

T-OMT: A Novel Opinion Mining Tool for Improving Global Customer Relationship Management

Mari Carmen Rodríguez-Gancedo¹, Javier Caminero¹, José Relaño¹, Carlos Picazo²

¹Telefónica R&D, Madrid, Spain; ²Polytechnic University of Madrid, Spain

Email: {mcrg, fjcjg}@tid.es

ABSTRACT

For IT companies, a quick reaction when a new product is being deployed into the market is crucial, so a very dynamic response is required, mainly analyzing the new channels used by their customers to express their satisfaction or concerns. This quick response can only be achieved if companies count with the support of powerful Opinion Mining Tools able to analyze the sentiment evolution over the new channels like social networks or corporate blogs and forums, then these have become the most common choice for obtaining a valuable feedback and traditional channels like filling out forms or call centers have been relegated to a secondary position. In the present paper, we present a novel and powerful tool for opinion mining developed by Telefonica R&D in the framework of the Render project, able to produce quick reports with user-friendly charts to analyze information coming from different sources and providing the company with a powerful weapon to be able to react to negative sentiment when releasing new products or services, and adapting them to the needs of the customers, that will actually be the final beneficiaries, then the products and services will be much more adapted to their current needs.

INTRODUCTION & MOTIVATION

Nowadays, companies involved in a global and dynamic market should be very careful with the needs of their actual customers to keep their loyalty. Therefore, new products to be launched must always take into account the current needs of the customers and the community related with them in order to be able to attract new customers with similar needs.

Nowadays, information changes very rapidly and social networks have become customer's preferred channel to share opinions, knowledge, advice and dissatisfaction, although for global companies like Telefonica with heterogeneous markets and customers, some customers still continue using the traditional channels like call centers, or personally visiting company stores, so an integration of the

different sources to express opinion must be taken into account for having an unbiased view of any situation.

Telefonica count with a corporate customer portal, where users are able to access all kind of information related with company products and services and also to make comments, suggestions or complaints about them, by means of emails that will be treated by very skilled operators. For information exchange about company's products and services, the customer portal also has independent blog and forums.

For any global company, it is important to address and anticipate user's needs, and that can only be achieved by analyzing the different interactions from the diverse channels where users express their concerns or suggestions about new products and services. Given the global scope of the company, a detailed analysis of the different markets where the company is present is a must. Specifically, Telefonica is present in 24 countries and sometimes product cannot be directly launched into the different countries and some adaptations are needed taken into account the idiosyncrasy of people.

Analyzing the available information can be quite challenging, then in some cases is somewhat structured, and in many others free text is usually the norm in the contacts found in open forums or social media.

RENDER PROJECT FRAMEWORK

The main motivation for Render project¹ starts from the unprecedented success of the Web for facilitating the publication, use and exchange of information on a global scale, on virtually every topic, and representing an amazing diversity of opinions, viewpoints, mindsets and backgrounds. Its design principles and core technological components have led to an unprecedented growth and mass collaboration. This trend is also finding increasing adoption in business environments. Nevertheless, the Web is also confronted with fundamental challenges with respect to the purposeful access, processing and management of these sheer amounts of information, whilst remaining true to its principles, and leveraging the diversity inherently unfolding through world wide scale collaboration.

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.

¹<http://www.render-project.eu/>

Render engages with these challenges by developing methods, techniques, software and data sets that leverages diversity as a crucial source of innovation and creativity, whilst providing enhanced support for feasibly managing data at very large scale, and for designing novel algorithms that reflect diversity in the ways information is selected, ranked, aggregated, presented and used. Render's information management solution scales not only to large amounts of data and users, but also to a plurality of points of views and opinions. This has been demonstrated through the usage of realistic data sources, including news streams covering over 5,000 sources worldwide with 100,000 items per day, (micro) blog streams adding up to more than a million posts per day, a full data stream from Wikipedia, and the Linked Open Data Cloud; through open source extensions to popular collaboration and communication platforms such as MediaWiki and Drupal; and through three high-profile case studies.

Render is helping to realize a world where information is acquired and shared in a fundamentally different manner than the consensual approach promoted by movements such as Web 2.0, and where communication and collaboration across the borders of social, cultural or professional communities are truly enabled via advanced Web technology, supporting one of the credos of European society: "United in diversity".

The Approach

Render provides a comprehensive conceptual framework and technological infrastructure for enabling, supporting, managing and exploiting information diversity in Web-based environments.

Diversity is a crucial source of innovation and adaptability. It ensures the availability of alternative approaches towards solving hard problems, and provides new perspectives and insights on known situations.

Equally important, embracing diversity in information management is essential for enhancing state-of-the-art technology in this field with novel paradigms, models, and methods and techniques for searching, selecting, ranking, aggregating, clustering and presenting information purposefully to users, thus alleviating critical aspects of information overload.

Render is developing concepts, methods, techniques and technology to:

- Collect and manage information sources rich in diversity so that this information is available in an effective form and can be processed efficiently in further steps. It is done by crawling, gathering, structuring and enriching various information sources with a great diversity basis, including sources relevant for the Render case studies. Render leverages very large amounts of content and metadata: news, blog and microblog streams, content and logs from Wikipedia, news archives, multimedia content and reader

comments, discussion forums and customer feedback databases from Telefonica, all together adding up to hundreds of millions of items, some even on a daily basis. This data can be managed by a highly scalable data management infrastructure, and enriched with machine-understandable descriptions and links referring to the Linked Open Data Cloud.

- Identify and extract the diversity embodied within the various information sources collected, and make the connections and references between different items and sources explicit.
- Represent and process diversely expressed information so as to explicate and conceptualize the results of the mining task, to enable the development of diversified information management algorithms and services. Render is developing novel, scalable techniques to reason upon opinions and viewpoints, and for diversity-aware information selection and ranking. Render also looks into proper means to make diversity information accessible to the end-user by providing sophisticated metaphors, interfaces and software tools to organize, display and visualize it. Raking algorithms take into account the viewpoints underlying different information items.
- Use diversity as integral concept of popular communication, collaboration and information sharing platforms, in the form of extensions to MediaWiki, Drupal, or Twitter. Render technology allows to explicitly link to items with a dissenting view, and thus to increase the diversity exposure of the wider Web audience.

T-OMT, TELEFONICA OPINION MINING TOOL Framework

Given the already mentioned difficulties and specific goals of IT companies for dealing with user's opinion and the mandatory quick responses times required, a new tool has been designed and implemented based on the Render project approach. This tool called T-OMT, which stands for Telefonica Opinion Mining Tool, constitutes one of the main use cases and outcomes of the project.

Large companies that provide several products and services need customer's feedback about their products and its impact on a global market with hundreds of millions of users. The customers publish their opinions about these products in different media (Twitter, email, surveys...) every day. The goal of T-OMT has been to collect massive datasets from these sources and use Render processes, giving the user analytic and useful information about the current opinion of the product. This tool uses Twitter as a primary opinion source and generates specific reports filtering the data by date, topics and language. These reports are showed also graphically, detailing the mentions of the topics, global sentiment, opinion geolocation, and details about the related topics.

T-OMT has been designed to satisfy the detected needs of the final users, allowing the analysis of several data sources provided by communications channels with customers and also potential customers.

T-OMT's stakeholders are the final users of Telefonica's corporate portal where the Render opinion mining capabilities are being deployed. The final users are mainly enterprises or departments both inside Telefonica group and also in another third party service provided designing applications for final customers focused mainly on these groups:

- 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 to allow an early correction.
 - Detection of brand perception and knowledge level in Internet.
 - Analysis of attributes associated to the brand.
- Business Intelligence, managing the extraction of knowledge through analysis of existing data in a company.

Architecture and Implementation Details

The Opinion Mining Tool consists of two main components, a backend for the data statistical analysis and a frontend as user interface.

The architecture of components is displayed in Figure 1, showing the different components for performing queries and displaying the right graphical charts as a result of the analysis.

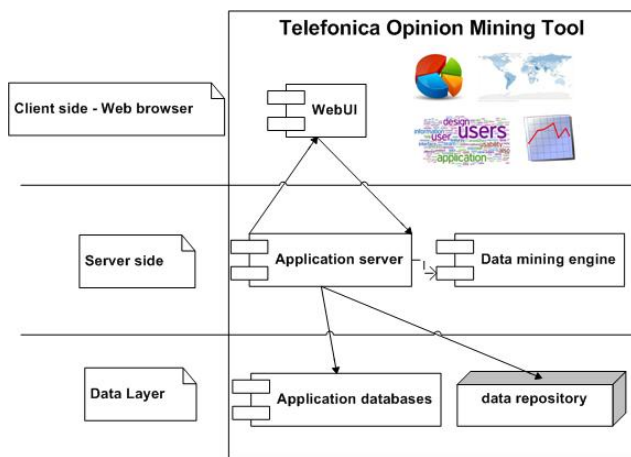


Figure 1: T-OMT Architecture diagram

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 is the chosen technology to build the interfaces and for the definition of the graphical aspects, CSS3 is used, since it is considered a standard for the design of web interfaces and finally, for the data processing of the front-end, JavaScript, adding a jQuery library has been the choice, for its simplicity, robustness and multidevice compatibility and able to deal with JSON objects, that are the outputs provided by the back-end. Additional software components and plugins are used, mainly for graphical widgets. GeoMap and line/pie charts are powered by Google charts libraries and their online technologies.

To be consistent with the main libraries and functionalities jQuery-ui is the library included for graphical web building. Its web browser compatibility, cross-device design and robustness make this plugin highly properly for this tool.

The server side is developed in Java6 and deployed in an Apache-Tomcat server. A ReSTful API implementing the real-time required functionality over AJAX methods is used as interface between the front-end and the server.

The application databases use MongoDB technology. It stores user information and processed reports saved by the user to be loaded without reprocessing the data in order to reduce response times and enhancing backend's scalability.

Main Toolkit Features

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 and also recovering previously saved reports, avoiding thus, the creation of repeated reports from the system.

The different filters make it possible to make a report generation according to the following criteria:

- User Group: Online / Offline
- Source Type: Twitter, Call-Center, Survey ...
- Country: Spain, Germany, UK ...
- Topic: individual or multiple topics can be added
- Time Period: a timeframe can be selected

Generated Reports Details

The generated report is shown through two main panels where all the information is shown (see Figures 2 and 3):

The first panel (Figure 2) is the "GeoMap" panel, showing a world map detailing the number of topic found in the different countries. On the top area of the panel three different charts are also displayed and that can provide with additional information by clicking on them:

- A 'mention chart', showing the number of mentions of a topic over the time (different topics can be displayed

simultaneously). In a new product launch, for example, this chart can help quickly visualize the evolution of the mentions of the product over the time. In this chart, the sentiment is not taken into account.

- An ‘overall sentiment’ chart, displaying graphically the positive and negative sentiment.
- A ‘geo-distribution’ chart, showing the geographical proportion of appearances of the selected topics

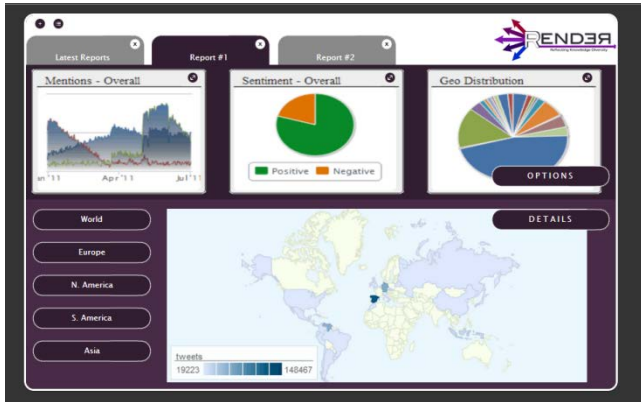


Figure 2: ‘GeoMap’ Graphical Output Panel

The second panel (Figure 3) shows details of the most frequent topics, and at the top the same three charts displayed on the previous panel are also shown.

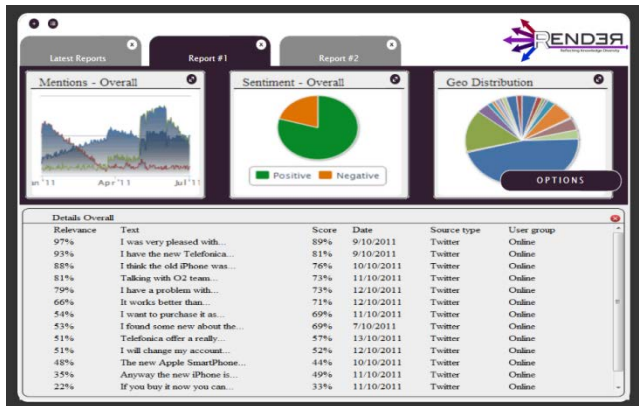


Figure 3: Details Output Panel

Additionally to the showed charts, some more could be interesting to be added:

- A ‘sentiment evolution chart’, showing the evolution of the sentiment of a topic over the time (different topics can be displayed simultaneously).
- A ‘detailed geo-distribution chart’, analyzing how the appearances grow up or decrease in a certain geographical area.

- A ‘related-topic word cloud’, showing the most frequent appearances in a bigger size and the less frequent in a smaller one.

CONCLUSIONS

In the framework of Customer Relationship Management, it is nowadays mandatory to have proper tools to be able to analyze and give proper quick responses to the needs of customers in a global and competitive world, where opinions and sentiment evolve very fast and new communication channels like the social networks are quickly replacing traditional channels like call centers.

In response to these needs, Render project has successfully provide a novel approach, and focused its efforts mainly in the new interaction channels, but without forgetting the traditional ones, then for global companies, in certain countries, traditional channels may still play an important role.

A new Opinion Mining Tool has been presented by Telefonica as a project use case and one of its major outcomes. This tool is able to generate reports analyzing massive information coming from different channels, including social networks and it shows valuable charts as a result of a deep analysis of the data and providing details as the evolution of the sentiment over the time and its geographical distribution, a useful and powerful tool particularly useful for global companies like Telefonica.

Thanks to the results provided by the tool, return of investment can be enhanced, missed business opportunities will be avoided and opinions, viewpoints, and customers ideas can be turned 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. T. O’Reilly. “What is Web 2.0?” Design Patterns and Business Models for the Next Generation of Software”, Sep. 2005. <http://oreilly.com/web2/archive/what-is-web-20.html>
2. Ha He, Hugo Haas and David Orchard. “Web Services Architecture Usage Scenarios”. Technical report - World Wide Web Consortium (W3C), 2004.
3. 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.
4. A. Thalhammer, I. Toma, R. Hasan, E. Simperl, and D. Vrandeic, “How to Represent Knowledge Diversity”, 10th International Semantic Web Conference ISWC’11, Germany, 2011.