskies won the Rose Bowl under Moon's leadership, and he was named Rose Bowl Most Valuable Player and the Pacific-8 Player of the Year. Overall, Moon passed for 3,277 yards and 19 touchdowns in his collegiate career.

Although Moon managed to win over Washington's fans, he failed to convince skeptical NFL scouts of his playing ability. His Rose Bowl performance notwithstanding, he was rated just the tenth best quarterback in the 1978 draft. "The stereotype was that he was a black quarterback and he was going to run around like a madman, but he wouldn't be able to throw very well," former Edmonton Eskimos and Houston Oilers coach Hugh Campbell told the Los Angeles Times. So, once again, Moon decided to prove himself elsewhere, signing with the Eskimos of the Canadian Football League.

During Moon's six seasons in Canada, he put up some stunning numbers: 21,230 yards passing and 1,700 yards rushing. He had back-to-back 5,000-yard passing seasons. His 5,649 yards passing over 15 games in 1983 remains an all-time high for pro football. In addition, the Eskimos won five straight Grey Cup trophies as champions of the CFL from 1978 to 1982.

By 1984, Moon had nothing left to prove. When his contract with Edmonton expired, seven NFL teams sought to sign him as a free agent. Moon initially leaned toward the Seattle Seahawks, which would allow him to return to his college town, but he eventually chose the Houston Oilers, the team that had hired his former Edmonton coach, Campbell. The Oilers tendered a five-year, $5.5 million contract which, at the time, made Moon the highest paid player in the NFL—before he even played in a league game.

When Moon joined Houston, it was the sorriest franchise in the NFL, having won only three games in the previous two seasons. "One of the challenges of Houston was to be part of a growing situation," Moon's agent, Leigh Steinberg, told the Houston Post. "He knew it would take longer [to be on a championship team], but when it came, he knew he would be an instrumental part of the building process."

In 1984 Moon was a rookie sensation. His six years in the CFL gave him a wealth of experience, and he threw for a then-Houston-record 3,336 yards on the season. Still the Oilers went 3-13, finishing last in their division. The next season, after the club won just five of its first 14 games, Campbell was fired and a defensive-oriented coach, Jerry Glanville, took over. "Those early years [in Houston] were really hard for me to deal with at first," Moon told the St. Louis Post Dispatch. "There were some uncertainties about my career here because of the coaching change. That left..."
Why QA

Search Engine Marketing
Why QA
Who is Warren Moon's agent?
Who is Warren Moon's agent?

**Short Answers**

**Answers 1-5**

- AGENT LEIGH STEINBERG
- MANNY RAMIREZ WILL CLARK STEVE
- QUARTERBACK WARREN
- CLARK STEVE YOUNG
- YOUNG WARREN
QA vs. SE

- Longer input
  - Keywords, natural language questions
  - Documents, short answer strings
- Shorter output

Saeedeh Momtazi | QA | 14 July 2011
QA vs. SE

longer input

Search Engine Marketing

Google

Yahoo!

Q&A
QA vs. SE

longer input

keywords

natural language questions
QA vs. SE

longer input

keywords

natural language questions

shorter output
**QA vs. SE**

- **longer input**
  - keywords
  - documents

- **shorter output**
  - natural language questions
  - short answer strings
Closed-domain
only answer questions from a specific domain.

Open-domain
answer any domain independent question.
Outline

1 Introduction

2 History

3 QA Architecture
   Factoid QA
   Opinion QA

4 Summary
History

- **BASEBALL [Green et al., 1963]**
  - One of the earliest question answering systems
  - Developed to answer user 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 in cities within Germany
History

- Closed-domain systems

- Extracting answers from structured data (database)

- Converting natural language question to a database query
History

- Closed-domain systems

- Extracting answers from structured data (database)

- Converting natural language question to a database query
Open-domain QA

Closed-domain QA $\Rightarrow$ Open-domain QA

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

Closed-domain QA $\Rightarrow$ Open-domain QA

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

- Covering many subjects
- Information constantly added and updated
- No manual work for building database
Open-domain QA

Closed-domain QA ⇒ Open-domain QA

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

- Covering many subjects
- Information constantly added and updated
- No manual work for building database

- Some times information is not up to date
- Some information is wrong
- More irrelevant information
Open-domain QA

Closed-domain QA ⇒ Open-domain QA

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

- Covering many subjects
- Information constantly added and updated
- No manual work for building database

- Some times information is not up to date
- Some information is wrong
- More irrelevant information

More complex systems are required
Open-domain QA

■ START [Katz, 1997]
  □ Utilized a knowledge-base to answer the user’s questions
  □ The knowledge-base was first created automatically from unstructured Internet data
  □ Then it was used to answer natural language questions
Open-domain QA

- **START [Katz, 1997]**
  - Utilized a knowledge-base to answer the user’s questions
  - The knowledge-base was first created automatically from unstructured Internet data
  - Then it was used to answer natural language questions
Open-domain QA

- START [Katz, 1997]
  - Utilized a knowledge-base to answer the user’s questions
  - The knowledge-base was first created automatically from unstructured Internet data
  - Then it was used to answer natural language questions

Saeedeh Momtazi | QA | 14 July 2011
Open-domain QA

- **START [Katz, 1997]**
  - Utilized a knowledge-base to answer the user’s questions
  - The knowledge-base was first created automatically from unstructured Internet data
  - Then it was used to answer natural language questions
IBM Watson

- Webpage: IBM Watson

- Video: IBM and the Jeopardy Challenge
  http://www.youtube.com/watch?v=FC3IryWr4c8&feature=relmfu
  http://www.youtube.com/watch?v=_1c7s7-3fXI

- Video: A Brief Overview of the DeepQA Project
  http://www.youtube.com/watch?v=3G2H3DZ8rNc
Who is Warren Moon’s Agent?

“Who is Warren Moon’s Agent?” → Question Analysis

Question Classification → Query Construction

Document Retrieval → Sentence Retrieval

Sentence Annotation → Answer Extraction

Answer Validation → “Leigh Steinberg”
“Who is Warren Moon's Agent?” → Question Analysis → Question Classification → Query Construction → Document Retrieval → Sentence Retrieval → Sentence Annotation → Answer Extraction → Answer Validation → “Leigh Steinberg”
“Who is Warren Moon's Agent?”

Question Analysis

Question Classification

Query Construction

Document Retrieval

Sentence Retrieval

Sentence Annotation

Answer Extraction

Answer Validation

“Leigh Steinberg”
“Who is Warren Moon’s Agent?”

Question Analysis

Question Classification

Query Construction

Document Retrieval

Sentence Retrieval

Sentence Annotation

Answer Extraction

Web

Answer Validation

“Leigh Steinberg”

Natural Language Processing

Information Retrieval

Information Extraction

Saeedeh Momtazi | QA | 14 July 2011
Question Analysis

- Target Extraction
  - Extracting the target of the question
  - Using question target at the query construction step
Question Analysis

- Target Extraction
  - Extracting the target of the question
  - Using question target at the query construction step

- Pattern Extraction
  - 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 at the answer extraction step
Question Analysis

- **Target Extraction**
  - Extracting the target of the question
  - Using question target at the query construction step

- **Pattern Extraction**
  - 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 at the answer extraction step

**Examples:**

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

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

**Answer Pattern:** “X was born in Y.”
Question Analysis

- Target Extraction
  - Extracting the target of the question
  - Using question target at the query construction step

- Pattern Extraction
  - 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 at the answer extraction step
Question Analysis

- Target Extraction
  - Extracting the target of the question
  - Using question target at the query construction step

- Pattern Extraction
  - 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 at the answer extraction step

- Parsing
  - Using a dependency parser to extract the syntactic relations between question terms
  - Using the dependency relation path between question words to extract the correct answer at answer extraction step
Question Classification

- Classifying the input question into a set of question types
- Defining a map between question types and available named entity labels
- Using question type to extract strings that have the same type as the input question at the answer extraction step
Question Classification

- Classifying the input question into a set of question types
- Defining a map between question types and available named entity labels
- Using question type to extract strings that have the same type as the input question at the answer extraction step

Example:

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

Type: Country
Question Classification

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

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

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

6 coarse- and 50 fine-grained classes

**Types**
- HUMAN
- ABREVIATION
- DESCRIPTION
- LOCATION
- NUMERIC
- ENTITY

- mountain
- state
- country
- city
Question Classification

- Using available classifiers based on the favorite model
  - SVM: SVM-light
  - Maximum Entropy: Maxent, Yasmet
  - Naive Bayes
Query Construction

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

- **Task:**
  - Assigning a higher weight to the question target
  - Using query expansion techniques to expand the query
    - Thesaurus-based query expansion
    - Relevance feedback
    - Pseudo-relevance feedback
Document Retrieval

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

- Task:
  - Reducing the search space for the subsequent modules
  - 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
Document Retrieval

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

- Using available information retrieval toolkits

Lucene™

Lemur
Sentence Retrieval

■ Task:
  □ Finding a small segment of text that contains 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

- Language model-based sentence retrieval

Query $Q$

$$P(Q|S_2)$$
Sentence Retrieval

- Language model-based sentence retrieval

Query likelihood model:

\[ P(Q|S) = \prod_{i=1}^{M} P(q_i|S) \]
Sentence Retrieval

- Challenge:
The term mismatch problem in sentence retrieval is more critical than document retrieval
Sentence Retrieval

- Challenge:
The term mismatch problem in sentence retrieval is more critical than document retrieval

- Approaches:
Sentence Retrieval

- **Challenge:**
  The term mismatch problem in sentence retrieval is more critical than document retrieval

- **Approaches:**
  - Query expansion, (Pseudo-)Relevance Feedback
Sentence Retrieval

- Challenge:
The term mismatch problem in sentence retrieval is more critical than document retrieval

- Approaches:
  - Query expansion, (Pseudo-)Relevance Feedback
    - Does not work at sentence-level retrieval
Sentence Retrieval

- Challenge:
  The term mismatch problem in sentence retrieval is more critical than document retrieval

- Approaches:
  - Query expansion, (Pseudo-)Relevance Feedback
  - Term relationship models
    - WordNet
    - Term clustering model
    - Translation model
    - Triggering model

**Does not work at sentence-level retrieval**
Sentence Annotation

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

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

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

Example (NER):

Sentence1: “Albert Einstein was born on 14 March 1879.”

Sentence2: “Albert Einstein was born in Germany.”

Sentence3: “Albert Einstein was born in a Jewish family.”
Sentence Annotation

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

Example (NER):

Sentence1: “Albert Einstein was born on 14 March 1879.”
  - Person Name
  - Date

Sentence2: “Albert Einstein was born in Germany.”
  - Person Name
  - Country

Sentence3: “Albert Einstein was born in a Jewish family.”
  - Person Name
  - Religion
Answer Extraction

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency pars of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation
Answer Extraction

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency pars of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

Example:

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

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency pars of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

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

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency parses of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

Example (Pattern):

Sentence1: “Albert Einstein was born on 14 March 1879.”

Sentence2: “Albert Einstein was born in Germany.”

Sentence3: “Albert Einstein was born in a Jewish family.”
Extracting candidate answers based on various informations:

- The extracted patterns from question analysis
- The dependency parses of question from the question analysis
- The question type from question classification
- All annotated data from sentence annotation

Example (Pattern):

Sentence2: “Albert Einstein was born in Germany.”

Sentence3: “Albert Einstein was born in a Jewish family.”
Answer Extraction

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency pars of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

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

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency parse of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

Example:

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

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

Question Type: LOCATION - COUNTRY
Answer Extraction

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency parses of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

Example (Pattern):

Sentence2: “Albert Einstein was born in Germany.”

Sentence3: “Albert Einstein was born in a Jewish family.”
Answer Extraction

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency parses of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

Example (NER):

Sentence2: “Albert Einstein was born in Germany.”
- Person Name
- Country

Sentence3: “Albert Einstein was born in a Jewish family.”
- Person Name
- Religion
Answer Extraction

- Extracting candidate answers based on various informations:
  - The extracted patterns from question analysis
  - The dependency parse of question from the question analysis
  - The question type from question classification
  - All annotated data from sentence annotation

Example (NER):

Sentence 2: “Albert Einstein was born in Germany.”

Sentence 3: “Albert Einstein was born in a Jewish family.”
Answer Validation

- Using the Web as a knowledge resource
- Sending question keywords and answer candidates to a search engine
- Finding the frequency of the answer candidate within the Web data
- Selecting the most likely answers based on the frequencies
Answer Validation

- Query model:
  - 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 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
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
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”
“Who is Warren Moon’s Agent?”

Corpus

Web

Answer Validation → “Leigh Steinberg”

Sentence Annotation

Sentence Retrieval

Document Retrieval

Query Construction

Question Classification

Question Analysis

Natural Language Processing

Information Retrieval

Information Extraction

Saeedeh Momtazi | QA | 14 July 2011
Fact vs. Opinion

- Factual questions

“Who is Warren Moon’s Agent?”

“In what country was Albert Einstein born?”

“When was Mozart 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?”

  “How do students feel about Microsoft products?”
Outline

1. Introduction

2. History

3. QA Architecture
   - Factoid QA
   - Opinion QA

4. Summary
Architecture

“What do people like about Wikipedia?”

→ Question Analysis
   ↓ Question Classification
   ↓ Query Construction
   ↓ Document Retrieval
   ↓ Sentence Retrieval
   ↓ Sentence Annotation
   ↓ Answer Extraction
   ↓ Answer Validation

Corpus

Web
“What do people like about Wikipedia?” → Question Analysis

↓

Question Classification

↓

Query Construction

↓

Document Retrieval

Corpus

Sentence Annotation

↓

Answer Extraction

↓

Answer Validation

Web
Architecture

“What do people like about Wikipedia?”

Question Analysis

Question Classification

Query Construction

Document Retrieval

Sentence Retrieval

Corpus

Sentence Annotation

Answer Extraction

Answer Validation

Web

Saeedeh Momtazi | QA | 14 July 2011
“What do people like about Wikipedia?”

Question Analysis → Question Classification

Query Construction

Corpus → Document Retrieval → Sentence Retrieval

Sentence Annotation → S Opinion Classification

Answer Extraction → S Polarity Classification

Web → Answer Validation
Architecture

“What do people like about Wikipedia?” → Question Analysis

↓

Query Construction

↓

Corpus → Document Retrieval → Sentence Retrieval

↓

Sentence Classification

↓

Sentence Annotation → S Opinion Classification

↓

Answer Extraction

↓

Answer Validation

↓

Web
“What do people like about Wikipedia?”

Question Analysis

Q Semantic Classification  Q Polarity Classification

Query Construction

Document Retrieval  Sentence Retrieval

Sentence Annotation  S Opinion Classification  S Polarity Classification

Answer Extraction  Answer Validation

Corpus

Web
“What do people like about Wikipedia?”

- **Question Analysis**
  - Q Semantic Classification
  - Q Polarity Classification

- **Query Construction**

- **Corpus**
  - Document Retrieval
  - Sentence Retrieval
    - S Opinion Classification
    - S Polarity Classification

- **Sentence Annotation**

- **Answer Extraction**

- **Web**
  - Answer Validation

**Natural Language Processing**

**Information Retrieval**

**Information Extraction**

---

Saeedeh Momtazi | QA | 14 July 2011
“What do people like about Wikipedia?”

Architecture

- Question Analysis
- Q Semantic Classification
- Q Polarity Classification
- Query Construction
- Document Retrieval
- Sentence Retrieval
- Sentence Annotation
- S Opinion Classification
- S Polarity Classification
- Answer Extraction
- Answer Validation

Natural Language Processing
Information Retrieval
Information Extraction
Opinion Mining
Q Polarity Classification

- Running in parallel with question classification

- Task:
  - Decide whether the input question has a positive or negative polarity
Q Polarity Classification

- Running in parallel with question classification

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

examples:

- “What do people like about Wikipedia?”
- “Why people hate reading Wikipedia articles?”
Q Polarity Classification

- Running in parallel with question classification

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

- Approaches:
  - Using a rule-based model based on subjectivity lexicon
  - Running a classifier trained on an annotated corpus
    - Training on overall vocabulary of the dataset
    - Training on all polarity expressions from the subjectivity lexicon
S Opinion Classification

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

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

<table>
<thead>
<tr>
<th>examples:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question: “What do people like about Wikipedia?”</td>
</tr>
</tbody>
</table>
S Opinion Classification

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

examples:

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.”


S4: “Wikipedia is a great way to access lots of information.”
S 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.”

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
S Polarity Classification

- **Task:**
  - Distinguishing between 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

1 Introduction

2 History

3 QA Architecture
   Factoid QA
   Opinion QA

4 Summary
Next Semester

Master Seminar on Question Answering Systems
Master Seminar on Question Answering Systems

- Question Analysis
- Q Semantic Classification
- Q Polarity Classification
- Query Construction
- Document Retrieval
- Sentence Retrieval
- S Opinion Classification
- S Polarity Classification
- Sentence Annotation
- Answer Extraction
- Answer Validation