

# 1. PUBLISHABLE SUMMARY

## **Project vision and objectives**

EUROSENTIMENT aims at creating a **shared pool of shared language resources for fostering sentiment analysis**, accessible by means of well-defined models and frameworks that leverage the promotion of SMEs in the emerging market of Sentiment Analysis products and services. The data pool covers 6 languages -English, Catalan, German, Italian, Portuguese and Spanish- and has been validated through opinion mining demonstrators in two different domains (i.e. hotel and electronic). The targeted users are B2B including service developers, content providers and language resource owners.

The **specific objectives** of the project are:

1. Provide semantic interoperability and connectivity between several multilingual sentiment analysis resources available online for the first time. Semantic interoperability is based on domain ontologies linked to a domain labelled WordNet and compatible with existing Linked Data initiative and EmotionML.
2. Reduce the cost of aggregating new language resources to the shared resource pool by providing best-practice guidelines and QA procedures based on a publicly available multilingual sentiment analysis corpus on two different domains.
3. Provide a self-sustainable and profitable framework for language resource sharing based on a community governance model, which offers contributors unwilling to grant free access the possibility of exploiting commercially the resources they provide.
4. Demonstrate the impact of the developed pool by providing public access to a multilingual demonstrator in the media domain, which show how the different resources can provide high quality results working with specialised language resources, integrate semantically their results and exploit these multilingual results with a semantic front-end.

## **Work performed (M1-M24)**

- Generation of a detailed collection of requirements and use cases for the shared Language Resource Pool, including interoperability, multilingual, security and semantic requirements.
- Review of Sentiment databases and governance models for shared data pools.
- High-level description of the software architecture of the EUROSENTIMENT LRP.
- Final version of Language Resource Model Specification for semantic interoperability.
- Final version of WordNet-Domains Extension for Sentiment Analysis.
- Final version of methodology, guidelines and QA procedures for language resource adaptation and preparation.
- Final version of adaptation of legacy language resources.
- Final version of optimising and verifying automatic acquisition, adaptation and cleanup for large scale language resources.
- Language Resource Pool setup with semantic, service access layer, management and services, accountability, IPR license model and secure access.
- Definition, first and final version of the project demonstrators.
- Evaluation Plan and activities for evaluating and assessing the project evolution and results.
- EUROSENTIMENT's User Group set up and maintenance.
- Project website (eurosentiment.eu) and dissemination material.
- Market Analysis, Exploitation and Liaisons Plan delivered; Liaisons with other projects.
- Dissemination Plan delivered. Dissemination events during the project lifetime.

- Language Resource Pool Management delivered to ensure future partnerships' structure, maintenance and security of the LRP.
- Refinements of the LRP Governance Model.
- Coordination activities and deliverables have been successfully achieved during the project.

### **Key Innovation**

Sentiment analysis can foster the development of new products and services. Nevertheless, the main obstacle to develop these services is the difficulty in accessing to multilingual language resources for sentiment analysis. The main barriers we have identified are:

1. The developed language resources remain scattered and restricted to their customers.
2. Lack of agreed language resource schemas and available multilingual language resources for sentiment analysis.
3. Atomised sentiment analysis projects resulting in reduced language resources visibility, accessibility and interoperability.

In this respect, EUROSENTIMENT innovates providing a domain-oriented shared language resource based on WordNetDomains and aligned with WordNet Affect. The pool is multi-lingual and based on linked data, providing a self-sustainable and profitable framework for language resource sharing.

### **Technical Approach**

EUROSENTIMENT proposes a twofold approach. First, since Opinion mining is strongly domain dependent, a semantic network relating entities and features provides a **uniform semantic interface for users**, and entities have uniform URI naming according to Linked Data conventions. In this way, different sentiment analysers (even in different languages) return semantic triples that can be easily combined and queried in a distributed fashion. Second, since WordNet is the most commonly used language resource, **WordNet-Domains has been extended for Sentiment analysis**, overcoming the limitations of SentiWordNet which is not domain driven. This domain orientation comes also from the analysis of the language resources provided by the consortium which are already domain-oriented. Once the resources are normalised, semantic and service access are provided following a licence model. In order to ensure not only the sustainability of the language resource pool but its business orientation, a community governance model has been defined and applied, following the successful community approach of (profitable) open source communities.

The Language Resource Pool has been built upon a Semantic and Service layer, allowing business users and content providers to access the shared pool through semantic Linked-data-based and REST-based requests following a license model. Service layer provides LRP protocol level access mechanisms in a seamless way, so third parties do not have to deal with issues such as interoperability, multilingualism, internal LRP transactions and secure and IPR functionalities by:

- Processing of datasets with arbitrary transformations in order to derive from source data many possible forms which are useful e.g. for specific reuses.
- Advanced search both at dataset and at specific record level including advanced operators and interactive functionalities
- Full set of APIs to programmatically interact with the repository
- Social aspects to foster crowdsourcing of comments and feedbacks. This allow users to comment at high granularity resources and foster their uses e.g. allowing users to leave examples.
- Recommendation aspects, algorithms that suggest resources given other resources or specific user input datasets.

### **Impact**

1. Improved European competitive position in a multilingual digital market through the provision of better products and services to citizens and businesses.
2. Novel forms of partnership between new programme entrants and established players, reduced development costs and shorter time-to-market, thus stimulating innovation and expanding markets.
3. Result-driven knowledge transfer between research centres (and their spin-offs) and progressive technology providers (especially SMEs), data brokers/aggregators and content providers.

### **Target groups and expected outcomes**

- **Service developers** being SME, large company or end users will be able to use EUROSENTIMENT dataset in order to develop new services by integrating the provided interfaces, based on a commercial license
- **Content providers** can use EUROSENTIMENT dataset in order to increase the value of their contents by aggregating consumers' trends and customers' perceptions
- **Language resource owners** will benefit from having a shared pool where their language resources are valorised thanks to the integration with the rest of resources, resulting in higher visibility and return of investment.

On the supplier side, this language resource pool is targeted at research centres willing to unveil their language resources, while keeping visibility, and SMEs willing to provide complementary language resources to the shared pool and exploiting these resources thanks to its increased visibility which will increment its business opportunities.

### **Official website and dissemination materials**

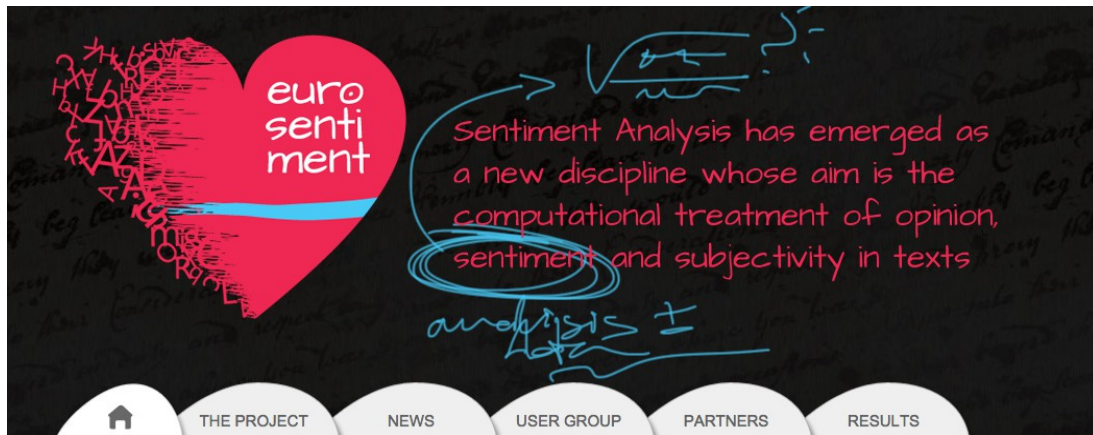
The website of the EUROSENTIMENT project is hosted by Paradigma and is located at <http://www.eurosentiment.eu>

The portal is the core activity for disseminating public information to the visitors and confidential information between the consortium partners. The web site functionality and structure has a background of close study on existing similar sites. Information on the goals and description of the project, as well as news, events and blog posts are publicly available and promoted in our site. Private access can be gained only by the consortium members, in order to collaborate on the common tasks and deliverables. The next illustration shows a sub-section of the home page of the EUROSENTIMENT's web-site which in addition to the information on the EUROSENTIMENT project shows news, blog as well as providing access to the main categories of information.

Dissemination material developed:

- Project logo
- Project website
- Project leaflet (English, Italian and Spanish versions)
- Project presentation (English, Italian and Spanish versions)
- Project poster (English)
- Project roll-up (English)

Publicity, including at a conference or seminar or any type of information or promotional material (brochure, leaflet, poster, presentation etc), has specified that the EUROSENTIMENT project has received Community research funding and displayed the European emblem.



### THE LANGUAGE RESOURCE POOL IS OPEN!

The first version of the EUROSENTIMENT Language Resource Pool (ELRP) is ready! Sign up today and start availing of the published resources and services for your sentiment extraction related project! Even better, create and manage your own resources and services to add a new Sentiment Analysis dimension to your portfolio!

**GO TO THE LANGUAGE RESOURCE POOL**

After you sign up, hands on! you can review the **Documentation**, the **Demo** and **Playground** to test all the features of the Language Resource Pool and start developing and using our services.

Keep in touch! email us at: [project@eurosentiment.eu](mailto:project@eurosentiment.eu)



#### NEWS

**EUROSENTIMENT participates in ISWC 2014**  
 PUBLISHED ON SEPTEMBER 16, 2014  
 9:00 AM  
 BY PARADIGMA  
 NO COMMENTS

EUROSENTIMENT project will have a strong presence in the 13th International Semantic

Illustration 1: EUROSENTIMENT web site <http://www.eurosentiment.eu>

The demonstrator is available at [http://portal.eurosentiment.eu/official\\_demo#](http://portal.eurosentiment.eu/official_demo#)



Home

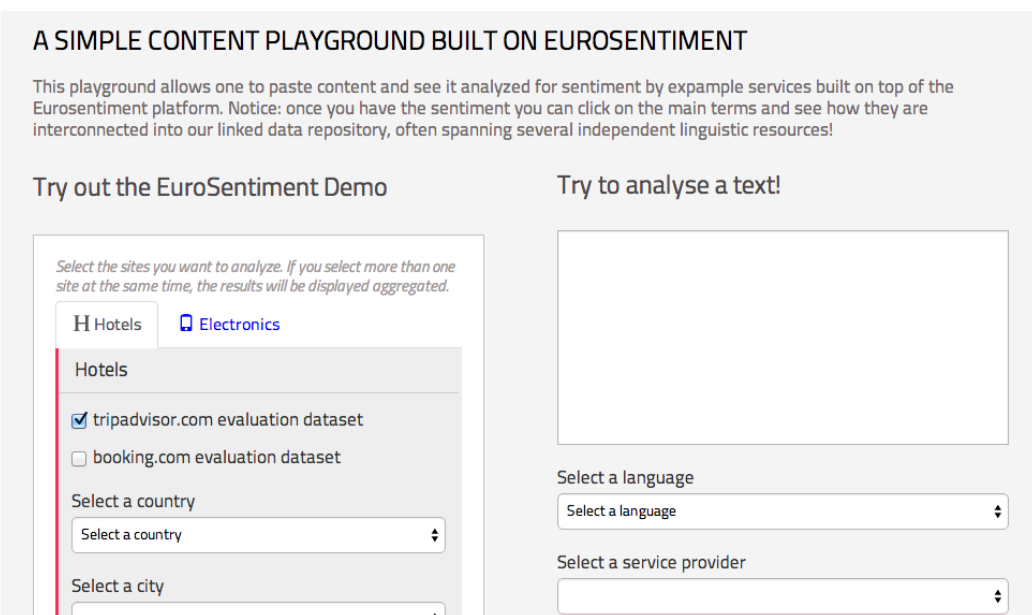


Illustration 2: EUROSENTIMENT demo web site [http://portal.eurosentiment.eu/official\\_demo#](http://portal.eurosentiment.eu/official_demo#)

The LRP management site is available at <http://portal.eurosentiment.eu>

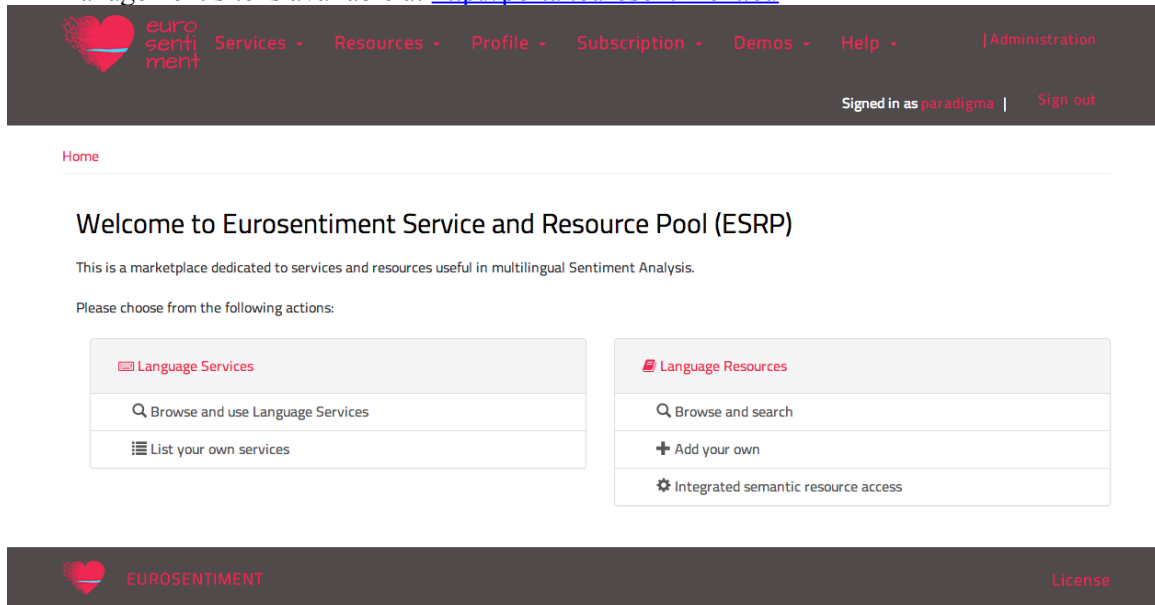


Illustration 3: EUROSENTIMENT LRP management site <http://portal.eurosentiment.eu>

Project leaflets: There exist a version in three languages:

- English: [http://eurosentiment.eu/wp-content/uploads/2014/09/EuroSentiment-leaflet\\_EN\\_v2.0.pdf](http://eurosentiment.eu/wp-content/uploads/2014/09/EuroSentiment-leaflet_EN_v2.0.pdf)
- Spanish: [http://eurosentiment.eu/wp-content/uploads/2012/12/EuroSentiment-leaflet\\_ES\\_v1.0.pdf](http://eurosentiment.eu/wp-content/uploads/2012/12/EuroSentiment-leaflet_ES_v1.0.pdf)
- Italian: [http://eurosentiment.eu/wp-content/uploads/2012/12/EuroSentiment-leaflet\\_IT\\_v1.0.pdf](http://eurosentiment.eu/wp-content/uploads/2012/12/EuroSentiment-leaflet_IT_v1.0.pdf)

Project presentations: There also exist a version in three languages:

- English: [http://eurosentiment.eu/wp-content/uploads/2012/12/Project\\_Presentation\\_EN\\_v1.0.pdf](http://eurosentiment.eu/wp-content/uploads/2012/12/Project_Presentation_EN_v1.0.pdf)
- Spanish: [http://eurosentiment.eu/wp-content/uploads/2012/12/Project\\_Presentation\\_ES\\_v1.0.pdf](http://eurosentiment.eu/wp-content/uploads/2012/12/Project_Presentation_ES_v1.0.pdf)
- Italian: [http://eurosentiment.eu/wp-content/uploads/2012/12/Project\\_Presentation\\_IT\\_v1.0.pdf](http://eurosentiment.eu/wp-content/uploads/2012/12/Project_Presentation_IT_v1.0.pdf)



Illustration 4: EUROSENTIMENT English leaflet

EUROSENTIMENT rollup:



## LANGUAGE RESOURCE POOL FOR SENTIMENT ANALYSIS IN EUROPEAN LANGUAGES

Establish a market for Semantically Interoperable Language Resources in Sentiment Analysis.

[www.eurosentiment.eu](http://www.eurosentiment.eu)

Demo:

[www.eurosentiment.eu/demo](http://www.eurosentiment.eu/demo)

### PARTNERS

{paradigma

EXPERT  
SYSTEM  
SEMANTIC INTELLIGENCE

FBK  
FONDAZIONE  
BRUNO KESSLER

Insight

SINDICATECH  
SEMANTIC & ANALYTICAL


POLITECNICA

Grant Agreement no: 296277  
Starting date: 01/09/2012  
Project duration: 24 months



*Illustration 5: EUROSENTIMENT Rollup*

Project poster:




**EUROSENTIMENT: Shared Language Resource Pool for Sentiment Analysis**

Co-funded by EU FP7-ICT 2011 SME Digital Contents and Languages

Project start: September 2012 (24 months)

<http://www.eurosentiment.eu>






Grant no. 296277


---

## Linked-Data based Domain-Specific Sentiment Lexicons



Gabriela Vulcu, Raul Lario Monje, Mario Munoz, Paul Buitelaar, Carlos A. Iglesias  
 Insight, Centre for Data Analytics, National University of Ireland, Galway, Ireland  
 Paradigma Tecnológico, Madrid, Spain  
 Universidad Politécnica de Madrid, Spain

**From review datasets to domain-specific lexicons**

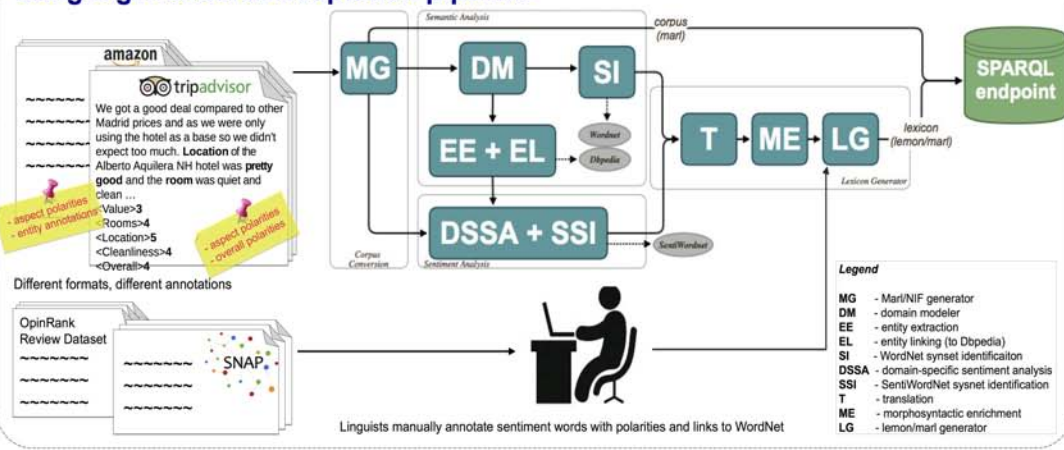


WordNet

OpinRank Dataset

**Language resource adaptation pipeline**

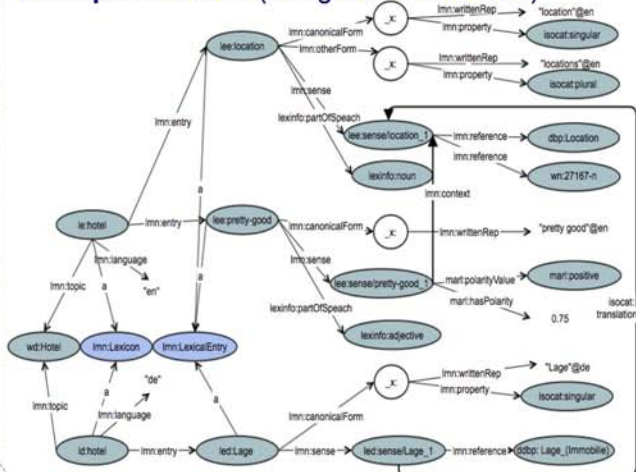


Linguists manually annotate sentiment words with polarities and links to WordNet

**Legend**

- MG - MarI/NIF generator
- DM - domain modeler
- EE - entity extraction
- EL - entity linking (to Dbpedia)
- SI - WordNet synset identification
- DSSA - domain-specific sentiment analysis
- SSI - SentiWordNet synset identification
- T - translation
- ME - morphosyntactic enrichment
- LG - lemon/marl generator

**Example lexicons (using lemon and MarI)**



**Sentiment words the 'electronics' domain.**

Sentiment	PolarityValue	Context Entity
"good"@en	0.5	"alarm"@en
"damaged"@en	-1.0	"apple"@en
"amazed"@en	1.0	"flash"@en
"expensive"@en	-0.5	"flash"@en
"annoying"@en	-0.75	"player"@en

**Sentiment word 'warm' in the 'hotel' domain.**

Sentiment	PolarityValue	Context Entity
"warm"@en	1.0	"pastries"@en
"warm"@en	1.0	"comfort"@en
"warm"@en	0.9	"restaurant"@en
"warm"@en	0.85	"service"@en
"warm"@en	0.45	"hotel"@en














Illustration 6: EUROSENTIMENT poster