OpenMinted

1. Deliverable T.4: Final Report

2. Project: BabelNet Extractor

3. Contact Person

Joan Codina
TALN group, Universitat Pompeu Fabra
joan.codina@upf.edu
+34 93 542 2241
C. Tànger, 122-140
08018 Barcelona

4. Final Report

The component is shared as a UIMA component (available in GitHub) and integrated together with BabelNet in a Docker image (also available in hub.docker.com)

Reasons to do a Docker component

The reason to share the component as a Docker image was for the difficulties to run BabelNet inside a java component. Even than BabelNet has a java API, the 26 Gbytes of the indexes have had to be distributed with the component and the best way to do that was using a Docker component.

Highlights and key success factors

The developed UIMA component detects BabelNet1[1] terms in text. The objective of this component is to scan a tokenized text to detect entries in BabelNet in the input document.

[1] R. Navigli and S. Ponzetto. **BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network.** Artificial Intelligence, 193, Elsevier, 2012, pp. 217-250

This component is the base of entity linking and word sense disambiguation as it detects the candidates to be disambiguated. The component produces WSD item annotations as defined in the DKPro WSD typesystem. Afterwards, disambiguation can be performed by other components (like DKPro WSD).

The component development consists in the integration within the OpenMinted platform of one existing solution already developed by the group. The solution has been adapted to the OpenMinted guidelines.
The component is a general purpose component and can be applied to any domain that is covered by BabelNet. It produces a set of outputs that can be reused by other components for disambiguation purposes, among others. The DKPro WSD tools expect as input the terms to be disambiguated, but lacks a component that is able to detect these terms. The BabelNet component integrates perfectly with the DKPro WSD tools, so once the candidates are detected, any kind of WSD or entity linking can be applied.

Unexpected events

- Even that everything went more or less smoothly the main problems were found in packaging the data into the docker and to build it with the right configuration.
- Also it was not clear the way that the docker had to communicate with the platform, but we could solve that by raising an issue in the OpenMINTED's issue tracker.

Balance of the amount of work to adapt and integrate Freeling into OMTD

- We found the task a bit bigger than expected due to the unexpected events: even that we already thought that some problems could arise, debugging into the docker was complex.

Justification of the work done

- We think that effort done has produced a reusable component, so it is justified, the integration of BabelNet into the platform allows the detection of entities in any of the languages supported by BabelNet .

Lessons learned from this project

- For us has been important to view and understand the way to organize and share applications and resources between researchers.

What we would have done differently and/or our recommendations for improvement:

- Timing has been very short.
- It might have been interesting to have a tutorial on how to integrate components, and to have a single skeleton of each kind of component.

5 Dissemination Report

The component has been uploaded in GitHub.com, and hub.docker.com

A blog post entry has been sent to Martine Oudenhoven.

Integration of BabelNet into the OpenMinTeD platform:

Within the context of OpenMinTeD, the integration of a UIMA component detects BabelNet terms in text has been completed. The objective of this component is to scan a tokenized text to detect entries in BabelNet in the input document. This component is the base of entity linking and word sense disambiguation as it detects the candidates to be disambiguated. The component produces WSD item annotations as defined in the DKPro WSD typesystem. Afterwards, disambiguation can be performed by other components (like DKPro WSD). The component, which has been shared as a Docker and the code released in Github with a GPL license, can be found in the following links:

https://hub.docker.com/r/taln/openminted_babelnet/
https://github.com/TalnUPF/OpenMinted_BabelNet

Adapt course:

The course "OpenMINTED: FreeLing and BabelNet Components". Has been uploaded to the adapt platform: http://courses.fosteropenscience.eu