

EUROSENTIMENT: Linked Data Sentiment Analysis

J. Fernando Sánchez-Rada¹, Gabriela Vulcu², Carlos A. Iglesias¹, and Paul Buitelaar²

¹ Dept. Ing. Sist. Telemáticos, Universidad Politécnica de Madrid,
{jfernando,cif}@gsi.dit.upm.es,
<http://www.gsi.dit.upm.es>

² Insight, Centre for Data Analytics at National University of Ireland, Galway
{gabriela.vulcu,paul.buitelaar}@insight-centre.org,
<http://insight-centre.org/>

Abstract. Sentiment and Emotion Analysis strongly depend on quality language resources, especially sentiment dictionaries. These resources are usually scattered, heterogeneous and limited to specific domains of application by simple algorithms. The EUROSENTIMENT project addresses these issues by 1) developing a common language resource representation model for sentiment analysis, and APIs for sentiment analysis services based on established Linked Data formats (lemon, Marl, NIF and ONYX) 2) by creating a Language Resource Pool (a.k.a. LRP) that makes available to the community existing scattered language resources and services for sentiment analysis in an interoperable way. In this paper we describe the available language resources and services in the LRP and some sample applications that can be developed on top of the EUROSENTIMENT LRP.

Keywords: Language Resources, Sentiment Analysis, Emotion Analysis, Linked Data, Ontologies

1 Introduction

This paper reports our ongoing work in the European R&D project EUROSENTIMENT, where we have created a multilingual Language Resource Pool (LRP) for Sentiment Analysis based on a Linked Data approach for modelling linguistic resources.

Sentiment Analysis requires language resources such as dictionaries that provide a sentiment or emotion value to each word. Just as words have different meanings in different domains, the associated sentiment or emotion also varies. Hence, every domain has its own dictionary. The information about what each domain represents or how the entries for each domain are related is usually undocumented or implied by the name of each dictionary. Moreover, it is common that dictionaries from different providers use different representation formats. Thus, it is very difficult to use different dictionaries at the same time.

In order to overcome these limitations, we have defined a Linked Data Model for Sentiment and Emotion Analysis, which is based on the combination of several vocabularies: the NLP Interchange Format (NIF) [1], to represent information about texts, referencing text in the web with unique URIs; the Lexicon Model for Ontologies (lemon) [2], to provide lexical information, and differentiate between different domains and senses of a word; Marl [5], to link lexical entries or senses with a sentiment; and Onyx [3], that adds emotive information.

The use of a semantic format not only eliminates the interoperability issue, but it also makes information from other Linked Data sources available for the sentiment analysis process. The EUROSENTIMENT LRP generates language resources from legacy corpora, linking them with other Linked Data sources, and shares this enriched version with other users.

In addition to language resources, the pool also offers access to sentiment analysis services with a unified interface and data format. This interface builds on the NIF Public API, adding several extra parameters that are used in Sentiment Analysis. Results are formatted using JSON-LD and the same vocabularies as for language resources. The NIF-compatible API allows for the aggregation of results from different sources.

The project documentation³ contains further information about the EUROSENTIMENT format, APIs and tools.

2 Language Resources

The EUROSENTIMENT LRP contains a set of language resources (lexicons and corpora). The available EUROSENTIMENT language resources can be found here.⁴ The user can see the domain and the language of each language resource. At the moment the LRP contains resources for electronics and hotel domains in six languages (Catalan, English, Spanish, French, Italian and Portuguese) and we are currently working on adding more language resources from other domains like telco, movies, food and music. Table 1 shows the number of reviews in each available corpus and the number of lexical entries in each available lexicon.

A detailed description of the methodology for creating the domain-specific sentiment lexicons and corpora to be added in the EUROSENTIMENT LRP was presented at LREC 2014 [4].

The EUROSENTIMENT demonstrator⁵ shows how users can benefit from the LRP, including an interactive SPARQL query editor to access the resources and a faceted browser.

3 Sentiment Services

In addition to a model for language resources, EUROSENTIMENT also provides an API for sentiment and emotion analysis services. Several already existing ser-

³ <http://eurosentiment.readthedocs.org>

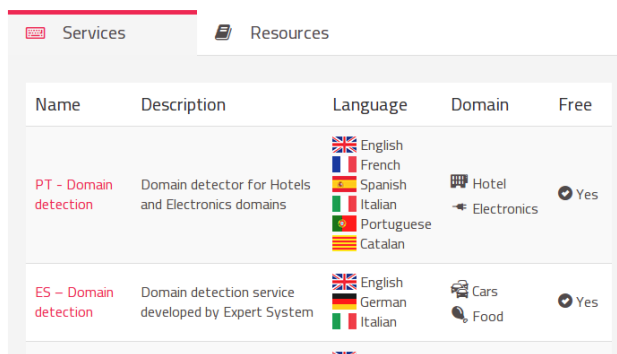
⁴ http://portal.eurosentiment.eu/home_resources

⁵ <http://eurosentiment.eu/demo>

| Lexicons | | | Corpora | | |
|------------|-------------------|-----------|------------|-------------------|-----------|
| Language | Domains | #Entities | Language | Domains | #Entities |
| German | General | 107417 | English | Hotel,Electronics | 22373 |
| English | Hotel,Electronics | 8660 | Spanish | Hotel,Electronics | 18191 |
| Spanish | Hotel,Electronics | 1041 | Catalan | Hotel,Electronics | 4707 |
| Catalan | Hotel,Electronics | 1358 | Portuguese | Hotel,Electronics | 6244 |
| Portuguese | Hotel,Electronics | 1387 | French | Electronics | 22841 |
| French | Hotel,Electronics | 651 | | | |

Table 1. Summary of the resources in the LRP

vices in different languages have been adapted to expose this API. Any user can benefit from these services, which are conveniently listed in the EUROSENTIMENT portal. At the moment, the following services are provided in several languages: language detection, domain detection, sentiment and emotion detection, and text analysis.



| Name | Description | Language | Domain | Free |
|-----------------------|---|--|--|------|
| PT - Domain detection | Domain detector for Hotels and Electronics domains | <ul style="list-style-type: none"> English French Spanish Italian Portuguese Catalan | <ul style="list-style-type: none"> Hotel Electronics | Yes |
| ES - Domain detection | Domain detection service developed by Expert System | <ul style="list-style-type: none"> English German Italian | <ul style="list-style-type: none"> Cars Food | Yes |

Fig. 1. The LRP provides a list of available services

4 Applications Using the LRP

To demonstrate the capabilities of the EUROSENTIMENT LRP, we open-sourced the code of several applications that make use of the services and resources of the EUROSENTIMENT LRP. The applications are written in different programming languages and are thoroughly documented. Using these applications as a template, it is straightforward to immediately start consuming the services and resources. The code can be found on the EUROSENTIMENT Github repositories.⁶

⁶ <http://github.com/eurosentiment>



Fig. 2. Simple service that uses the resources in EUROSENTIMENT to analyse opinions in different languages and domains

Acknowledgements

This work has been funded by the European project EUROSENTIMENT under grant no. 296277

References

- Hellmann, S., Lehmann, J., Auer, S., Nitzschke, M.: Nif combinator: Combining nlp tool output. In: Knowledge Engineering and Knowledge Management, pp. 446–449. Springer (2012)
- McCrae, J., Spohr, D., Cimiano, P.: Linking lexical resources and ontologies on the semantic web with lemon. In: Antoniou, G., Grobelnik, M., Simperl, E., Parsia, B., Plexousakis, D., De Leenheer, P., Pan, J. (eds.) The Semantic Web: Research and Applications, Lecture Notes in Computer Science, vol. 6643, pp. 245–259. Springer Berlin Heidelberg (2011)
- Sánchez-Rada, J.F., Iglesias, C.A.: Onyx: Describing emotions on the web of data. In: ESSEM@ AI* IA. pp. 71–82. Citeseer (2013)
- Vulcu, G., Buitelaar, P., Negi, S., Pereira, B., Arcan, M., Coughlan, B., Sánchez-Rada, J.F., Iglesias, C.A.: Generating Linked-Data based Domain-Specific Sentiment Lexicons from Legacy Language and Semantic Resources. In: th International Workshop on EMOTION, SOCIAL SIGNALS, SENTIMENT & LINKED OPEN DATA, co-located with LREC 2014, LREC2014, Reykjavik, Iceland (May 2014)
- Westerski, A., Iglesias, C.A., Tapia, F.: Linked Opinions: Describing Sentiments on the Structured Web of Data. In: Proceedings of the 4th International Workshop Social Data on the Web (2011)