

# Opinion Mining Tools for the Analysis and Adaptation of Corporate SFEs

Mari Carmen Rodríguez-Gancedo<sup>1</sup>, Javier Caminero<sup>1</sup>, Carlos Picazo<sup>2</sup>, Álvaro Hernández<sup>2</sup>

<sup>1</sup>Telefónica R&D, Madrid, Spain; <sup>2</sup>Polytechnic University of Madrid, Spain

Email: {mcrg, fjcg}@tid.es

## ABSTRACT

Nowadays, companies concerned about their corporate reputation should do a sentiment analysis, obtaining data from the different channels from where their customers can express their opinions and concerns about the portfolio of product and services of the company. For providing this feedback, customers are no longer so interested in traditional channels like call centers or written forms, and instead new channels like social networks or corporate blogs and forums are becoming the preferred choice. To analyze these data and consequently adapt company applications is crucial to have at your disposal tools able to perform this analysis. In this paper, novel applications that are being developed by Telefonica R&D in the framework of the RENDER project are described, showing their capability to take data from different sources like the social networks and effectively producing automatic reports.

## INTRODUCTION & MOTIVATION

For companies committed to provide their customers with the best user experience in their services and products, it is usually difficult to meet all the varying issues, topics and problems given the large amounts of heterogeneous and rapidly changing information to be taken into account in all communication channels, in which customers share their opinions, knowledge, advice and dissatisfaction.

Customer satisfaction is an important goal for any company: the way the user experiences products and services conditions their next decisions or actions. Telefónica uses different channels for this interaction with the customers to ease this bidirectional and detailed communication in which feedback to customers is provided either physically or virtually.

In Telefónica's customer portal, users are able to access all kind of information related to all products and services and also to make complaints about them, by means of mails to skilled operators. Moreover, in the portal it is possible to find open independent forums in which customers and users exchange information about its products and services.

Other more conventional channels are a *call center*, where

operators deal with concerns of customers related to products and services of the company, and also *paper surveys* can be filled out by the customers after visiting a shop.

But nowadays, more important than conventional channels is the presence in the social networks providing another alternative communication channel with the customers, or potential customers.

Records of interactions from these channels hold different types of information: requests, information about products and services, complaints and surveys, and also opinions, advices, remarks and knowledge about the different products. These records tend to be somewhat structured, however free text is often used to capture non-structured aspects of the communication with a customer, which is usually the norm in the contacts found in open forums.

Leveraging diversely expressed information created through these channels can be a mean to improve the exploitation of incoming information and to forecast future decisions, so efficiency and dynamism in Customer Relationships Management can be increased by means of the application of topic detection and opinion mining techniques.

It is necessary to address the growing needs of multinational enterprises to exploit the 'wisdom' of their large customer base (expressed as a vast array of opinions, viewpoints, suggestions and ideas) as a mean to optimally respond to market demands and developments. In RENDER project<sup>1</sup>, Telefónica is developing a novel approach to customer relationship management as a first step towards the implementation of a more comprehensive global enterprise crowd sourcing strategy.

An Opinion Mining tool will be set up to satisfy the detected needs of the final users, allowing the analysis of several data sources provided by communications channels with customers and potential customers.

In this sense, it is important to note that the data are provided by real sources, i.e. traditional sources but also the social network channels like Twitter, with real users that have real problems and that our main objective by the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

<sup>1</sup> <http://www.render-project.eu/>

moment is not to implement a working solution covering all the sources (because nowadays it would not be viable), but a proof of concept solution to show the capacity of the opinion mining techniques to extract interesting information from the chosen sources.

The stakeholders for these tools are the final users of Telefónica's corporate portal where the RENDER opinion mining capabilities are going to be deployed. The final users are mainly enterprises or departments in major corporations focused on:

- Social media marketing, providing mainly social media profiles of the company. The main functions are:
  - Community Management.
  - Evaluation of the impact of new products/services or advertisement campaigns.
  - Influence level of different online media.
  - Competition products analysis.
- Corporate Reputation, checking between several issues, the online reputation of the company. The online reputation is centered in:
  - Detection of negative opinions and enabling early correction.
  - Detection of brand perception and knowledge level in Internet.
  - Analysis of attributes associated with the brand.
- Business Intelligence, managing the extraction of knowledge through analysis of existing data in a company.

Managers do not have any tool to allow them to search for topics and track their evolution in the channels dominated by diversity. So, in RENDER project, we have the opportunity to research and work in a useful challenge.

### OPINION MINING TOOLKIT ARCHITECTURE

One of the goals of RENDER is to mine the communication means offered to customers as a way to capture the diversity of their opinions. The application of the methods and tools developed in RENDER will provide means to search and visualize certain topics and track their evolution. These methods will enrich the currently available evidences in corporate decision-making.

The Opinion Mining Tool consists of two main components, a *backend* for the generation of models and a *frontend* as user interface.

The architecture diagram showing the different components is displayed in Figure 1: the primary goal of the interface is carried out by the *diversity analysis component*, which can be invoked by other RENDER components in order to satisfy the full set of requirements for the opinion analysis workspace. For system demonstration purposes, a graphical user interface (GUI) is also included in the architecture.

The backend is driven by an embedded relational database to store the items and metadata. The search indexing and feature construction part is driven by a Miner infrastructure to ensure handling of datasets bigger than the available

main memory. The models themselves are stored in-memory, since they need to be re-trained often. This does not pose significant challenges for scalability, since the required storage for the model is usually proportional to the size of the feature space and not to the item count. The system also made a caching of the generated feature vectors for the items, enabling in this way an efficient re-training of the models when new labels are available.

The main controller of the system takes care of the models maintenance. Since real-time data updates and instant re-labeling are required without interrupting the requests from users, concurrent model training has been implemented. For instance, when the system receives a new label, it first checks whether the model is currently in the process of being re-trained. If it is so, the label is put into a queue, waiting for the next pass. In this way, adding label is a non-blocking process for the model training. When the model training is finished, the new re-trained model immediately replaces the old model, so that all subsequent classifications are executed using the new model. This ensures that the search operation is also non-blocking with regards to the model training.

Training times are usually in the range of a second for several thousand examples – just long enough to be a perceptible delay in blocking mode. In non-blocking mode, this delay is barely noticeable, since adding a label and issuing a new search query with classification are two separate interactions and the re-training usually finishes just before the user starts with the search and classification.

The architecture is presented in the following diagram:

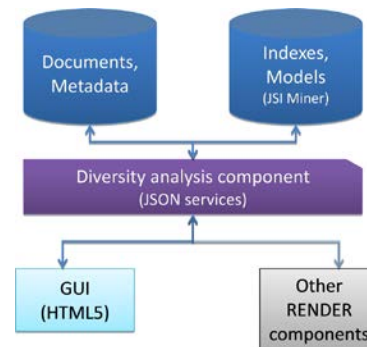


Figure 1: Architecture diagram

### OPINION MINING TOOLKIT FRONT-END

After discarding some other possibilities and having into account that the front-end of the application will have a web interface and it will require some data processing, HTML5 seems to be the best option, because it is a state-of-the-art, polyvalent, extended and essential language to build interfaces. For the definition of the graphical aspects, CSS3 will be used, since it is considered a standard for the design of web interfaces.

For the data processing of the front-end, JavaScript will be the choice, for its simplicity, robustness and multidevice compatibility. Considering that the data to be received by

the front-end will be simple (string and numeric values), Javascript is enough to handle these data and to properly format them for their presentation.

In addition to native Javascript, a jQuery library will be also added. In this way, the framework can offer a lot of options and advantages for the data processing, and above all, for handling the different elements of the webpage. One of its best features is the DOM management with CSS-like references, simplifying a lot the operations over the interface. Besides, it offers a native support for handling JSON objects, which are the output format provided by the backend. jQuery is widely used and it is compatible with the most popular browsers.

As a conclusion and taking into account the previously mentioned reasons, we could state that for the development of the prototype intended to visualize the information of RENDER, the most suitable technologies are HTML5, CSS3 and Javascript (including jQuery). According to the visualized information, the user will be able to take different decisions. The requirements of portability, compatibility and robustness will be also satisfied thanks to these technologies.

### Toolkit Features

In this section, the implemented functionalities are presented, describing the selected solutions able to deal with the topic search and evolution features.

The system interface will be used as a decision maker tool, so in the first place, a panel to filter the information to be treated is required. This panel will allow the application to generate reports according to the restrictions inserted by the user.

Another functionality of the panel will be to recover reports previously saved, offering in this way continuity to the task and facilitating the comparison among previous reports, avoiding the creation of repeated reports from the system.

Different filters can be used, making it possible to sort out the information either by 'user group', like online or offline, or by the source type of the information like twitter, email, call-center, or by the language of the information.

Another useful filter is able to sort out by topics the information to be searched by the user, being also able to deal with several topics at the same time, spreading or reducing in this way the work focus. Finally, the user can also sort out the data by date using the Time-frame thanks to the date-picker selector powered by JQuery.

In summary, it is possible to filter the information using different criteria:

- User Group: Online / Offline
- Source Type: Twitter, Call-Center, Survey, etc
- Language: English, Spanish, etc.
- Topic: The topics to be included in the report can be multiple, making it possible to extend or reduce the focus of work.

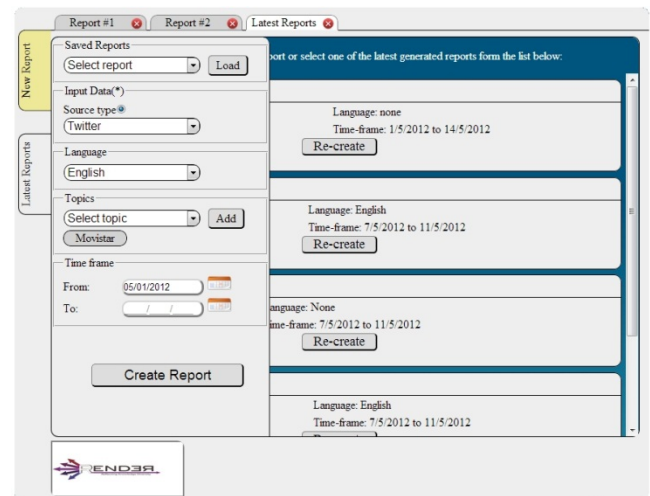
- Time Period: dates can be selected in the Time-frame with a "date-picker" provided by jQuery.

### Graphical Interface Overview

The interface consists of two main panels where the contents are shown (see Figure 2):

The left panel is devoted to set up the filters that make it possible the generation of new reports, with the additional functionality of displaying the already generated and saved reports. From a functional point of view, this panel is intended to introduce data and retrieve report generation requests from the user. Therefore, the panel is the graphic representation for the application input and it should be adaptable to possible future functionalities.

The right panel represents the output information generated by the application. This information can be of different nature if necessary. The application generated contents are stored whereas the user does not request them to be deleted.



**Figure 2: Front-End Graphical User Interface**

In the option filter panel, the different configurable features are distributed according to their nature. Each different feature is represented into a separated frame (shown in Figure 3):

- Saved reports: Through a list, it allows the selection of saved reports. In case of using this tool to load the report into the report information panel, it is not required to use the other filtering options.
- Input Data: It makes a distinction between two different categories, i.e. user group and source type. It shows a selection of the different possible options of each of these two types through a listing.
- Language: It allows the language selection for the filtering through a pull-down list with the different languages accepted by this application. It only allows the selection of an element from that list or none.
- Topics: A selector shows a list of the different possible topics. A side button allows that the selected topic was

used as a filter. It will represent through labels the different selected topics.

- Time-Frame: This filter consist of two filtering elements, one to set up the starting date and the other the end date. The search will be restricted to the timeframe between the starting and the ending dates.
- Create Report: This button executes the report generation according to the established preferences in the rest of the filters.

**Figure 3: Report filter option panel**

In the report information panel (see Figure 4) two areas are always present. The first one is where the tabs are found and below the content of the selected tab is located. Each tab shows an identifier that has been automatically produced in the generation process. The identifiers represent a content associated to the tab overall the current session. Each tab has a button represented by a red circle with a white multiplication sign inside that can be used to delete the tab and therefore its associated content (unless it had been previously saved).

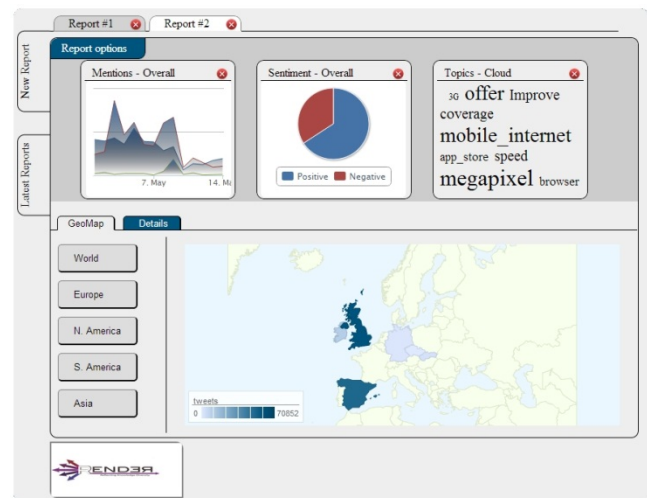
In this version of the front-end a time-growing graph shows the relationship between the data selected by the filter and the rest.

An abstract view of the proportion of the data evaluated as positive, and those evaluated as negative is also shown.

Another view is the ‘topics-cloud’ where the most relevant words appear in different sizes to represent the most recurrent ones over the rest.

## CONCLUSIONS

More and more companies adopt crowd sourcing to leverage the wisdom of their global customer base by using customer feedback for their product and services.



**Figure 4: Report information panel**

Many investments have been made in the deployment of Web 2.0-like approaches to customer relationship management, often, however, without the technology to manage the huge amounts of diversely expressed information generated in discussions forums, online customer portals, wikis, blogs, and media portals.

This lack of appropriate technology impacts on the return of investment, and leads to missed business opportunities from a product and service perspective.

Telefónica R&D has adapted RENDÉR’s concept and technology to successfully develop novel customer management solutions that are able to turn the opinions, viewpoints, and ideas of its customers into a competitive advantage.

## ACKNOWLEDGEMENTS

This work received funding from the European Commission’s Seventh Framework Program under grant agreement number 257790 (FP7-ICT-2009-5).

## REFERENCES

1. Ha He, Hugo Haas and David Orchard. “Web Services Architecture Usage Scenarios”. Technical report - World Wide Web Consortium (W3C), 2004.
2. M. Damova, K. Simov, Z. Tashev, and A. Kiryakov, "FactForge: Data Service or Diversity through Inferred Knowledge over LOD," in Proceedings of AIMSA’2012, Bulgaria, 2012.
3. A. Thalhammer, I. Toma, R. Hasan, E. Simperl, and D. Vrandecic, “How to Represent Knowledge Diversity”, 10th International Semantic Web Conference ISWC’11, Germany, 2011