Graph-Based Entity Linking

A. Naderi, J. Turmo, H. Rodríguez

TALP Research Center Universitat Politecnica de Catalunya

February 20, 2014



Table of contents

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



Introduction	
00000	

Importance of World Knowledge

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





Introduction	
00000	

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.





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



- but the high cost of manual elicitation to create KBs forces toward automatic acquisition from text.
- This requires two main abilities:



- 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



- 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).



Introduction	
000000	

Entity Linking (EL)

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





イロト イヨト イヨト

• Automatic link generation for entity references in news articles.



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



- 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.



- 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





16/29

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.

イロト イヨト イヨト



Our Graph-Based Approach for EL



EL Task Definition

An example Mono-Lingual EL query is

```
<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





Graph Generation: initial graph



A graph for query name Picasso



イロト イヨト イヨト

Graph Generation

Each edge in this graph has two characteristics:

- Attributes (e.g., Place_of_Birth, Age)
- 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.



Introd	uction

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.



Common properties between query and candidates



イロト イヨト イヨト

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



$$s_{c_k} = \begin{cases} \sum_{P_{c_k}^i \in P_{c_k}} \sum_{w_j \in P_{c_k}^i} w_j & \text{if } P_{c_k} \neq \emptyset \\ 0 & \text{if } P_{c_k} = \emptyset \end{cases}$$







A sample view of our graph structure

$$\begin{split} s_{c_1} &= s_q \cdot \left(w_q^2 + w_{c_1}^1 \right) \\ s_{c_2} &= s_q \cdot \left(w_q^2 + w_{c_2}^1 \right) + s_q \cdot \left(w_q^1 + w_{c_2}^2 \right) \\ s_{c_3} &= 0, \end{split}$$

Initial score for the candidates is zero and for the query is 1.



Assuming:

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

the link between m_q and KB is obtained as follows:

$$link(m_q, KB) = \begin{cases} c & \text{if } \exists c \in C : s_c = \max(S) \ge \beta \\ \text{NIL} & otherwise \end{cases}$$





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.
- E. González, H. Rodriguez, J. Turmo, P.R. Comas, A. Naderi, A. Ageno, E. Sapena, M. Vila, M.A. Marti. 2012. The TALP participation at TAC-KBP 2012. Text Analysis Conference, Gaithersburg, ML USA.



Introduction	

Thank You

