Natural Language Processing

Question Answering

Potsdam, 21 June 2012

Saeedeh Momtazi
Information Systems Group
Outline

1 Introduction

2 History

3 QA Architecture
Outline

1 Introduction

2 History

3 QA Architecture
Motivation

- Finding small segments of text which answer users’ questions
Motivation

Who is Warren Moon's agent?

Booking Warren Moon Appearances, Contact Warren Moon Agent...
Call 1-888-248-7141 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.teesworld.com/sports/.../warren-moon.php - Cached - Similar -

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...
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, ...
panda.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, 1990 - Warren Moon's agent went on the offensive after another day of tense 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 to agree to a deal with the Kansas City Chiefs...
Motivation

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
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

1 Introduction

2 History

3 QA Architecture
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)

- Converting natural language questions to database queries

Labor intensive to build

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

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

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
IBM Watson

- Playing against two greatest champions of Jeopardy

- Challenges
  - Knowledge
  - Speed
  - Confidence

Building Watson: An Overview of the DeepQA Project


AI Magazine, 2010
Outline

1. Introduction
2. History
3. QA Architecture
“Who is Warren Moon’s Agent?”

Question Analysis

Question Classification

Query Construction

Document Retrieval

Sentence Retrieval

Sentence Annotation

Answer Extraction

Answer Validation

“Leigh Steinberg”
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:

“John grills a fish on an open fire.”

Cook Food Heating-Instrument
Q Analysis: Semantic Role Labeling

- Frame assignment
- Role labeling

Example:

“Jim flew his plane to Texas.”
- Jim: Driver
- flew: OPERATE
- his plane: Vehicle
- to Texas: Goal

Example:

“Alice destroys the item with a plane.”
- Alice: Destroyer
- destroys: DESTROYING
- the item: Undergoer
- with a plane: Instrument
Q Analysis: Semantic Role Labeling

- Finding the question’s head verb

Example:

“Who purchased YouTube?”

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.”
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: “Albert Einstein was born in 14 March 1879.”
- Person
- Date

S2: “Albert Einstein was born in Germany.”
- Person
- Country

S3: “Albert Einstein was born in a Jewish family.”
- Person
- Religion
Question Classification

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

6 coarse- and 50 fine-grained classes

Types

- HUMAN
- ABBREVIATION
- DESCRIPTION
- LOCATION
- NUMERIC
- ENTITY
Question Classification

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

6 coarse- and 50 fine-grained classes

Types
- HUMAN
- ABBREVIATION
- DESCRIPTION
- LOCATION
- NUMERIC
- ENTITY
  - City
  - Country
  - State
  - Mountain
  - Other

Saeedeh Momtazi | NLP | 21.06.2012
## Question Classification

<table>
<thead>
<tr>
<th>Question</th>
<th>Type</th>
<th>Sub-type</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Who killed Gandhi?”</td>
<td>HUMAN</td>
<td>Individual</td>
</tr>
<tr>
<td>“Who has won the most Super Bowls?”</td>
<td>HUMAN</td>
<td>Group</td>
</tr>
<tr>
<td>“What city did Duke Ellington live in?”</td>
<td>LOCATION</td>
<td>City</td>
</tr>
<tr>
<td>“Where is the highest point in Japan?”</td>
<td>LOCATION</td>
<td>Mountain</td>
</tr>
<tr>
<td>“What do sailors use to measure time?”</td>
<td>ENTITY</td>
<td>Technique</td>
</tr>
<tr>
<td>“Who is Desmond Tutu?”</td>
<td>DESCRIPTION</td>
<td>human</td>
</tr>
</tbody>
</table>
Question Classification

Coarse Classifier

{ABREVIATION} {DESCRIPTION} {LOCATION} {NUMERIC} {ENTITY}

Fine Classifier

{HUMAN} {STATE} {MOUNTAIN} {CITY} {COUNTRY}

Question

{OTHER}
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

![Graph showing the number of candidates against confidence](image-url)
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
Comparing language modeling with traditional methods
Sentence Retrieval

- Comparison of the effects of text length on information retrieval

<table>
<thead>
<tr>
<th></th>
<th>MAP</th>
<th>P @ 10</th>
<th>P @ 20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>0.191</td>
<td>0.232</td>
<td>0.200</td>
</tr>
<tr>
<td>750 bytes</td>
<td>0.064</td>
<td>0.186</td>
<td>0.149</td>
</tr>
<tr>
<td>500 bytes</td>
<td>0.055</td>
<td>0.166</td>
<td>0.142</td>
</tr>
<tr>
<td>250 bytes</td>
<td>0.036</td>
<td>0.136</td>
<td>0.117</td>
</tr>
<tr>
<td>Sentences</td>
<td>0.030</td>
<td>0.098</td>
<td>0.081</td>
</tr>
</tbody>
</table>

based on work by Vanessa Murdock

- Main problem of sentence retrieval: sentence brevity
Sentence Retrieval

- Approaches to overcome the sentence brevity problem:
  - 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

\[ \text{Query } Q \]

\[ \theta_{S_2} \]

\[ P(Q|S_2) \]

\[ Q \]

\[ P(Q|S_2) \]

\[ W_1, \ldots, W_j, \ldots \]
Sentence Retrieval

- Translation Model

\[ Q \]

\[ q_1: t_1, ..., t_k \]

\[ ..., q_i: t_1, ..., t_k \]

\[ w_1, ..., w_j, ... \]

\[ \theta_{S_2} \]

\[ P(Q|S_2) \]

Saeedeh Momtazi | NLP | 21.06.2012
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

Query $Q$

$P(Q|S_2)$

$\theta_{S_2}$

$w_1, \ldots, w_j, \ldots$
Sentence Retrieval

- Class Model

Query $Q$

$P(Q|S_2)$

$\theta_{S_2}$

$C_{w_1}, \ldots, C_{w_j}, \ldots$

$w_1 \rightarrow C_{w_1}$

$w_i \rightarrow C_{w_i}$

$w_n \rightarrow C_{w_n}$

Saeedeh Momtazi | NLP | 21.06.2012
Annotating relevant sentences using linguistic analyses

- Named entity recognition
- Syntactic parsing
- Semantic role labeling

Example:

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

- Annotating relevant sentences using linguistic analyses
  - Named entity recognition
  - Syntactic parsing
  - Semantic role labeling

Example (NER):

Sentence 1: “*Albert Einstein* was born in *14 March 1879*.”

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

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

- Extracting candidate answers based on various information

  - Question
    - Question Analysis: patterns
    - Question Analysis: syntactic parse
    - Question Analysis: semantic roles
    - Question Classification: question type

  - Sentence
    - Sentence Annotation: all annotated data
Answer Extraction

Question Processing

Information Retrieval

Question

Features

Is ac type correct?
expected answer type matching

Is ac supported?
ac context matching

Answer Candidate Generation
(ac: all basic noun phrases)

Answer Candidate Ranking
(Features integration)

Sent_0   ac_00, ac_01, ac_02, ac_03, ...
Sent_1   ac_10, ac_11, ac_12, ac_13, ...
Sent_2   ac_20, ac_21, ac_22, ac_23, ...
...
...

Answer
Answer Extraction

- Using extracted patterns

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

- Using extracted patterns

Example (Pattern):

Sentence 1: "Albert Einstein was born in 14 March 1879."

Sentence 2: "Albert Einstein was born in Germany."

Sentence 3: "Albert Einstein was born in a Jewish family."
Answer Extraction

- Using question type and entity type

Example:

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

Question Type: LOCATION - Country
Answer Extraction

- Using question type and entity type

Example (NER):

Sentence 1: “Albert Einstein was born in 14 March 1879.”
- Person Name
- Date

Sentence 2: “Albert Einstein was born in Germany.”
- Person Name
- Country

Sentence 3: “Albert Einstein was born in a Jewish family.”
- Person Name
- Religion
Answer Extraction

- Using syntactic parsing
  - Different wordings possible, but similar syntactic structure

Q: **Who founded the Black Panthers organization?**

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

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

based on work by Dan Shen
Answer Extraction

- Using syntactic parsing
  - Many syntactic variations → need robust matching approach

Q: **Who founded the Black Panthers organization?**

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

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

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

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

- Using semantic roles

Example:

“Who purchased YouTube?”

Example:

“In 2006, YouTube was purchased by Google for $1.65 billion.”
Answer Extraction

- Comparing answer extraction features
Answer Validation

- Using Web as a knowledge resource for validating answers

- Required steps
  - Query creation
  - Answer rating
Answer Validation

- Query creation
  - 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

- Answer rating
  - 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
“Who is Warren Moon’s Agent?”

Question Analysis → Question Classification → Query Construction

Corpus

Document Retrieval → Sentence Retrieval → Sentence Annotation

Web

Answer Extraction → Answer Validation → “Leigh Steinberg”

Natural Language Processing

Information Retrieval

Information Extraction