OpeNER demo: Open Polarity Enhanced Named Entity Recognition

Aitor García-Pablos, Montse Cuadros
Vicomtech-IK4 research centre
Mikeletegi 57, San Sebastian, Spain
{agarciap, mcuadros}@vicomtech.org

German Rigau
IXA Group
Euskal Herriko Unibertsitatea,
San Sebastian, Spain
{german.rigau}@ehu.es

Abstract

OpeNER is a project funded by the European Commission under the 7th Framework Programme. Its acronym means Open Polarity Enhanced Named Entity Recognition. OpeNER main goal is to provide a set of open and ready to use tools to perform some NLP tasks in six languages including English, Spanish, Italian, Dutch, German and French. In order to display these OpeNER analysis output in a format suitable for a non-expert human reader we have developed a Web application to display this content in different ways. This Web application should serve as a demonstration of some of the OpeNER modules capabilities after the first year of development.

.**Keywords:** DEMO, OpeNER, named entity recognition, named entity linking, sentiment analysis

1. Introduction

OpeNER[1] is a project funded by the European Commission under the FP7 (7th Framework Programme, project reference: 296451). Its acronym means Open Polarity Enhanced Named Entity Recognition. It is a two year duration project which officially started at July 2012, and finishes at July 2014. In OpeNER are collaborating six partners from Italy, Holland and Spain.

OpeNER main goal is to provide a set of open and ready to use tools to perform some NLP tasks in six languages including English, Spanish, Italian, Dutch, German and French, free and easy to adapt for SMEs to integrate them in their workflow. Some of these NLP task are Named Entity Recognition and Classification, Named Entity Linking and Sentiment Analysis.

The OpeNER project analysis tools produce their output in XML following the KAF[2] format, acronym of Kyoto Annotation Framework (Bosma, et al., 2009). The KAF documents generated by the OpeNER NLP modules are very rich and complex XML files. Obviously, this makes difficult for a human reader to interpret the results by reading the KAF document without any further post-processing.

In order to display these KAF documents in a format suitable for a non-expert human reader we have developed a Web application capable of displaying the content of a KAF document in different ways. This Web application should serve as a demonstration of some of the OpeNER modules capabilities and to explore the results stored in KAF format.

2. Visualization of KAF documents

KAF documents generated by OpeNER analysis modules are XML files with a complex structure. KAF documents are structured in different layers, each one corresponding to a different NLP task (e.g. Tokenization, Part-of-Speech tagging, Named Entity Recognition, etc.). In this format they are difficult to read and interpret by a human reader.

```
[…]
<text>
    <wf wid="w1" sent="1" para="1" offset="0"
        length="3">The</wf>
    <wf wid="w2" sent="1" para="1" offset="4"
        length="4">city</wf>
</text>
<terms>
    <!--The-->
    <term tid="t1" type="close" lemma="the"
        pos="D" morphofeat="DT">
        <span>
            <target id="w1" />
        </span>
    </term>
    <!--city-->
    <term tid="t2" type="open" lemma="city"
        pos="N" morphofeat="NN">
        <span>
            <target id="w2" />
        </span>
    </term>
</terms>
[…]
```

Table 1. A snippet of a KAF document

Among other information contained in a KAF document, there are some aspects that should be displayed in a way suitable for a human. These aspects are:

Detected named entities: Inside a KAF document the detected named entities are stored in their own layer, with a pointer to the span of terms that hold the mention in the text. To visualize this information the text must be reconstructed from its tokens adding an appropriated markup to the named entities span. This markup should allow displaying the text in a Web browser with the

words belonging to a named entity highlighted (for example, rendering them with a different color).

Named entity types: KAF also contains information about the type of detected entities. This information can be used to display the different entity types highlighted in different ways (for example, a color code for each entity type).

Linked entities: In case of a named entity is disambiguated and linked to an external knowledge-base (e.g. to DBpedia). KAF contains the linked resource URI. This information can be used to add hyperlinks to the linked resource when a linked named entity mention occurs. Another approach could be exploiting the linked resource attributes to gather and display further information about the entities.

Polarity words: KAF contains information about the detected polarity of the words in a text. This information in inside the "term" layer, and can be used to highlight positive and negative words in a text.

3. A Web demo for OpeNER

The main approach to display all this information in a more human-readable way is to reconstruct the original text from the KAF document, token by token, highlighting relevant information via HTML markup. Each token is checked against different KAF layers to find out if it belongs to a relevant category (i.e. a named entity, a positive or negative word, etc.). The final result of this highlighted reconstruction is displayed to the user, which can then read the full text and asses if the highlighted words are correctly annotated or not.

The Web application has been organized in three different menus:

NERC/Sentiment demo contains a set of texts for each of the six languages handled by OpeNER. Each text has been preprocessed with the OpeNER NLP tools, and the resulting KAF has been stored. These KAF documents are employed to show different aspects of the analysis.

EMM comparison demo is a named entity recognition comparison between OpeNER and the JRC European Media Monitor News Explorer. The EMM News Explorer gathers and analyzes news from a many sources and in a lot of languages. The EMM News Explorer Web site contains information about the different entities detected in the texts of the news inside each cluster. These entities are classified as "person", "place" and "other". Some of those entities are disambiguated and linked to their respective Wikipedia page. As OpeNER can do a very similar task, it is a good chance to try to compare OpeNER with another working system.

Custom text demo allows writing or copying a piece of text and launching the OpeNER NLP analysis on it. The result is then displayed in the same way than the "NERC/Sentiment demo", with the different analyzed information contained in KAF highlighted in the text.
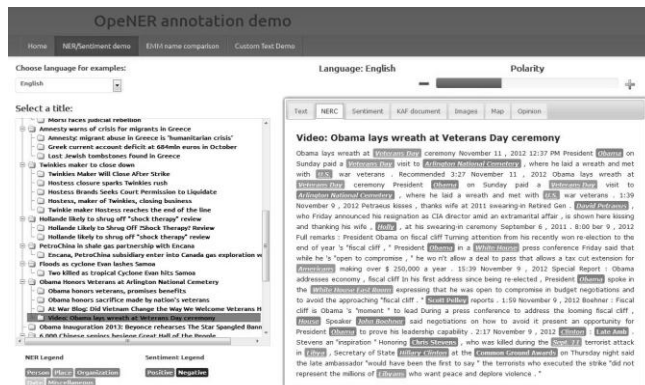

Figure 1. A screenshot from the demo showing highlighted text

The prototype for the demo[3] has been designed as a Web application. It means that it can be accessed from a Web browser with no other requirement for an end user. All the OpeNER NLP modules are deployed as web services[4], which are accessed from the demo, and can be also invoked from any other custom code.
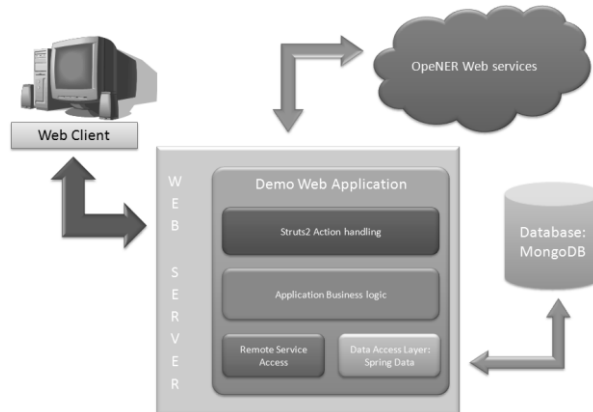

Fig. 2. Basic architecture overview

Acknowledgements

References

Rodrigo Agerri, Montse Cuadros, Sean Gaines and German Rigau (2013). OpeNER: Open Polarity Enhanced Named Entity Recognition, Sociedad Española para el Procesamiento del Lenguaje Natural,Volume 51

Bosma, W., Vossen, P., Soroa, A., Rigau, G., Tesconi, M., Marchetti, Aliprandi, C. (2009). KAF: a generic semantic annotation format. In Proceedings of the GL2009 Workshop on Semantic Annotation.

---

[3] The prototype can be accessed from http://demo-opener.rhcloud.com

[4] The services can be accessed from http://opener.olery.com