Graph-Based Entity Linking

A. Naderi, J. Turmo, H. Rodríguez

TALP Research Center
Universitat Politecnica de Catalunya

February 20, 2014
Table of contents

1 Introduction
   - Entity Linking
   - EL applications
   - EL problem in real-life entities
   - Our proposal

2 Our Approach
   - Candidate Generation
   - Graph Generation
   - Graph-based Ranking

3 Evaluation
   - Framework
Introduction
Importance of World Knowledge

- Recently, the needs of world knowledge for Artificial Intelligence applications are highly increasing.
Importance of World Knowledge

- Recently, the needs of world knowledge for Artificial Intelligence applications are highly increasing.

- Knowledge Bases (KB) are involved to keep and categorize entities and their relations as a part of world knowledge.
Way to prevent high cost of manual elicitation

but the high cost of manual elicitation to create KBs forces toward automatic acquisition from text.
Way to prevent high cost of manual elicitation

- but the high cost of manual elicitation to create KBs forces toward automatic acquisition from text.

- This requires two main abilities:
Way to prevent high cost of manual elicitation

- but the high cost of manual elicitation to create KBs forces toward automatic acquisition from text.

- This requires two main abilities:
  1) extracting relevant information of mentioned entities from the unstructured text (Slot Filling), and
Way to prevent high cost of manual elicitation

- but the high cost of manual elicitation to create KBs forces toward automatic acquisition from text.

- This requires two main abilities:
  1) extracting relevant information of mentioned entities from the unstructured text (Slot Filling), and
  2) linking these entities with entries in the KB (Entity Linking).
Entity Linking (EL)

- Also known as record linkage, grounding, or name entity disambiguation.

France reject Irish replay hopes

The French Football Federation (FFF) has dismissed the Republic of Ireland’s request for a rematch of their controversial World Cup play-off game.

The Republic pleaded for the FFF to offer a rematch after Fifa ruled that the result would stand.

France striker Thierry Henry, who handled the ball in the build-up to the winning goal, had earlier said a replay would be the fairest solution.
EL applications

- **Automatic link generation** for entity references in news articles.
EL applications

- **Automatic link generation** for entity references in news articles.
- **E-mail processing** to process messages and identify references to people in the contact list.
EL applications

- **Automatic link generation** for entity references in news articles.
- **E-mail processing** to process messages and identify references to people in the contact list.
- **To monitor events in the companies** to monitor events like merges of companies or new product releases.
EL applications

- **Automatic link generation** for entity references in news articles.

- **E-mail processing** to process messages and identify references to people in the contact list.

- **To monitor events in the companies** to monitor events like merges of companies or new product releases.

- **In automating corporate customer care** to process inquiries or complaints using the information provided by the customers.
EL Problems

George H. W. Bush (41st U.S. president)

George W. Bush (43rd U.S. president)

George Bush (bibalical scholar)

George Washington Bush (the first black settler in U.S. state of Washington)

George P. Bush (U.S. Navy Reserve officer)
Our Proposal for EL

We take advantage of a graph-based structure for EL. The advantages are:

- It is used to present facts in visual form.
- To be used to make facts clearer and more understandable.
- Is a convincing structure that can show and compare relationships and changes.
- Is a compact way to convey information.
- Is forceful that emphasizes main points.

Some EL research are presented using graph-based approaches but in comparison to other approaches it is not enough.
Section 2

Our Graph-Based Approach for EL
An example Mono-Lingual EL query is

```xml
<query id="EL000304">
  <name>Barnhill</name>
  <docid>eng-NG-31-100578-11879229</docid>
  <startoffset>xxx</startoffset>
  <endoffset>yyy</endoffset>
</query>
```

The answer is:

- the ID of the KB entry to which the name refers, or
- a “NILxxxx” ID
Our Approach

1. Query
2. Candidate Generation
3. Graph Generation
4. Graph-based Ranking
5. Answer
6. Ref. KB
A graph for query name Picasso
Graph Generation

Each edge in this graph has two characteristics:

1. Attributes (e.g., Place_of_Birth, Age)
2. An assigned weight

In the case of weights, we manually assign different weights for the edges considering the reliability of information that are extracted for each candidate.
Graph Generation

- Then, the system tries to find common properties between candidates and query.
- And, a path is generated between the candidate and query through these common properties.
Graph-Based Ranking

- $C = (c_1, c_2, \ldots, c_n)$ is the set of candidate nodes.
- $q$ is the query node in the graph $G$.
- $P_{c_k} = (P^1_{c_k}, P^2_{c_k}, \ldots, P^m_{c_k})$ is the set of paths between $q$ and $c_k$.
- $P^i_{c_k}$ is represented by the sequence of weights corresponding to the edges in the path, $P^i_{c_k} = \langle w_1, w_2, \ldots, w_r \rangle$.
- $s_{c_k}$ is the score of the candidate node $c_k$, then:
Graph-Based Ranking

\[ s_{c_k} = \begin{cases} \sum_{P^i_{c_k} \in P_{c_k}} \sum_{w_j \in P^i_{c_k}} w_j & \text{if } P_{c_k} \neq \emptyset \\ 0 & \text{if } P_{c_k} = \emptyset \end{cases} \]
Graph-Based Ranking

A sample view of our graph structure

\[
\begin{align*}
sc_1 &= sq \cdot (w_q^2 + wc_1) \\
sc_2 &= sq \cdot (w_q^2 + wc_1) + sq \cdot (wc_2^1 + w_q^2) \\
sc_3 &= 0,
\end{align*}
\]

Initial score for the candidates is zero and for the query is 1.
Graph-Based Ranking

Assuming:
- $m_q$ as the query name,
- $S = \{s_{c_k}\}$,

the link between $m_q$ and KB is obtained as follows:

$$
\text{link}(m_q, KB) = \begin{cases} 
  c & \text{if } \exists c \in C : s_c = \max(S) \geq \beta \\
  \text{NIL} & \text{otherwise}
\end{cases}
$$
Evaluation
For testing our Baseline EL system, we have participated in the TAC-KBP English Mono-Lingual Entity Linking (MLEL) evaluation in the years 2012.

Thank You