AGDISTIS - Agnostic Disambiguation of Named Entities Using Linked Open Data

Ricardo Usbeck[1,2,*], Axel-Cyrille Ngonga Ngomo[1], Sören Auer[1], Daniel Gerber[1], and Andreas Both[2]

[1] University of Leipzig, Germany
{usbeck|ngonga|auer|dgerber}@informatik.uni-leipzig.de
[2] R & D, Unister GmbH, Leipzig, Germany

Abstract. AGDISTIS is an Open Source Named Entity Disambiguation Framework able to link entities against every Linked Data Knowledge Base.

Overview

Over the last decades, several billion Web pages have been made available on the Web. The ongoing transition from the current Web of unstructured data to the Data Web yet requires scalable and accurate approaches for the extraction of structured data in RDF (Resource Description Framework). One of the key steps towards extracting RDF from natural-language corpora is the disambiguation of named entities. While several approaches aim to address this problem, they still achieve poor accuracy on Web/news data (WND). We address this drawback by presenting AGDISTIS, a novel knowledge-base-agnostic approach for named entity disambiguation that performs well on WND. Our approach combines the HITS algorithm with label expansion strategies and string similarity measures. Based on this combination, it can efficiently detect the correct URIs for a given set of named entities within an input text. We evaluate our approach on four different datasets against state-of-the-art named entity disambiguation frameworks. Our results indicate that we outperform the state-of-the-art approach by up to 16% F-measure. Moreover, FOX [2] is a modern Named Entity Recognition Framework based on AGDISTIS. More information about the project can be found here: `http://aksw.org/projects/AGDISTIS`

Knowledge Base

The important data for runnning AGDISTIS is stored in a Lucene 4.4.0 Index that can be found here: `http://139.18.2.164/rusbeck/index_dbpedia_39_en.7z`.

Web Service

We deployed AGDISTIS as a RESTful service reachable via the following command:

```
curl −−data−urlencode
"text='The <entity>University of Leipzig</entity> in
<entity>Barack Obama</entity>.'" −d type=agdistis
http://139.18.2.164:8080/AGDISTIS
```

```
curl −−data−urlencode "text@test.txt" −d type=agdistis
http://139.18.2.164:8080/AGDISTIS
```

AGDISTIS also provides a Wrapper for DBpedia Spotlight [1]. Just change the "type" to "spotlight" instead of "agdistis". Please note that every entity you need disambiguated must be recognized beforehand.

Run your own webservice

For running AGDISTIS on your machine go to the root directory and of AGDISTIS and execute

```
mvn tomcat:run
```

Now a webservice is running on `localhost:8080`

Running from source

The easiest way of running AGDISTIS from source is to have a look at the Java Source `/src/test/java/AGDISTISTest.java`.

We hope you will enjoy using AGDISTIS!

References

1. Pablo N. Mendes, Max Jakob, Andres Garcia-Silva, and Christian Bizer. Dbpedia spotlight: Shedding light on the web of documents. In *Proceedings of the 7th International Conference on Semantic Systems (I-Semantics)*, 2011.
2. Axel-Cyrille Ngonga Ngomo, Norman Heino, Klaus Lyko, René Speck, and Martin Kaltenböck. Scms–semantifying content management systems. In *The Semantic Web–ISWC 2011*, pages 189–204. Springer, 2011.