Assessment of historical figures using BabelNet

Szymon Roziewski, Maciej Kowalski

National Information Processing Institute, Natural Language Processing Laboratory, Warsaw
Motivation

- Building a simple Sentiment Analysis System.
- Providing sentiment analysis of famous people based on BabelNet synsets. Actually, they could be fictional.
- To have some fun :)
Motivation

Why assessment of historical figures?

- To get a fast sentiment feedback about people playing a role in History.
- Resulting sentiment might be useful for another purposes: search engines, marketing companies etc.

"Always walk through life as if you have something new to learn and you will."

- Vernon Howard
Method

- We used BabelNet for the core of the system.
- BabelNet provides synset tree for a given entity.
- Each tree leaf carries short gloss of its own definition.
- We collect closest synsets to build a small corpus of text surrounding given term.
- We perform sentiment analysis based on sentiment dictionaries:
  - lexicon dictionary
  - SentiWordNet
BabelNet

- Huge multilingual encyclopedic dictionary and ontology
- Offers a large network of concepts and named entities with multilingual lexical coverage
Sentiment Dictionaries

- Lexicon Dictionary – lexicon built at University of Illinois Chicago
  It is divided into positive and negative words
- SentiWordNet (Princeton) – a list of words described by two real numbers: the positive and negative impact

Basic forms of words obtained by using morfologik library.
### BabelNet query for Julius Caesar

#### Noun

1. **Julius Caesar, Gaius Julius Caesar, Caesar**
   - Conqueror of Gaul and master of Italy (100-44 BC)
   - ID: [00014550](https://babelnet.org/)| Named Entity

2. **Julius Caesar (play)**
   - The Tragedy of Julius Caesar is a tragedy by William Shakespeare, believed to have been written in 1599.
   - ID: [00343295](https://babelnet.org/)| Named Entity

3. **Julius Caesar (1953 film)**
   - Julius Caesar is a 1953 Metro-Goldwyn-Mayer film adaptation of the play by Shakespeare, directed by Joseph L. Mankiewicz, who also wrote the uncredited screenplay, and produced by John Houseman.
   - ID: [01000470](https://babelnet.org/)| Named Entity

4. **Julius Caesar (TV miniseries)**
   - Julius Caesar is a 2002 mini-series about the life of Julius Caesar.
   - ID: [02021562](https://babelnet.org/)| Named Entity

5. **Julia Caesaris (sister of Julius Caesar), Julius Caesar (sisters of)**
   - Julia is the name of two daughters of praetor Gaius Julius Caesar and Aurelia Cotta, the parents of dictator Gaius Julius Caesar.
   - ID: [03177744](https://babelnet.org/)| Named Entity

6. **Julius Caesar (judge)**
   - Sir Julius Caesar was an English judge and politician who sat in the House of Commons at various times between 1589 and 1622.
   - ID: [02140630](https://babelnet.org/)| Named Entity
BabelNet query for Julius Caesar
BabelNet query for Julius Caesar
Sentiment Analysis of Julius Caesar

Positive :)
Implementation

- The process has been automated by using Java along with BabelNet API, for querying the semantic network.
- Parameters used for the synset selection:
  - language = EN
  - BabelPOS = NOUN
  - BabelSenseSource = WIKIDATA
### Sentiment Analysis Results

<table>
<thead>
<tr>
<th>Historical Figure</th>
<th>Positive impact</th>
<th>Negative impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albert Einstein</td>
<td>0.523</td>
<td>0.477</td>
</tr>
<tr>
<td>Isaac Newton</td>
<td>0.7</td>
<td>0.3</td>
</tr>
<tr>
<td>Al Capone</td>
<td>0.304</td>
<td>0.696</td>
</tr>
<tr>
<td>Marylin Monroe</td>
<td>0.602</td>
<td>0.398</td>
</tr>
<tr>
<td>Billy the Kid</td>
<td>0.601</td>
<td>0.399</td>
</tr>
<tr>
<td>Julius Caesar</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>Cinderella</td>
<td>0.4776</td>
<td>0.5224</td>
</tr>
</tbody>
</table>
Future Work

- There were no gold-standard sets to compare with, we assume such method can be developed further.
- Implementing disambiguating mechanism in order to select the exact synset carrying glosses related to historical figure in question.
Thank you!

Ośrodek Przetwarzania Informacji
Państwowy Instytut Badawczy

al. Niepodległości 188 b
00-608 Warszawa

tel.: +48 22 570 14 00
e-mail: opi@opi.org.pl