NLP
Question Answering

*System Architecture*
System Architecture

• Many questions can be answered by traditional search engines
What is the Largest City in Northern Afghanistan?

... Afghanistan, Kabul, 2,450 ... Administrative capital and largest city (1997 est ... Undetermined. Panam...
System Components

- **Source identification**
  - semi-structured vs. text sources

- **Query modulation**
  - best paraphrase of a NL question given the syntax of a search engine
  - Example: Who wrote Hamlet → author | wrote Hamlet

- **Document retrieval**

- **Sentence ranking**
  - n-gram matching, Okapi

- **Answer extraction**
  - question type classification
  - phrase chunking

- **Answer ranking**
  - question type, proximity to query words, frequency
What is the largest city in Northern Afghanistan?

Query modulation

(largest OR biggest) city “Northern Afghanistan”

Document retrieval

www.infoplease.com/cgi-bin/id/A0855603
www.washingtonpost.com/wp-dyn/print/world/

Sentence ranking

Gudermes, Chechnya's second largest town ... location in Afghanistan's outlying regions within three miles of the airport at Mazar-e-Sharif, the largest city in northern Afghanistan

Answer extraction

Gudermes
Mazar-e-Sharif

Answer ranking

Mazar-e-Sharif
Gudermes
Question Type Classification

• Can help find the right answers in the text
• Example
  – Who wrote Anna Karenina?
  – Looking for a PERSON/INDIVIDUAL/WRITER
### SYN-classes (IBM AnSel)

<table>
<thead>
<tr>
<th>QA-token</th>
<th>Question type</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLACE</td>
<td>Where?</td>
<td>In the Rocky Mountains</td>
</tr>
<tr>
<td>COUNTRY</td>
<td>Where? What country?</td>
<td>United Kingdom</td>
</tr>
<tr>
<td>STATE</td>
<td>Where? What state?</td>
<td>Massachusetts</td>
</tr>
<tr>
<td>PERSON</td>
<td>Who?</td>
<td>Albert Einstein</td>
</tr>
<tr>
<td>ROLE</td>
<td>Who?</td>
<td>Doctor</td>
</tr>
<tr>
<td>NAME</td>
<td>Who? What? Which?</td>
<td>The Shakespeare Festival</td>
</tr>
<tr>
<td>ORG</td>
<td>Who? What?</td>
<td>The U.S. Post Office</td>
</tr>
<tr>
<td>DURATION</td>
<td>How long?</td>
<td>For 5 centuries</td>
</tr>
<tr>
<td>AGE</td>
<td>How old?</td>
<td>30 years old</td>
</tr>
<tr>
<td>YEAR</td>
<td>When? What year?</td>
<td>1999</td>
</tr>
<tr>
<td>TIME</td>
<td>When? What time?</td>
<td>In the afternoon</td>
</tr>
<tr>
<td>DATE</td>
<td>When? What date?</td>
<td>July 4th, 1776</td>
</tr>
<tr>
<td>VOLUME</td>
<td>How big? How large?</td>
<td>3 gallons</td>
</tr>
<tr>
<td>AREA</td>
<td>How big? How large?</td>
<td>4 square inches</td>
</tr>
<tr>
<td>LENGTH</td>
<td>How long? How big?</td>
<td>3 miles</td>
</tr>
<tr>
<td>WEIGHT</td>
<td>How heavy? How big?</td>
<td>25 tons</td>
</tr>
<tr>
<td>NUMBER</td>
<td>How many?</td>
<td>1,134.5</td>
</tr>
<tr>
<td>RATE</td>
<td>How much? What percentage?</td>
<td>50 per cent</td>
</tr>
<tr>
<td>MONEY</td>
<td>How much?</td>
<td>4 million dollars</td>
</tr>
</tbody>
</table>
UIUC Question Types

- **ENTITY**: entities
  - animal: animals
  - body: organs of body
  - color: colors
  - creative: inventions, books and other creative pieces
  - currency: currency names
  - dis.med.: diseases and medicine
  - event: events
  - food: food
  - instrument: musical instrument
  - lang: languages
  - letter: letters like a–z
  - other: other entities
  - plant: plants
  - product: products
  - religion: religions
  - sport: sports
  - substance: elements and substances
  - symbol: symbols and signs
  - technique: techniques and methods
  - term: equivalent terms
  - vehicle: vehicles
  - word: words with a special property

- **ABBREVIATION**: abbreviation
  - abb: abbreviation
  - exp: expression abbreviated

- **DESCRIPTION**: description and abstract concepts
  - definition: definition of sth.
  - description: description of sth.
  - manner: manner of an action
  - reason: reasons

- **HUMAN**: human beings
  - group: a group or organization of persons
  - ind: an individual
  - title: title of a person
  - description: description of a person

- **LOCATION**: locations
  - city: cities
  - country: countries
  - mountain: mountains
  - other: other locations
  - state: states

- **NUMERIC**: numeric values
  - code: postcodes or other codes
  - count: number of sth.
  - date: dates
  - distance: linear measures
  - money: prices
  - order: ranks
  - other: other numbers
  - period: the lasting time of sth.
  - percent: fractions
  - speed: speed
  - temp: temperature
  - size: size, area and volume
  - weight: weight
UIUC Examples

- **NUM:date**
  - When did Rococo painting and architecture flourish?

- **LOC:country**
  - What country’s national passenger rail system is called Via?

- **HUM:ind**
  - Who invented Make-up?

- **DESC:desc**
  - What is the origin of the word ‘`attic`’?

- **DESC:desc**
  - What did Delilah do to Samson’s hair?

- **ENTITY:animal**
  - What kind of animals were in the Paleozoic era?

- **HUM:ind**
  - Which of the following was Rhodes Scholar?

- **HUM:ind**
  - Who comprised the now-defunct comic book team known as the Champions?

- **DESC:manner**
  - How do you make a paintball?

- **LOC:state**
  - What U.S. state is Fort Knox in?

- **ENTITY:animal**
  - What is a female rabbit called?

- **LOC:mount**
  - Where is the highest point in Japan?

- **DESC:desc**
  - Where do chihuahuas come from?

- **LOC:other**
  - Where does Barney Rubble go to work after he drops Fred off in the “Flintstones” cartoon series?
UIUC papers on question classification

- Xin Li and Dan Roth. Experimental Data for Question Classification

- Xin Li, Dan Roth. Learning Question Classifiers: The Role of Semantic Information
  - [http://l2r.cs.uiuc.edu/~danr/Papers/LiRo05a.pdf](http://l2r.cs.uiuc.edu/~danr/Papers/LiRo05a.pdf)
  - [http://cogcomp.cs.illinois.edu/page/software_view/LBJava](http://cogcomp.cs.illinois.edu/page/software_view/LBJava)

- Data set
  - training ([http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/QC/train_5500.label](http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/QC/train_5500.label))
Techniques for Question Classification

• Classification task
  – Use standard techniques

• Regular expressions
  – WHO is|was -> PERSON
  – WHICH STATE -> STATE
Query Reformulation

<table>
<thead>
<tr>
<th>No.</th>
<th>Operator</th>
<th>Paraphrase</th>
<th>TRDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>IDENTITY</td>
<td>What country is the &quot;biggest producer&quot; of tungsten</td>
<td>0.837</td>
</tr>
<tr>
<td>1</td>
<td>DEL.WH</td>
<td>country is the &quot;biggest producer&quot; of tungsten</td>
<td>1.755</td>
</tr>
<tr>
<td>2</td>
<td>DEL.AUX</td>
<td>What country the &quot;biggest producer&quot; of tungsten</td>
<td>1.322</td>
</tr>
<tr>
<td>3</td>
<td>DEL.ART</td>
<td>What country is &quot;biggest producer&quot; of tungsten</td>
<td>1.436</td>
</tr>
<tr>
<td>4</td>
<td>DEL.PREP</td>
<td>What country is the &quot;biggest producer&quot; of tungsten</td>
<td>1.436</td>
</tr>
<tr>
<td>5</td>
<td>DEL.STOP</td>
<td>What country is the &quot;biggest producer&quot; of tungsten</td>
<td>0.837</td>
</tr>
<tr>
<td>6</td>
<td>REPL.1N</td>
<td>What (&quot;administrative district&quot; OR &quot;political unit&quot; OR people OR region OR &quot;geographical area&quot; OR &quot;rural area&quot;) is the &quot;biggest producer&quot; of tungsten</td>
<td>1.181</td>
</tr>
<tr>
<td>7</td>
<td>REPL.2N</td>
<td>What country is the &quot;biggest producer&quot; of (&quot;metallic element&quot; OR wolfram OR &quot;atomic number 74&quot;)</td>
<td>1.419</td>
</tr>
</tbody>
</table>

[Radev et al. 2001]
Passage Retrieval

• **Features**
  – Proper nouns that match the query
  – Near each other
  – Entities that match the expected answer type
Answer Retrieval

• Use NER to identify the matching phrases
  – E.g., “January 1, 1951” as a DATE

• Features
  – Distance to query words
  – Answer type
  – Wordnet similarity
  – Redundancy
Redundancy (1/2)

What is the capital of Spain?
- Madrid is the capital of Spain
- En route to Spain's capital of Madrid
- Madrid, Spain's capital city is situated almost at the geographical epicentre of the country
- The capital of Spain is Madrid
- Madrid, Spain's sunny capital
- Madrid became Spain's capital
- In 1561, it was elevated to status as Spain's capital city
- Madrid has been the capital of Spain since 1562
- Madrid, the physical and cultural capital of Spain
Redundancy (2/2)

• When did French revolutionaries storm the Bastille?
  – The storming of the Bastille occurred in Paris on the morning of 14 July 1789
  – The storming of the Bastille, 14 July 1789
  – The storming of the Bastille prison on July 14th 1789 was an event that paved the way to further civil disorder and upheaval in France.
  – French revolutionaries storm Bastille, 1789
  – The storming of the Bastille (Louis XVI's prison) by French revolutionaries took place on July 14th, 1789

• Who killed Mahathma Gandhi?
  – Mohandas Karamchand Gandhi (often called Mahatma Gandhi) was assassinated on 30 January 1948, shot at point-blank range by Nathuram Godse.
  – Nathuram Godse killed Gandhi
  – Godse killed Gandhi
  – Mahatma Gandhi was assassinated on 30 January 1948, shot at point-blank range by Nathuram Godse
  – Mohandas Gandhi was shot dead by Nathuram Godse, a Hindu extremist
  – Godse assassinated Mahatma Gandhi on January 30, 1948, approaching him during the evening prayer, bowing, and shooting him three times at close range with a Beretta semi-automatic pistol.
NLP