

REACTION at the Entity Linking task in KBP 2011

David S Batista
dsbatista@inesc-id.pt
INESC-ID

Silvio Moreira
samir@inesc-id.pt
INESC-ID

Mário J Silva
mjs@inesc-id.pt
INESC-ID

Francisco M Couto
fcouto@di.fc.ul.pt
LaSIGE

Paula Carvalho
pcc@inesc-id.pt
INESC-ID

Abstract

We adopted a simple and naive approach based solely on queries over an index of the Knowledge Base text. The queries consist of information gathered from the query string and the support document. This paper describes the prototype used to submit the runs. It is an early version of the prototype we have envisioned, which is, currently a work-in-progress.

1. Introduction

Entity Linking is the process of associating an entity mentioned in a text to an entry, representing that entity, in a knowledge base. It can be used to automatically augment text with links, and greatly improve a web user experience. For instance, when reading on-line news, if the main actors in the story are linked to a knowledge base, the user can simply mouse-over their names to obtain additional information. An information extraction system can make use of such links to infer facts or answer questions.

2. System

Due to the late start in the development of the system that we would create to participate in the Entity Linking task, it was simply based on Lucene to index the Knowledge Base text about entities and an analysis of the supporting document associated with each query.

We indexed the Knowledge Base with a version of Lucene¹ with support for BM25 extensions². Using as gold standard the data from the 2010 edition of Knowledge Base Population Task, we tuned the BM25 parameters k_1 to 0.1 and b to 1.0.

We used a named-entity recognition (NER) system built in-house, REMBRANDT [Cardoso(2008)], which uses Wikipedia as a raw knowledge resource and explores the Wikipedia document structure to classify all kinds of named entities in the text.

For each query string we extracted possible alternative names based on Wikipedia redirection pages. Then the support document associated to the query was analyzed by REMBRANDT

to gather the following named-entities: persons, organizations, places. We also extracted a window of ten tokens surrounding the mention of the query string in the support document. The query string, the alternative names, the gathered named entities and the context of occurrence of the query string are then combined with a Boolean OR operator to perform a query over the collection. The entity corresponding to the top document returned by Lucene is selected as the answer to the query.

3. Results

The performance results obtained with the run that we submitted were:

MicroAverage : 0.124

B³Precision : 0.086

B³Recall : 0.108

B³F1 : 0.096

References

Nuno Cardoso. REMBRANDT - Reconhecimento de Entidades Mencionadas Baseado em Relações e ANálise Detalhada do Texto. In *Encontro do Segundo HAREM, PROPOR 2008*, 2008.

¹<http://lucene.apache.org/>

²<http://nlp.uned.es/~jperezi/Lucene-BM25/>