

PROJECT FINAL REPORT

Grant Agreement number: 296277

Project acronym: EUROSENTIMENT

Project title: Language Resource Pool for Sentiment Analysis in European Languages

Funding Scheme: STREP

Period covered: from September 2012 to August 2014

Name of the scientific representative of the project's co-ordinator¹, Title and Organisation:

Mr. Daniel Molina, Paradigma Tecnológico S.L.

Tel: +34 91 352 59 42

Fax: +34 91 715 89 66

E-mail: daniel.molina@paradigmatecnologico.com

Project website address: <http://www.eurosentiment.eu>

¹ Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

Table of contents

1	Final publishable summary report	3
1.1	Executive summary	3
1.2	Project context and objectives	3
1.3	Main S&T results	6
1.4	Potential impact. Main dissemination and exploitation activities	7
1.5	Project website and contact details	11
2	Use and dissemination of foreground	14
3	Report on societal implications	31

1 Final publishable summary report

1.1 Executive summary

This document describes the Project Final Report of EUROSENTIMENT project according to the template provided by the European Commission to detail project objectives, results, dissemination and exploitation activities and potential impact. Complementary information at both management and financial level is included in the deliverable D1.4: EUROSENTIMENT Second Periodic Project Report.

The document is structured as follows. Section 1.2 details the project context, key innovation, objectives, research focus and technical approach. Section 1.3 describes the main S/T results of the project. Section 1.4 details the impact of EUROSENTIMENT (together with Part B2 of this document) and the main dissemination and exploitation activities (Workshop, Hackathon, dissemination material), liaisons and coordination of standardisation efforts. Section 1.5 depicts the project website and partners' contacts.

Section 2 shows the use and dissemination of foreground, including the list of scientific publications of the project and dissemination activities and the exploitable foreground.

Section 3 includes a set of tables concerning societal implications required by the Commission.

1.2 Project context and objectives

Context

The web contains a wealth of product reviews, opinions and sentiments that users publish freely in blogs, review sites and social networks. This user-generated content (UGC) represents a valuable source of information for individuals, businesses and governments. Nevertheless, user generated content proliferation has led to an explosion of customer reviews and opinions, which makes infeasible its manual processing. Thus, **Sentiment Analysis** has emerged as a new discipline whose aim is the computational treatment of opinion, sentiment and subjectivity in texts, often available in so-called social media. Sentiment analysis, also called Opinion mining, combines different techniques from natural language processing, computational linguistics, text analytics and semantic analysis, in order to extract and identify subjective information in source materials. Some of the main business applications of sentiment analysis are brand and reputation management, social media monitoring, mood analysis, advertisement optimisation or product comparison.

Given the wide range of business oriented applications of sentiment analysis products, there exists a starting market for such products, with a reduced time to market. Sentiment analysis can foster the development of new products and services, such as predicting stock market evolution measuring the emotions expressed in Twitter or weighting opinions according to the social relationships of a user. 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:

- Even though this field has fostered the innovation in a range of European SMEs with strong semantic and linguistics skills, the developed language resources remain restricted to their customers. The main reason is their fear for losing competitiveness or missed returns on research / business investment.
- Lack of agreed language resource schemas for sentiment analysis and not normalised magnitudes for measuring sentiment strength.
- Lack of available multilingual language resources for sentiment analysis, being most of them available only for English.

- Atomised sentiment analysis projects lead by SMEs and research centers which are often accessible for a small community, resulting in reduced language resources visibility, accessibility and interoperability.
- Scattered language resources should be bundled together in order to achieve the added value for the user that comes from accessing a single large online language resource pool.
- One of the factors that make opinion mining difficult is that sentiment and subjectivity are quite context-sensitive and quite domain dependent. Even the same expression (“Go read the book”) can indicate different sentiments for book reviews (positive) or movie review (negative). Even though available language resources for Sentiment Analysis, such as WordNet-Affect or SentiWordNet provide prior polarity lexicon (fixed polarity), adjective polarity is often domain dependent and target dependent. For example, cold pizza (negative) and cold coke (positive) in the food domain.

Taking into account the relevance of these barriers, it is obvious that many of them convey at the same time great **opportunities**:

- Social media is becoming an integral part of modern society and its adoption is constantly growing. Thus, there is huge room for innovation for high-tech SMEs exploiting the power of user generated data with technologies such as sentiment analysis as far as they can access available language resources.
- The strength of European language industry is also an opportunity for the development of opinion mining services, with a growth rate of 10% in spite of the economic crisis [EU09]. Thus, it is not surprising that many of the available language resources for Sentiment Analysis, such as Wordnet-Affect, SentiWordNet or DBPedia [Bi09] have been developed in Europe. In addition, European SMEs have filled the gap between the state of the art in Opinion mining and client requirements and have developed high-quality solutions supported by rich language resources, combining manual and automatic methods.
- The lack of available language resource pools as well as specifications for sharing these resources is a competitive advantage, since being the first in providing such resources will encourage its adoption and usage.
- There is a growing demand for multilingual resources. Recently, most of the Internet use growth was supported by non-native English speakers: starting 2000, for non-English speaking regions, the growth has surpassed 3,000% to compare with 342 % of the over-all growth².

Based on the detection of these opportunities as well as the expertise developed from the beneficiaries that form the project consortium, the **main concept** of the project is to provide a **shared language resource pool for fostering sentiment analysis**. To this end, a detailed collection of requirements of language resource providers and users will be performed with the aim of satisfying their needs in terms of multilingual, quality and domain coverage. Then, next step will be providing interoperability between resources, cleaning and aligning them in order to provide a homogeneous interface. EUROSENTIMENT proposes a twofold approach. First, since Opinion mining is strongly domain dependent, a semantic network relating entities and features will provide a uniform semantic interface for users, and entities will have uniform URI naming according to Linked Data conventions. In this way, different sentiment analysers (even in different languages) will return semantic triples that can be easily combined and even more, queried in a distributed fashion. Second, since WordNet is the most commonly used language resource, WordNet-Domains will be 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

²<http://www.internetworldstats.com/stats.htm>

service access will be 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 will be defined and applied, following the successful community approach of (profitable) open source communities.

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.

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.

Objectives

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.

Research focus: Sentiment Analysis

Sentiment Analysis usually follows these phases: **(i) Entity extraction** (also called Item extraction) is the process of extracting the subject matter where opinions have been expressed; **(ii) Feature extraction** is the identification of the features of the entities on which users have expressed their

opinions in their reviews, comments and feedbacks. For example, if the entity is a car, the features could be its price, design or safety; **(iii) Sentiment / Affect quantification** of opinions. This task classifies opinions about features and entities as positive, negative or neutral, and measures the strength of this polarity. This concept has evolved to identify the affective (also so-called emotional) state of the user (anger, sadness, love, fear disgust, etc.). Emotional states are particularly useful for understanding people sentiment in public opinion, for example. Finally, **(iv) Opinion processing and visualisation** can help end users to take advantage of the gathered opinions, for different tasks such as product comparison, supplying complementary information to customers in e-business or media contents.

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.

1.3 Main S&T results

The project has produced valuable output for highly innovative services with focus on language resources pool sharing.

The target groups of this resource pool are: (i) service developers, (ii) content providers and (iii) language resource owners. 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. Finally, language resource owners will benefit from having a shared pool where there

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.

The project participants contributed with high-quality language resources for Sentiment Analysis in 6 European languages and up to 8 domains.

Specific S/T outcomes of the project are

- 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.

Public S/T Software and Resources

- GitHub repositories (UPM):
 - EuroSentiment MARL Generator: Portal to convert legacy data to semantic formats (e.g. Marl and Onyx in JSON-LD) (GitHub³ / Website⁴).
 - SEAS⁵: Sentiment Analysis Services framework, open source reference implementation of EUROSENTIMENT.
 - SAGA⁶: Gate module for Sentiment Analysis.

³<https://github.com/gsi-upm/eurosentiment-generator>

⁴<http://demos.gsi.dit.upm.es/eurosentiment/marlgenerator>

⁵<https://github.com/gsi-upm/SEAS>

⁶<https://github.com/gsi-upm/SAGA>

- EuroSentiment Playground⁷ (UPM)
- Domain modeller component⁸ (NUIG)
- Entity extraction and entity linking component (NUIG):
 - aela-eurosentiment0.3.tar.gz⁹
 - gazetteer.tar.gz¹⁰
 - neighboursIndex.tar.gz¹¹
- Synset Identification component¹² (NUIG)
- Domain-Specific Sentiment Analysis component¹³ (NUIG)
- Lemon/marl generator¹⁴ (NUIG)
- Translation component¹⁵ (NUIG)
- Morphosyntactic component¹⁶ (NUIG)
- SentiBoost¹⁷ (FBK)
- EUROSENTIMENT Use Case¹⁸ in W3C Ontolex Community¹⁹
- MARL ontology²⁰ (UPM)
- ONYX ontology²¹ (UPM)
- WordNet-Affect Taxonomy²² (UPM)

1.4 Potential impact. Main dissemination and exploitation activities

Impact

The strategic impact of the project can be summarised as follows:

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.

⁷<http://demos.gsi.dit.upm.es/eurosentiment-playground>

⁸<https://github.com/insight-unlp/domainmodeller>

⁹<https://dl.dropboxusercontent.com/u/17176685/EUROSENTIMENT-code/aela-eurosentiment0.3.tar.gz>

¹⁰<https://dl.dropboxusercontent.com/u/17176685/EUROSENTIMENT-code/gazetteer.tar.gz>

¹¹<https://dl.dropboxusercontent.com/u/17176685/EUROSENTIMENT-code/neighboursIndex.tar.gz>

¹²<https://www.dropbox.com/s/henii3iyigjf92l/wnsd-1.2-full.tar.gz>

¹³<https://dl.dropboxusercontent.com/u/40572132/EuroSentiment.zip>

¹⁴<https://dl.dropboxusercontent.com/u/17176685/EUROSENTIMENT-code/LemonMarlGenerator.zip>

¹⁵[git://github.com/moses-smt/mosesdecoder.git](https://github.com/moses-smt/mosesdecoder.git)

¹⁶<https://dl.dropboxusercontent.com/u/17176685/EUROSENTIMENT-code/morphosyntactic.zip>

¹⁷<https://bitbucket.org/hardest/sentiboost>

¹⁸<http://www.w3.org/community/ontolex/wiki/Eurosentiment>

¹⁹http://www.w3.org/community/ontolex/wiki/Main_Page#Use_Cases

²⁰<http://www.gsi.dit.upm.es/ontologies/marl>

²¹<http://www.gsi.dit.upm.es/ontologies/onyx>

²²<http://www.gsi.dit.upm.es/ontologies/wnaffect>

EUROSENTIMENT impact to the actors in the language resources for sentiment analysis value chain:

- **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.

Main Dissemination and exploitation activities

EUROSENTIMENT Workshop

The Consortium of the project celebrated the Workshop of the project as part of the 5th International EMOTION, SOCIAL SIGNALS, SENTIMENT & LINKED OPEN DATA Workshop (ES3LOD²³), arranged as a side event of LREC 2014 Conference²⁴ held in Reykjavik, Iceland, on June 2014.

EUROSENTIMENT proposed a workshop as joint initiative with Opener and TrendMiner projects for LREC 2014 titled: SALD: Sentiment Analysis and Linked Data. This workshop was accepted, and LREC organisation proposed to merge it with ES3LOD Workshop chaired by Björn Schuller.



Illustration 1: ES3LOD front page

EUROSENTIMENT participation in the Workshop

During the workshop, EUROSENTIMENT team participated in a varied set of dissemination activities in order to spread the work about the project to potential users and client, and reach potential collaborators:

- Tutorial: “Linked Data for Language Technologies”: Gabriela Vulcu presented the “EUROSENTIMENT” use-case of modelling language resources with NIF, Lemon and Marl

²³<http://emotion-research.net/sigs/speech-sig/es3lod>

²⁴<http://lrec2014.lrec-conf.org/en>

in the talk “Generating Linked-Data based Domain-Specific Sentiment Lexicons from Legacy Language and Semantic Resources”.

- During the Session 1 -Markup and Linked Data- Gabriela Vulcu presented the paper “Generating Linked-Data based Domain-Specific Sentiment Lexicons from Legacy Language and Semantic Resources” – Gabriela Vulcu, Paul Buitelaar, Sapna Negi, Bianca Pereira, Mihael Arcan, Barry Coughlan, J. Fernando Sánchez and Carlos A. Iglesias”. The presentation described the research challenges of EUROSENTIMENT and how they were addressed in terms of Language Format Heterogeneity for Language Resources. The presentation also presented the Pipeline for Language Resource Adaptation developed by NUIG in the context of EUROSENTIMENT. Gabriela Vulcu detailed every component of the Pipeline and showed several examples of work.



Illustration 2: Gabriela Vulcu presenting EUROSENTIMENT during the ES3LOD Workshop

- During Session Keynote II and Plenary Discussion, Carlos A. Iglesias presented “A linked data approach for describing sentiments and emotions”. His presentation covered the following aspects:
 - Emotion & Sentiment Analysis: E&S models and Challenges
 - EUROSENTIMENT: Linked Data perspective
 - Vocabularies: Marl & Onyx
 - Service and Annotation - NIF
 - Lexical model: Lemon
 - W3C CG on LD models for E & S Analysis
- Carlos A. Iglesias chaired the Session 4 of the workshop, where papers selected covering the topic of Social Networks were presented.
- EUROSENTIMENT team presented a poster at the Data Challenge at the 3rd Workshop on Linked Data in Linguistics "Linked-Data based Domain-Specific Sentiment Lexicons" - Gabriela Vulcu, Raul Lario Monje, Mario Munoz, Paul Buitelaar and Carlos A. Iglesias. The poster was awarded as best poster in the LREC Conference.

- EUROSENTIMENT booth at the LREC EC-village²⁵: Francesco Danza, Gabi Vulcu and Paul Buitelaar presented EUROSENTIMENT to attendees of LREC. The project raised interest and many people asked about how EUROSENTIMENT addresses the multilingual sentiment analysis and how the Language Resources and Services marketplace works. Several interesting prospects were done during the conference and the booth allowed to severely increase the visibility of the activity performed in the project.
- Discussion about W3C Linked Data Models for Emotion and Sentiment Analysis Community Group²⁶ proposed by EUROSENTIMENT's partners NUIG/Insight, UPM, Paradigma. Chairs: J. Fernando Sánchez and Bjorn Schuller. This slot served as a starting point for the Community Group, general ideas were shared and several people contributed and joined the Group. After this kick-off event, several audio conferences have followed and the activity of the Group has officially started.



Illustration 3: EUROSENTIMENT team at the poster session in LREC.

²⁵http://lrec2014.lrec-conf.org/media/filer_public/2013/10/22/hlt_village_at_lrec_2014.pdf

²⁶<http://www.w3.org/community/sentiment>



Illustration 4: EUROSENTIMENT booth with Francesco Danza, Gabriela Vulcu and Paul Buitelaar.

EUROSENTIMENT hackathon

EUROSENTIMENT hackathon²⁷ was organised by UPM last July 14th in Madrid. The goal was to use existing services and resources for sentiment analysis to create original applications and services. To this end, participants had access to EUROSENTIMENT portal and APIs.

Participants were organised into teams of 3 people and developed applications. After development time, all teams made a brief presentation of their application or service.

The heterogeneity the ideas showed the many different applications of Sentiment Analysis, and the possibilities that a multilingual platform like EuroSentiment has to offer.

Five teams: ISIS, who tried to compare how different forums felt about certain topics (apple vs android); Rush, who used the API and data from Reddit to measure how each release of a new game was seen by the public; Merryweather, who came up with the idea of measuring sentiment about certain topics in mass media and social media (such as Twitter) to compare them and detect possible censorship and the influence of media; Enjoy the silence, who built the “Synesthesia” website that

²⁷<http://www.eventbrite.com/e/hackathon-de-analisis-de-sentimientos-eurosentiment-tickets-12134967013>

detected emotions in the lyrics of a song and used that information to create an original visualisation of the emotion which allowed deaf people to “feel” the music in a different way; and last but not least ELJEPE, whose original idea was to compare how people in different countries express themselves in Twitter, based on the use of foreign words and sentiment in each tweet.

After an exhausting yet challenging time, the participants themselves voted the application by Merryweather as the best. This team was formed by Pablo Sanz, Eloy Cepa y Eloy Ramírez, each of whom got a Leap Motion as a price.



Illustration 5: EUROSENTIMENT hackathon: J. Fernando (UPM) introducing the event.



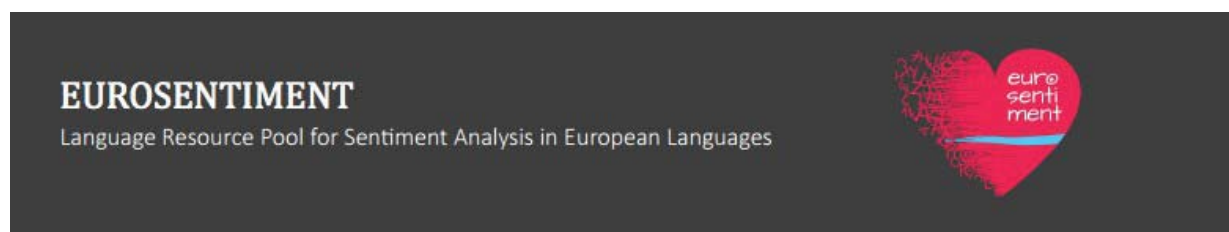
Illustration 6: EUROSENTIMENT Hackaton: Awarded team

Dissemination material

To address the audience as one project, it is important to create a corporate identity among all the partners. A basic step to achieve such identity is through the use of a complete set of materials that clearly represents the goals of the project and link every partner to the image of the project. The basic package for the EUROSENTIMENT project during the first year has consisted of:

- Project logo
- Project website
- Brochure
- 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.



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 will cover 6 languages -English, Catalan, German, Italian, Portuguese and Spanish- and will be 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:

- 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.
- 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.
- 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.
- Demonstrate the impact of the developed pool by providing public access to a

Project title
Language Resource Pool for Sentiment
Analysis in European Languages

Grant Agreement no: 296277
Call: FP7-ICT-2011-SME-DCL
Starting date: 01/09/2012
Duration: 24 months

Coordinator
Paradigma Tecnológico (Spain)

Contact Person
Daniel Molina
daniel.molina@paradigmatecnologico.com

Consortium
Fondazione Bruno Kessler (Italy)
Expert System (Italy)
Sindice (Ireland)
Universidad Politécnica de Madrid (Spain)
National University of Ireland, Galway
(Ireland)

Illustration 7: 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

Demo:

www.eurosentiment.eu/demo

PARTNERS




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



Illustration 8: 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>

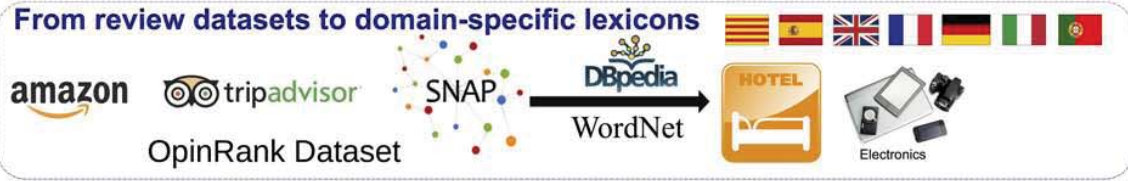


Grant no. 296277

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 Politecnica de Madrid, Spain

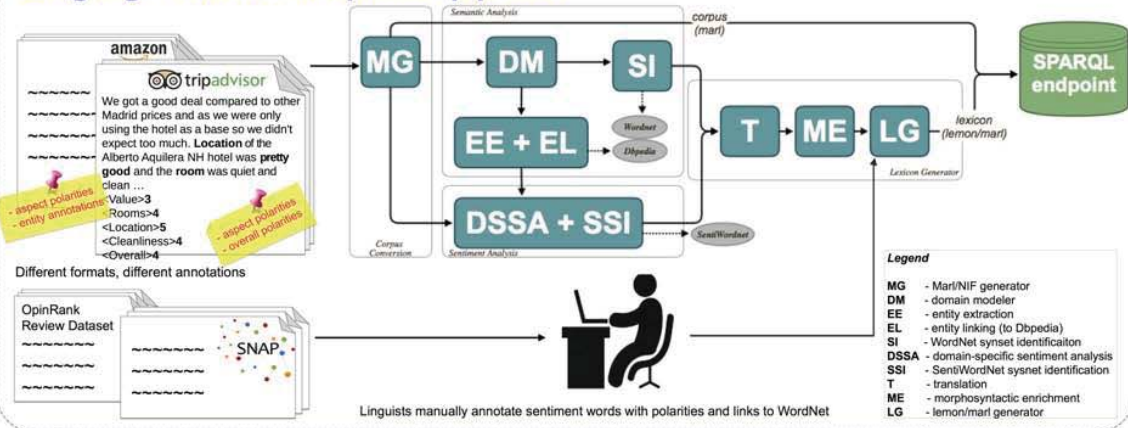
From review datasets to domain-specific lexicons



amazon tripadvisor SNAP DBpedia WordNet HOTEL Electronics

OpinRank Dataset

Language resource adaptation pipeline



amazon, tripadvisor, SNAP, OpinRank Review Dataset, WordNet, SPARQL endpoint

MG, DM, SI, EE + EL, DSSA + SSI, T, ME, LG

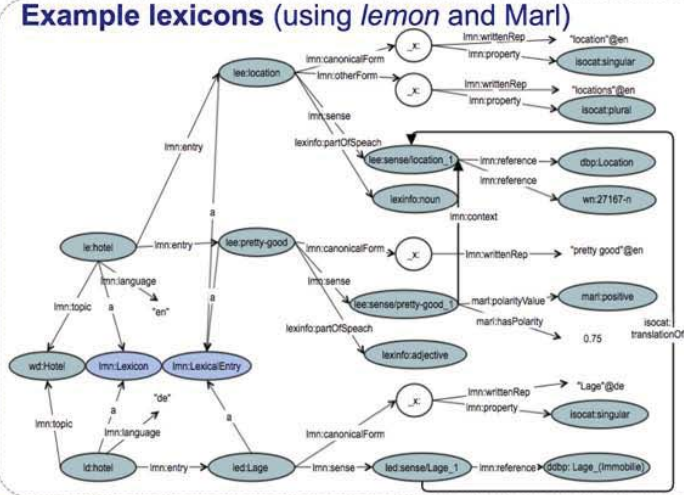
corpus (marl), lexicon (lemon/marl)

Legend

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

Linguists manually annotate sentiment words with polarities and links to WordNet

Example lexicons (using lemon and Marli)



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 9: EUROSENTIMENT poster

Dissemination and exploitation activities

The dissemination activities have included the following international events, conferences and papers:

- BigData Spain: event organised by Paradigma Tecnológico in November 16th, 2012. <http://www.bigdataspain.org/en>
- FBK: Invited talk "Emotions and Creative Language" at 1st Workshop on Practice and Theory of Opinion Mining and Sentiment Analysis September 21, 2012 in Vienna, Austria.
- FBK: Invited talk at the "The Creative Web: Computational Creativity as a Web-Service" An international symposium hosted by the Web Science & Technology Division* of KAIST, the Korea Advanced Institute of Science and Technology, December 3 – 4, 2012.
- FBK: Conference Paper Carlo Strapparava, Oliviero Stock and Ilai Alon corpus-based Explorations of Affective Load Differences in Arabic-Hebrew-English In proceedings of 24th International Conference on Computational Linguistics, Mumbai, India, 8-15 December 2012.
- NetMob 2013 (USA): Paradigma Tecnológico was selected to present a poster for the "NetMob 2013" conference at the MIT Media Lab,. The Business Manager of EUROSENTIMENT attended to network both telco industry and research groups in order to broaden potential users for the Resource Pool and the User Group of the project with more target users such as language resource owners.
- MUMIA COST action: In June 2013 Paul Buitelaar (DERI-NUIG) presented EUROSENTIMENT project (objectives, partners, its use of lemon) at a meeting of the MUMIA COST action on Multilingual and Multifaceted Interactive Information Access in Tallin, Estonia.
- ESWC Conference: In early June 2013 Paul Buitelaar (DERI-NUIG) mentioned EUROSENTIMENT project objectives, its use of lemon) in the context of a presentation on the outcomes of the [Monnet project](#) at the EU project networking meeting at [Extended Semantic Web Conference \(ESWC\)](#) in Montpellier, France.
- EUROSENTIMENT User Group meeting: During the plenary meeting in Rome, members of the User Group were invited to establish a first eye-to-eye contact, present the project and the User Group strategy.
- OpeNER's hackaton and User Group meeting: J. Fernando Sánchez (UPM), attended the Hackaton around Natural Language Technology and Text Mining large datasets organised in Amsterdam by [OpeNER](#), a FP7 project liaised to EUROSENTIMENT. On July 3rd J. Fernando attended the user group meeting. Fernando introduced EUROSENTIMENT and discussed future promising cooperation lines with OpeNER.
- Metaforum 2013: International conference on powerful language technologies for the multilingual information society, the data value chain and the information market place. EUROSENTIMENT was represented by a poster on September 19/20th. <http://www.meta-net.eu>
- 2nd Workshop on Linked Data in Linguistics (LDL-2013): The paper "Linguistic Linked Data for Sentiment Analysis", by Paul Buitelaar, Mihael Arcan, Carlos Iglesias, Fernando Sánchez and Carlo Strapparava, has been accepted. <http://ldl2013.org>

- TASS 2013: Workshop for Sentiment Analysis and online reputation analysis focused on Spanish language, organised as a satellite event of the annual [SEPLN Conference](#). EUROSENTIMENT was represented by a poster on September 20th in Madrid.
- Foundation of W3C Community for Sentiment Analysis: EUROSENTIMENT project has promoted a new W3C Community Group to foster sentiment analysis research: [Linked Data Models for Emotion and Sentiment Analysis Community Group](#). Topics to be addressed are:
 - Definition of a Linked Data based vocabulary for emotion and sentiment analysis.
 - Requirements beyond text-based analysis, i.e. emotion/sentiment analysis from images, video, social network analysis, etc.
 - Clarifying requirements and the need for consensus as e.g. systems currently use widely varying features for describing polarity values (1-5, -2/-1/0/1/2, positive/neutral/negative, good/very good etc.).
 - Marl and Onyx are vocabularies for emotion and sentiment analysis that can be taken as a starting point for discussion in the CG.
- EDF 2014: During 19-20th March, EUROSENTIMENT project was present in the [European Data Forum](#) in Athens. The project set up an exhibition booth and presented the initial version of the Language Resource Pool and the [demonstrator](#) delivered in February.
- API Days Mediterranea in Barcelona (May, 2014): Paradigma gave a talk on Semantic Web, Linked Data and Construction of Opinion and more specifically about EUROSENTIMENT. The overall level of talks, good organisation and interesting conversations with colleagues contributed to a positive balance of the event. Paradigma introduced EUROSENTIMENT key points and introduce theoretical concepts of Semantic Web, Linked Data and NLP. they emphasised a couple of case studies to illustrate how to create own language resources that may be useful in sentiment analysis algorithm. <http://mediterranea.apidays.io>
- LREC 2014 (June, 2014): The project had a significant presence in the 9th Edition of the Language Resources and Evaluation Conference (LREC), held held at the Harpa Conference Centre in Reykjavik (Iceland) during the week of 26 – 31 May 2014. EUROSENTIMENT had a booth at the event, where the details of the project could be presented to numerous people interested in the Language Resource Pool solution. <http://lrec2014.lrec-conf.org/en>
- 3rd Workshop on Linked Data in Linguistics LREC 2014, side event (June, 2014): EUROSENTIMENT partners NUIG/Insight, UPM and Paradigma presented a poster at the Data Challenge at the 3rd Workshop on Linked Data in Linguistics "Linked-Data based Domain-Specific Sentiment Lexicons" - Gabriela Vulcu, Raul Lario Monje, Mario Munoz, Paul Buitelaar and Carlos A. Iglesias <http://ldl2014.org>
- Co-Chairman in ES3LOD Workshop on EMOTION, SOCIAL SIGNALS, SENTIMENT & LINKED OPEN DATA LREC 2014, side event (June, 2014): EUROSENTIMENT workshop was celebrated as a side event of LREC 2014 conference held in Reykjavik, Iceland. Activities of EUROSENTIMENT in the workshop included:
 - Tutorial: "Linked Data for Language Technologies": Gabriela Vulcu presented the "EUROSENTIMENT" use-case of modelling language resources with NIF, Lemon and Marl in the talk "Generating Linked-Data based Domain-Specific Sentiment Lexicons from Legacy Language and Semantic Resources".
 - Carlos A. Iglesias presented the Keynote II (A linked data approach for describing sentiments and emotions): "Generating Linked-Data based Domain- Specific Sentiment Lexicons from

Legacy Language and Semantic Resources” – Gabriela Vulcu, Paul Buitelaar, Sapna Negi, Bianca Pereira, Mihael Arcan, Barry Coughlan, J. Fernando Sánchez and Carlos A. Iglesias

- Discussion about W3C Linked Data Models for Emotion and Sentiment Analysis Community Group proposed by EUROSENTIMENT’s partners NUIG/Insight, UPM, Paradigma. Chairs: J. Fernando Sánchez and Bjorn Schuller. <http://emotion-research.net/sigs/speech-sig/es3lod>
<http://lrec2014.lrec-conf.org/en>
- EUROSENTIMENT Hackathon: On July 14th EUROSENTIMENT project celebrated a hackathon in UPM facilities in Madrid, where five groups of students competed to develop the best application to use the EUROSENTIMENT Portal in as little as 6 hours. Five teams: ISIS, who tried to compare how different forums felt about certain topics (apple vs android); Rush, who used the API and data from Reddit to measure how each release of a new game was seen by the public; Merryweather, who came up with the idea of measuring sentiment about certain topics in mass media and social media (such as Twitter) to compare them and detect possible censorship and the influence of media; Enjoy the silence, who built the “Synesthesia” website that detected emotions in the lyrics of a song and used that information to create an original visualisation of the emotion which allowed deaf people to “feel” the music in a different way; and last but not least ELJEPE, whose original idea was to compare how people in different countries express themselves in Twitter, based on the use of foreign words and sentiment in each tweet. The heterogeneity of these ideas shows the many different applications of Sentiment Analysis, and the possibilities that a multilingual platform like Eurosentiment has to offer. After an exhausting yet challenging time, the participants themselves voted the application by Merryweather as the best. This team was formed by Pablo Sanz, Eloy Cepa y Eloy Ramírez, each of whom got a Leap Motion as a price.
- CNL 2014: EUROSENTIMENT sponsored Fourth Workshop on Controlled Natural Language on 20-22nd August 2014. This event, co-located with COLING 2014, took place in Galway, Ireland. The workshop on controlled natural language (CNL) embraces all approaches based on natural language: simplified language, plain language, formalised language, processable language, fragments of language, phraseologies, conceptual authoring, language generation, and guided natural language interfaces.
- ISWC14: Gabriela Vulcu gave a hands-on session about EUROSENTIMENT Language Resources processing and use at the "Building the Multilingual Web of Data: A Hands-on tutorial”
- David Moreno. (2014). Design and development of a sentiment analysis service based on a Big Data infrastructure. Master thesis, Universidad Politécnica de Madrid, E.T.S.I. Telecomunicación.
- Keynote at ES³LOD 2014 - LREC 2014 Carlos Iglesias (Universidad Politécnica de Madrid, Spain), A linked data approach for describing sentiments and emotions
- J. Fernando Sánchez-Rada, Marcos Torres, Carlos A. Iglesias, Roberto Maestre & Raquel Peinado (2014). A Linked Data Approach to Sentiment and Emotion Analysis of Twitter in the Financial Domain. In Second International Workshop on Finance and Economics on the Semantic Web (FEOSW 2014)
- J. Fernando Sánchez-Rada & Carlos A. Iglesias (2013). Onyx: Describing Emotions on the Web of Data. In Proceedings of the First International Workshop on Emotion and Sentiment in Social and Expressive Media: approaches and perspectives from AI (ESSEM 2013), pages 71-82. Torino, Italy: CEUR-WS

- J. Fernando Sánchez-Rada, Carlos A. Iglesias & Ronald Gil. A Linked Data Approach for Annotating Sentiments and Emotions in Multimedia, submitted, 2015.

Liaisons and Cooperation with other projects

EUROSENTIMENT has maintained fruitful contacts with similar research projects covering the topic of sentiment analysis. Several meetings have been held with this projects and as a result, different ways of collaborations were performed at research and dissemination levels. The projects were exploitation of synergies and technical concertation took place were:

- **OpeNER²⁸**: OpeNER's main goal is to provide a set of ready to use tools to perform some natural language processing tasks, free and easy to adapt for Academia, Research and Small and Medium Enterprise to integrate them in their workflow. Four sentiment propagation lexicons from OPENER were added to the EUROSENTIMENT Language Resource Pool. In order to integrate it, NUIG defined a new type of language resource indicating that such type of resources have to be expressed in RDF and that they should be directly loaded to the SPARQL endpoint with no processing in between. The graphURIs for the new resources are as follows:
 - en: <<http://www.eurosentiment.eu/dataset/general/en/opener/0043/lexicon>>
 - es: <<http://www.eurosentiment.eu/dataset/general/es/opener/0044/lexicon>>
 - fr: <<http://www.eurosentiment.eu/dataset/general/fr/opener/0045/lexicon>>
 - it: <<http://www.eurosentiment.eu/dataset/general/it/opener/0046/lexicon>>
- **LIDER²⁹**: The project's mission is to provide the basis for the creation of a Linguistic Linked Data cloud that can support content analytics tasks of unstructured multilingual cross-media content. EUROSENTIMENT offered language resources transformed into Linguistic Linked Data (mainly sentiment corpora with NIF/Marl/Onyx-based transformation) and LLD-based language resources (mainly sentiment lexicons using lemon/Marl) to be used/integrated into LIDER. Vice versa we already used in EuroSentiment other LLD resources such as Babelnet and WordNet-RDF in generating the (multilingual) sentiment lexicons.
- **TrendMiner³⁰**: Its goal is to deliver. innovative, portable open-source real-time methods for cross-lingual mining and summarisation of large-scale stream media. EUROSENTIMENT and specially NUIG has maintained frequent contact with the technical managers of the project in order to explore synergies. Shared dissemination events have been performed.
- **MONET³¹**: The use of lemon (developed in MONNET) in EuroSentiment for the representation of sentiment lexicons has provided an innovative use case and has significantly increased the academic and industrial impact of the Monnet research outcomes. At the same time, the use of lemon provided a well-established and stable basis for the Linked Data strategy in the EuroSentiment project.

Coordination of standardisation efforts

EUROSENTIMENT project has promoted a new W3C Community Group to foster sentiment analysis research: Linked Data Models for Emotion and Sentiment Analysis Community Group¹. Topics addressed are:

Definition of a Linked Data based vocabulary for emotion and sentiment analysis.

Requirements beyond text-based analysis, i.e. emotion/sentiment analysis from images, video, social network analysis, etc.

²⁸www.opener-project.eu

²⁹www.lider-project.eu

³⁰<http://www.trendminer-project.eu>

³¹<http://www.monnet-project.eu>

Clarifying requirements and the need for consensus as e.g. systems currently use widely varying features for describing polarity values (1-5, -2/-1/0/1/2, positive/neutral/negative, good/very good etc.).

Marl and Onyx are vocabularies for emotion and sentiment analysis that can be taken as a starting point for discussion in the CG.

Fernando J. Sanchez from UPM is chairing this group, and several audio conferences have been held to organise the activity of the group.

The LIDER project on Linguistic & Multilingual Linked Data started a W3C Community Group which is also of relevance to EuroSentiment. In particular the current discussion in LIDER and EuroSentiment around licences for Linguistic Linked Data is of interest to both projects. <http://www.w3.org/community/ld4lt>: This group aims to consult with current and potential users of linguistic data to assemble user cases and requirements for Language Technology Applications that use Linked Data. The results will be used to guide future interoperability, research and development activities spanning the language technology and linked data domains, including via the OntoLex and BP-MLOD community groups.

Potential users are companies and public bodies involved in content management, the language services and localisation industry and other applications of content analytics techniques used in search, recommender systems, sentiment analysis and terminology management.

The group will engage with users through surveys and international road-mapping events organised by an EU-funded R&D consortium called LIDER. Contributions can be provided openly through engagement with the Community Group or anonymously if preferred. The anonymised outcomes of this exercise, in terms of use case and requirements priorities, technology gaps and interoperability roadblocks will be made publicly available via this community group by May 2014.

1.5 Project website and contact details

Project website

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

This website is the core activity for disseminating public information and results of the project. 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. The following screenshot 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.

The overall organisation of the web portal is:

- Home: includes links to all sections in the web portal. A quick view of the project is presented, plus links to EUROSENTIMENT LRP, social networks, a view of people involved and a view of the project news.
- The project: this section contains a general view of EUROSENTIMENT vision and goal.
- News: includes information about news about project progress, dissemination events, Sentiment Analysis, etc.
- User Group: this section includes information about the User Group of the project.
- Partners: a brief description of all project beneficiaries is depicted in this section.
- Results: this section includes public deliverables, dissemination material, public software and resources, publications and articles and events of the project.



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



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

EUROSENTIMENT is part of the LT-Innovate network, which gives the project a valuable visibility of the activities performed and the outcomes reached since the very beginning of the project.

Url	Description
http://eurosentiment.eu	Official web portal of the project
http://lt-innovate.eu/page/eu-projects	LT-Innovate is the Forum for Europe's Language Technology Industry
http://www.w3.org/community/ontol ex/wiki/Eurosentiment	W3C page for EUROSENTIMENT
http://eurosentiment.readthedocs.org /en/master	Read The Docs: EUROSENTIMENT project documentation
http://www.openeducationeuropa.eu/ es/node/124394	European Commission site for Open Education Europe

Web pages with presence of EUROSENTIMENT

Language Resource Pool and Demo sites

The demonstrator is available at http://portal.eurosentiment.eu/official_demo#

[Home](#)

A SIMPLE CONTENT PLAYGROUND BUILT ON EUROSENTIMENT

This playground allows one to paste content and see it analyzed for sentiment by example services built on top of the Eurosentiment platform. Notice: once you have the sentiment you can click on the main terms and see how they are interconnected into our linked data repository, often spanning several independent linguistic resources!

Try out the EuroSentiment Demo

Select the sites you want to analyze. If you select more than one site at the same time, the results will be displayed aggregated.

H Hotels 📁 Electronics

Hotels

tripadvisor.com evaluation dataset

booking.com evaluation dataset

Select a country

Select a city


Try to analyse a text!

Select a language

Select a service provider

Illustration 11: EUROSENTIMENT demo web site http://portal.eurosentiment.eu/official_demo

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


[Services](#) - [Resources](#) - [Profile](#) - [Subscription](#) - [Demos](#) - [Help](#) - [Administration](#)

Signed in as [paradigma](#) | [Sign out](#)

[Home](#)

Welcome to Eurosentiment Service and Resource Pool (ESRP)

This is a marketplace dedicated to services and resources useful in multilingual Sentiment Analysis.

Please choose from the following actions:

🗂️ Language Services

📚 Language Resources


 EUROSENTIMENT
[License](#)

Illustration 12: EUROSENTIMENT LRP management site <http://portal.eurosentiment.eu>

Contact details

Entity	Person	eMail
Paradigma Tecnológico	Daniel Molina	daniel.molina@paradigmatecnologico.com
Foundazione Bruno Kessler	Michele Mostarda	mostarda@fbk.eu
Expert System	Francesco Danza	fadanza@expertsystem.it
Sindicetech	Giovanni Tummarello	giovanni@sindicetech.com
Technical University of Madrid	Carlos A. Iglesias	cif@gsi.dit.upm.es
Insight Centre for Data Analytics, NUI Galway	Paul Buitelaar	paul.buitelaar@deri.org

2 Use and dissemination of foreground

EUROSENTIMENT includes a comprehensive dissemination plan, detailed in deliverable D8.7, in which an overall strategy is designed to provide visibility and facilitate the inclusion of resources outside the consortium by disseminating the results of the project. The dissemination activities during the project include: technology fairs, scientific conferences and publications, collaborating with other projects, providing contributions to selected standardisation bodies; planning the exploitation of the results, training external industrial and academic learners, and demonstrating the results to external parties.

The dissemination strategy of EUROSENTIMENT consists of three fundamental strategies: Awareness, Engagement and Demonstration.

The objective of Awareness actions is to spread the word about the EUROSENTIMENT project, inform possible users and clients, and reach potential collaborators, for academia (publications, academic presentation and events) and practitioners (professional presentations and events).

Along with the awareness actions, EUROSENTIMENT promotes Engagement actions through the Users Group set up in the project. One of the main goals of the User Group is to encourage members to participate in the actions conducted by the consortium and other related tasks that may reflect positively in the outcomes of the project.

Internet: EUROSENTIMENT is heavily linked to the Internet in many aspects. It aims at interconnecting parties to share resources, parties that develop their activities on the Internet to one extent or another, and that probably use data that has been created or transferred through the Internet.

For this reason, and many others, the promotion of EUROSENTIMENT during the project has been conducted on the Internet through different channels. Apart from project website and dissemination material detailed in the previous section, the following channels can be mentioned:

Social networks: There is an obvious shift towards the so-called social networks. Plenty of users are substituting their RSS feeds or traditional subscriptions with subscriptions to certain users, channels or aggregators in the handful of social networks they are present in.

For EUROSENTIMENT we are exploiting this new behaviour by sharing the public news related to the project through these channels:

- Twitter (<https://twitter.com/eurosentiment>)
- Scoop.it (<http://www.scoop.it/t/eurosentiment>)