

SEMO: A Framework for Customer Social Networks  
Analysis based on Semantics.



Ángel García-Crespo  
Phone: +34 91 624 9417  
Fax: +34 91 624 9129  
angel.garcia@uc3m.es  
Computer Science Department  
Universidad Carlos III de Madrid  
Avda. de la Universidad 30, Leganés, Madrid, Spain

Ricardo Colomo-Palacios  
(Corresponding Author)  
Phone: +34 91 624 5958  
Fax: +34 91 624 9129  
ricardo.colomo@uc3m.es  
Computer Science Department  
Universidad Carlos III de Madrid  
Avda. de la Universidad 30, Leganés, Madrid, Spain

Juan Miguel Gómez-Berbís  
Phone: +34 91 624 5958  
Fax: +34 91 624 9129  
juanmiguel.gomez@uc3m.es  
Computer Science Department  
Universidad Carlos III de Madrid  
Avda. de la Universidad 30, Leganés, Madrid, Spain

Belén Ruiz-Mezcua  
Phone: +34 91 624 9968  
Fax: +34 91 624 9129  
belen.ruiz@uc3m.es  
Computer Science Department  
Universidad Carlos III de Madrid  
Avda. de la Universidad 30, Leganés, Madrid, Spain

#### Acknowledgements

This work is supported by the Spanish Ministry of Industry, Tourism, and Commerce under the EUREKA project SITIO (TSI-020400-2009-148), SONAR2 (TSI-020100-2008-665), INNOVA 3.0 (TSI-020100-2009-612) and GO2 (TSI-020400-2009-127).

SEMO: A Framework for Customer Social Networks Analysis based on Semantics.

Abstract

The increasing importance of the Internet in most domains has brought about a paradigm change in consumer relations. The influence of Social Networks has entered the Customer Relationship Management domain under the coined term CRM 2.0. In this context, the need to understand and classify the interactions of customers by means of new platforms has emerged as a challenge for both researchers and professionals world-wide. This is the perfect scenario for the use of SEMO, a platform for Customer Social Networks Analysis based on Semantics and emotion mining. The platform benefits from both semantic annotation and classification and text analysis, relying on techniques from the Natural Language Processing domain. The results of the evaluation of the experimental implementation of SEMO reveal a promising and viable platform from a technical perspective.

*Keywords:* Social networks, Customer Relationship Management, Semantics, Emotions, Natural Language Processing.

---

## **1. INTRODUCTION**

The dramatic spread of the Internet in society has substantially changed the forms of communication, entertainment, knowledge acquisition and consumption. There is a constant increase in the number of people who consult the Internet as a medium for answering their queries, and who use the Internet as a new form of communication. A shift in the Web content consumer-producer paradigm is making the Web a means for conversation, cooperation and mass empowerment. Emerging killer applications combine sharing information and social dimension, undermining the very principles on which content has relied for decades, namely information asymmetry and top-down content delivery. Social interactions have recently found an exceptional vehicle in the recent breed of user generated content aware technologies encompassed by the "Web 2.0" buzzword (O'Reilly, 2005). These technologies have forced some organizations and initiatives to make an adoption which enables them to meet their business challenges and obtain a competitive advantage. But mostly, they have provided a platform to foster social critical mass,

particularly due to the amount of metadata they have generated to provide tags, picture sharing environments, social bookmarks, blogs and music preferences. According to O'Reilly (2005), a fundamental principle of Web 2.0 is that users add value by generating content through these applications, resulting in network effects among the community of users.

According to a study by McKinsey consultants (McKinsey, 2007) where 2,847 executives were interviewed, respondents informed that Web 2.0 technologies are strategic and that they plan to increase investments in those technologies. Moreover, they stated that they are using Web 2.0 technologies to communicate with customers and business partners, and to encourage collaboration inside the company. More precisely, executives' blogs are also frequently mentioned as a channel for communicating with customers and, in some cases, as a channel for airing criticism.

This new web offers limitless opportunities for companies to engage their customers (Eikelmann et al., 2008). For example, the Southwest airline blog has received more than 6,300 comments since it started in April 2006 in response to little more than 250 posts. Rather than ignoring or fearing criticism or opinion generated in Web 2.0 forums, companies should seize Web 2.0 tools to respond and gain competitive advantage (Eikelmann et al., 2008). The usefulness of this novel web structure has also been demonstrated in the development of customer management tools. Studies by Forrester consultancies confirm that Customer Relationship Management (CRM) applications have adopted the importance of Web 2.0 in CRM environments. These studies indicate that innovative businesses are using Web 2.0 tools to: collaborate on sales, customer service, and marketing collateral; connect social networking tools into a business environment to help identify leads better; and utilize community networks to better provide service to customers (Marson, 2008), and they also show that CRM professionals must find innovative ways to engage with emerging "social consumers" (Band, 2008).

Thus, the SEMO framework is proposed, in this environment, in which the importance of Web 2.0 is steadily rising in the domain of customer relations. The objective of this framework is to exploit the advantages of the Social Web by means of the use of semantic technologies, in relation to Customer Relationship Management.

The remainder of the paper is structured as follows: Section 2 provides an overview of the state of the art in semantics, social networks and the use of emotions analysis in CRM Systems. Section 3 presents SEMO, the solution proposed in

this paper, including features and architecture. Section 4 illustrates a use scenario for SEMO. Section 5 presents experimental set up. Finally, Section 6 discusses the conclusions drawn and future work to be done

## **2. STATE OF THE ART**

Understanding the needs of customers and offering value-added services are recognized as factors that determine the success or failure of companies (King and Burgess, 2008). The purpose of CRM is to identify, acquire, serve, and retain profitable customers by interacting with them in an integrated way across a range of communication channels (Mahdavi *et al.*, 2008). The origins of what is today known as CRM stem from the Relationship Marketing field (Levitt, 1983). Relationship Marketing is an integrated effort to identify, build up and maintain a network with individual customers for the mutual benefit of both sides (Shani and Chalasani, 1992, p. 34).

The increasing capacities of technology have triggered the diversification of CRM, thereby extending its philosophy. Thus, the isolated approach of dealing with customer relationships has evolved into a philosophy aimed at creating an integrated view of the customer throughout the enterprise, where legacy systems were connected and which today provides the building blocks for comprehensive integrated CRM systems (Bueren *et al.*, 2005). CRM applications take full advantage of technological innovations with their ability to collect and analyze data regarding customer patterns, interpret customer behavior, develop predictive models, respond with timely and effective customized communications, and deliver product and service value to individual customers (Chen and Popovich, 2003).

In order to build longstanding and worthwhile relationships with customers, it is necessary to serve each customer in his preferred way and channel (Davenport *et al.*, 2001). The most common forms of customer interaction are the following (Chang *et al.*, 2008): (1) Face-to-face interaction with retail personnel; (2) Calls to customer service centers and conversations with customer service representatives; (3) Comments on company websites; and (4) Opinions expressed through e-mail. From the technical viewpoint, the infrastructures of CRM solutions are focused on Internet technology, among other support structures (Chen and Popovich, 2003). In this Internet scenario, Web 2.0 has turbocharged the whole notion of 'word-of-mouth', circumventing traditional marketing by letting individuals talk directly to each other about their passions, their

buying preferences and their pet peeves (Eikermann *et al.*, 2008). Thus, according to corporate studies, there is a continuously increasing volume of commercial CRM tools which incorporate and stimulate the use of social networks for global client management (Band, 2008; Maoz, 2008; Marson, 2008;). For example, according to O'Reilly (2007), Salesforce.com demonstrates how the web can be used to deliver software as a service, in enterprise scale applications such as CRM. Originating from this combination of technologies and philosophies, a number of authors have begun to employ the term "CRM 2.0", for example (Stone, 2009).

In addition, the study of emotions in customer behaviour has an established tradition (Huang, 2001) with important and numerous contributions to the literature (Bagozzi *et al.*, 1999; Gountas and Gountas, 2007; Schoefer and Diamantopoulos, 2008; Van Dolen *et al.*, 2004; Zeelenberg and Pieters, 2004). Philosophers and psychologists have extensively studied emotions, outlining diverse theories regarding their composition and typologies. As a result of these complex studies, there are numerous definitions of emotion. For the objective of this work, whose aim is to establish the emotions of customers within Web 2.0, the definition provided by Izard (1977) will be adopted. For this author, emotion is composed of three aspects: a) the experience or conscious feeling of emotion, b) the processes that occur in the brain and nervous system, and c) the observable extensible patterns of emotion. For our purposes, written patterns will be used. We can observe, in the domain of customer emotions, the study by Laros and Steenkamp (2005), which proposes classifying emotions into three levels. The first level represents the balance of emotions, that is, positive and negative effect. The next level is considered as the basic emotion level, and the lowest subordinate level consists of groups of individual emotions that form a category named after the most typical emotion of that category. Figure 1 shows the hierarchy of consumer emotions.

INSERT FIG 1 HERE

Given SEMO's objective, it is necessary to outline the key concept of an ontology. Ontologies (Fensel, 2002) are the technological cornerstones of the Semantic Web because they provide structured vocabularies that describe a formal specification of a shared conceptualization. The term "Semantic Web" was coined by Berners-Lee, Hendler and Lassila (2001) to describe the evolution from a document-based web

towards a new paradigm that includes data and information for computers to manipulate. In this application environment, SEMO requires two types of ontologies: Firstly, an ontology which allows the classification of emotions, in particular, customer emotions; Secondly, an ontology which models the different aspects related to CRM.

In relation to ontologies of emotions, there are diverse valid research initiatives in distinct application fields (e.g. Francisco *et al.*, 2007; López *et al.*, 2008; Mathieu, 2005) and, additionally, there is a W3C Emotion Markup Language Incubator Group, working on the definition of valid representations of those aspects of emotional states that appear to be relevant for a number of use cases in emotion scenarios. Undoubtedly, with the objective of taking advantage of the possibilities of combining current ontologies of emotion and the hierarchy of consumer emotions identified by Laros and Steenkamp (2005), the work of García-Crespo *et al.* (2008) proposes an ontology adapted to customer emotions.

In the second place, the aim is to annotate all the elements in the framework, taking a CRM ontology as a base. Within the CRM field in the last few years, ontologies have been constructed for Customer Complaint Management (Jarrar, 2008) as well as efforts focusing on employees' point of view (Van Damme *et al.*, 2007) or from a universal viewpoint (Magro and Goy, 2008) attempting to combine the set of problems and foci of CRM strategies.

Another of the elements required to achieve the objectives of the current work is to carry out an analysis of texts for the classification of the various opinions available in Web 2.0 environments. The analysis of texts in a CRM environment has a longstanding field of studies associated with it as well as available tools (Chang *et al.*, 2008; Kazmer, Burnett and Dickey, 2007; Linoff and Berry, 2002). Within the text analysis literature, many researchers have devoted themselves to developing techniques for exploring, extracting, mining, and aggregating opinions and sentiments. This research domain has become known as Sentiment Analysis or Opinion Mining. For a review of this research field see Takashi and Manabu (2006). SEMO builds on some of the benefits of previous works (Danisman and Alpkocak, 2008; Strapparava and Mihalcea, 2008) and presents a novel solution in which authors use the Open Social Network Dataset and a hierarchy of consumer emotions to face the challenges of the interactive characteristics of the Social Web and the Semantic Web. This promising new solution identifies not only the valence of the emotion, but the basic emotion of the user, providing a significant

contribution for the current literature. Several works have been devoted to opinion mining in the web, but the novel contribution of SEMO is that it looks for emotions and classifies them into basic emotions.

### **3. SEMO**

In this section, we will define the SEMO approach, which is based on extracting features from Social Networks and relating them to Consumer Emotions which will be the basis for a CRM-based strategy in order to maximize customer satisfaction. In the following, we will discuss the bridge between Social Networks and structured semantics, presenting a structured mechanism which acts as the theoretical basis of the framework, the Open Social Network Dataset (OSND), and finally, proposing an architecture for SEMO.

#### *3.1. The Open Social Network Dataset (OSND)*

OpenSocial is an application programming interface to build social applications across the Web, in other words, a common set of APIs for social applications across multiple websites. With standard JavaScript and HTML, developers can create applications that access a social network's friends and update feeds (OpenSocial, 2008).

OpenSocial is currently being developed by Google in conjunction with members of the web community. The ultimate goal is for any social website to be able to implement the APIs and host 3rd party social applications. There are many websites implementing OpenSocial, including Engage.com, Friendster, hi5, Hyves, imeem, LinkedIn, MySpace, Bebo, Ning, Oracle, orkut, Plaxo, Salesforce.com, Six Apart, Tianji, Viadeo, and XING (OpenSocial, 2008).

OpenSocial is not a social network itself; rather it is a set of three common APIs that allow developers to access the following core functions and information on social networks:

- People and Friends data API: allows client applications to view and update People Profiles and Friend relationships using AtomPub GData APIs with a Google data schema. These applications can request a list of a user's Friends and query the content in an existing Profile.
- Activities data API: allows client applications to view and publish "actions" in the OpenSocial platform using AtomPub GData APIs with a Google data schema. This API allows the creation of new entries, editing or deletion

of existing entries, and the capability to view lists of entries.

- Persistence data API: allows client applications to view and update key/value content using AtomPub GData APIs with a Google data schema. Applications can edit or delete content for an existing application, user, or gadget instance, and query the content in an existing feed.

OSND is a lightweight ontology based on the information extracted from the Open Social network source. It is constructed using the information from a set of social networks, obtaining a structured version of user profiles, getting a list of user friends per user and following their friend connections in order to get detailed profiles. We can determine which people are friends of a user and how important or close they are.

In our particular scenario, OSND is focused on "opinions" or concepts related to products. However, another fundamental feature is the possibility of tagging the content in all these applications. Tags are freely chosen keywords describing a particular resource. They offer a simple way of retrieving content (e.g. retrieval of my interesting communities in LinkedIn with the tag Semantics). These tag sets and their assignments to objects are envisaged as subjective conceptualizations, being potentially aggregated to a flat bottom-up categorization or folksonomy. Folksonomies are said to be an interesting emergent attempt for information retrieval (Shadbolt *et al.*, 2006) but serve different purposes for ontologies, such as attempts to define parts of the data world more carefully and to allow mappings and interactions between data held in different formats. Hence, ontologies are defined through a careful, explicit process that attempts to remove ambiguity, whereas the definition of a tag is a loose and implicit process where ambiguity might well remain. Finally, the inferential process applied to ontologies is logic-based and uses operations such as "join". The inferential process used on tags is statistical in nature and employs techniques such as clustering.

Nevertheless, in the past few years, there have been successful attempts at enriching tags with hierarchical relations (Schmitz, 2006) and the creation of faceted ontologies (Heyman and Garcia-Molina, 2006). Furthermore, Giunchiglia, Marchese and Zaihrayeu (2007) describe the theory of formal classification, where labels are translated to a propositional concept language. Each node is associated



with a normal formula that describes the content of the node, capturing the knowledge that implicitly exists within simple classification hierarchies.

Hence, we can build an application that easily works across all the OpenSocial partners, and people who have an account in any social network supporting OpenSocial can use our solution for email ranking and filtering, taking advantage of the information in his/her social network.

### 3.2. Building Up the OSND

Building the OSND is based on collaborative data filtering and rating in which we follow an integrated approach for combining three types of techniques to improve its construction from the tag sets gathered from the aforementioned Web 2.0 social networks such as Engage.com, Friendster, hi5, Hyves, imeem, LinkedIn, MySpace, Bebo, Ning, Oracle, Orkut, Plaxo, Salesforce.com, Six Apart, Tianji, Viadeo, and XING.

The three techniques we apply are as follows:

- Applying the Vector Space Model: The Vector Space Model (Salton et al., 1975) is an algebraic model used for information filtering, information retrieval, indexing and relevancy rankings. It represents natural language documents (or any objects, in general) in a formal manner through the use of vectors (of identifiers, such as, for example, index terms) in a multi-dimensional linear space. Documents are represented as vectors of index terms (keywords). The set of terms is a predefined collection of terms; for example the set of all unique words occurring in the document corpus. Relevancy rankings of documents in a keyword search can be calculated, using the assumptions of document similarities theory, by comparing the deviation of angles between each document vector and the original query vector where the query is represented as the same kind of vector as the documents.
- Using Latent Semantic Analysis (LSA) (Deerwester et al., 1990) for analyzing relationships between a set of documents and the terms they contain by producing a set of concepts related to the documents and terms. LSA uses a term-document matrix which describes the occurrences of terms in documents. A typical example of the weighting of the elements of the matrix is the TF-IDF (Term Frequency-Inverse Document Frequency): the element of the matrix is proportional to the number of times the

terms appear in each document, where rare terms are up-weighted to reflect their relative importance.

- Validating the set of terms pertaining to the OSND with online lexical resources, such as WorldNet. Dictionaries are generally considered as a valuable and reliable source containing information about the relationships among terms (e.g. synonyms). In addition, WorldNet can add conceptual meaning to the tags and there is an RDF transcript available.

Fundamentally, the coupling of the three techniques firmly based on Information Retrieval literature provides a two-pronged approach to retrieve and accurate OSND: selecting and extracting the most accurate tags from the pool of Web 2.0 applications user generated content and creating a "metadata cloud" which encapsulates the subjective meaning and intention the user conveyed through the tagging process. The OSND, hence, represents a valuable piece of knowledge which could be envisaged as a projection of the subjective mindset of the user.

### 3.3. Architecture

In this section, we will show the SEMO architecture by introducing a number of software components that use the technologies described in previous sections. Given that a software architecture is the set of connections, components and interfaces in which the software system is organized, we will elaborate on how the architecture supports a number of functionalities from that viewpoint.

The SEMO architecture is composed of several self-contained software modules or subsystems as is discussed in the following:

- Open Social Network Dataset Crawler: The OSND crawler was explained in previous sections. It finds, classifies and generates a lightweight ontology, the OSND where-by hoarding semi-structured information and processing it. The OSND is the entry point of a huge amount of information which can be dispersed and difficult to find in Social Networks since these structures are not fully opened and work mostly as "Chinese walls" regarding data integration. An implementation of this crawler can be found in the work of Rivera *et al.*, (2008). The OSND crawler output is sent to the Sentiment Analysis engine. Fundamentally, the OSND Crawler works based on the algorithm described in section 3.2, which means building

up the OSND. The core algorithm working behind the Dataset crawler is the combination of the three techniques: Applying the Vector Space Model, using the Latent Semantic Analysis (LSA) and validating the terms through a thesaurus. The goal of this combination is twofold. On the one hand, it is aimed at retrieving all necessary information or, at least, the information which is highly tuned with the domain ontology (the Customer Emotion Ontology, which will be explained in the following architectural component, which is based on the hierarchy of consumers' emotions shown in section 2. On the other hand, since we are working on a close domain where emotions can be summarized with a particular and finite set of language expressions, the core strategy is to identify those terms from the Social Networks pool of data, structure it and build up the OSND as a lightweight ontology. The efficiency of the OSND has been validated by the evaluation success we achieved, as we show in section 5.1.

- Sentiment Analysis Engine: This component uses Sentiment Analysis techniques over OSND data. For that, semantics play a key role and are exploited as follows. At this stage, the OSND has a number of concepts related to the domain from the set of resources where the OSND has been pooling out data i.e. Social Networks. This harvested data must be checked, validated and put into context for a proper knowledge base where Customer, Emotions and Products are related and, precisely, their relationships bring -added value to the system. Hence, the core knowledge base of the SEMO framework is a Customer Emotion Ontology (CEO). We have analyzed, designed and, finally, implemented the Customer Emotion Ontology to populate instances of this ontology associating the concepts of the hierarchy with the sets of structured data. The ontology is based on the hierarchy of emotions shown in section 2, but it is implemented with the Ontology Web Language (OWL), a family of knowledge representation languages for authoring ontologies, endorsed by the World Wide Web Consortium. In particular, since we will be using Description Logics as the underlying framework to reason with, we implemented the ontology in its OWL-DL flavor. Once this step is done, the Sentiment Analysis Engine applies sentiment criteria to relate the OSND data to the CEO terms, populating the ontology instances, in a two-pronged process: first the concepts are populated and then the relationships. For example, if user Mateusz

Heinz is criticizing the Blackberry Bold in his Facebook, his OSND will reflect this opinion. The Sentiment Analysis Engine will create an instance of the Customer concept of the ontology, "Mateusz Heinz", a number of Emotion concept instances and, eventually, a Product instance "Blackberry Bold" related to both the Customer instance and the Emotions, delivering the populated ontology. Hence, the output of the Sentiment Analysis Engine is the populated CEO.

- Product Feedback Manager (PFM): Once the ontology is populated, we have a knowledge-base where we use inference based on the underlying logical formalism of the ontology. The PFM uses a combination of inference based on Description Logics (DL) (Baader et al., 2003), a family of knowledge representation languages which can be used to represent the concept definitions of an application domain in a structured and formally well-understood way in classical querying to structured data structures.
- Ontology Repository: This component deals with the ontology storage. Ontology Repositories are software components that deal with scaling, loading and inferencing of real ontologies. Ontology Repositories extensive performance figures, which we summarize in the table below, have been recently in SEKT EU Project: Deliverable D2.6.3. They provide a comparison of the tools in terms of scalability, speed, and inference capabilities.

**INSERT TABLE 1 HERE**

Hence, we have decided to use KAON2 as the SEMO architecture ontology repository given its excellent performance in terms of scalability and inference, and also because its API is very accurate and easy to learn and is thus reliable for developers

**INSERT FIGURE 2 HERE**

The actual dynamics of the architecture are as follows. The OSND provides a structured information dataset in the form of a "lightweight ontology" which works as the input of the Sentiment Analysis Engine which will classify most of the terms, assigning them an "emotional" (hence, sentiment-based) category, yielding a populated ontology of consumer emotions and products. This OSND works with fixed channels (sites) and

mines information about given users that participate in the process, granting access to their profiles (if they are not open) or using open content (often anonymous). In fact, the mining must be based on a particular set of fixed channels where both data quality and availability are ensured. Hence the mining process starts when these channels provide a number of datasets from which a subset of knowledge-driven statements can be extracted or directly inferred. In addition, if the channels are not providing heterogeneous but interest-savvy data streams from the user perspective, these channels will have to be replaced by those providing value-added information.

The correspondence of products and emotions is a knowledge-intensive tool where a number of value-added relationships can be extracted, which is precisely the role of the Product Feedback Manager.

It is of the utmost importance to stress the role and relevance of semantics in the SEMO framework, particularly as the cornerstone of the PFM, the core of the SEMO framework. We consider emotional aspects as the primary aspect for semantic matchmaking: if a particular Consumer Emotion Ontology (CEO) instance does not provide any valuable relationship between customers and emotions for a particular product, then it is not usable and other, non-functional aspects are irrelevant. We define an emotion-based model, in which a particular set of emotions of a Product  $P$  denotes a sequence of emotions  $\Sigma = (e_0, \dots, e_N)$ . Analogously, we understand a particular solution of a Product Feedback  $\Pi = (s_0, \dots, s_N)$  as a sequence of states from the initial state into a state of the world wherein the objective of implementing the feedback is solved. A functional description  $P$  formally describes the possible emotions a particular product triggers after its commercial release.

Hence, we define  $P$  over a signature  $\beta$ , and use ontologies  $\Omega$ , as the background knowledge.  $P$  consists of a set of variables, a pre-condition  $\mu^{\text{pre}}$  that constrains the possible emotions and a post-condition  $\mu^{\text{post}}$  that constrains the number of emotions. The formal meaning of  $P$  is logically described by the implication semantics between the precondition and the post-condition.

To sum up, in order to deal with emotion descriptions in terms of model-theoretic semantics, we present this as a DL formula  $\mu^P$  of the form  $\mu^{\text{pre}} \Rightarrow \mu^{\text{post}}$ . Then,  $\Pi \models P$  is given if and only if every state of Product Feedback is represented by a  $\beta$ -interpretation of that which is a model of  $\mu^P$ . That means, in a nutshell, that a set of Product Feedback states can address emotions of particular Product  $P$  space.

INSERT FIGURE 3 HERE

The Figure 3 shows the logical structure of the system. A three-layered architecture was selected because of its adaptability, flexibility and reusability (Eckerson, 1995). The system was developed in an incremental and evolutionary manner that needed those main characteristics to provide successful fulfillment. A three-layered architecture also offers the advantage of easing the localization of errors, since it avoids the transfer of errors between layers.

In the upper layer, the Presentation Layer, both a Graphical User Interface (GUI) Web Front and a Web Services access component were included. The Web Service component provides the extra functionality of communication with loosely-coupled external systems, using programmatic interfaces, which benefits the interoperability of SEMO as a whole

The Application Layer is the core software layer of the SEMO architecture. It encapsulates the Business Logic of the architecture through a number of loosely-coupled software components such as the OSND Crawler, the Sentiment Analysis Engine and the Product Feedback Manager, which have been previously detailed in this section.

Finally, the Data Layer is the logical semantic storage backbone, where both the Customer Emotion Ontology (CEO) schema and its populated instances are stored in the KAON2 ontology repository which we implemented. The Ontology Repository provides the four basic create, read, update and delete (CRUD) functions for persistent storage together with reasoning and querying functionalities, with the RDF storage performance and figures shown in Table 1.

In this section, we have provided a detailed description of the SEMO architecture from a logical-functional, layer-oriented and architecture dynamics perspective. In the following section, we elaborate on a use case scenario showing the advantages of the SEMO approach as well as on the benefits of using semantics as its backbone technology.

#### **4. USE CASE**

To explain the realization of SEMO in a functional environment, as referred to in Section 3, a use case will be included. The manufacturing company of the mobile phones GoingWithU would like to launch a new model. The new product under consideration is DJPhone, a mobile phone with capacities for sequencing and recording music. The company

has a set of Beta Testers available, and the aim is to include them in the co-creation process of the new model in the final part of the design. The organization sends the application to the set of customers and encourages them to participate as a group by means of the notepad social tool and testimonies regarding their experience of using the mobile phone. It is assumed that the users make use of the social networks such as Orkut or LinkedIn to carry out the interaction.

Once the testing period of DJPhone is complete, SEMO is employed by the GoingWithU team to carry out an emotional categorization of the commentary of the Beta Testers, following the classification of Laros and Steenkamp (Laros and Steenkamp, 2005). Thus, by means of OpenSocial, SEMO accesses the data which the customers have introduced into the Social Networks mentioned above. The analysis results are generated using Natural Language Processing technologies, being transformed into elements which populate the Customer Emotions Ontology (CEO) creating a number of CEO instances. For example, the analysis of a number of comments from the users may imply Frustration (with a comment like *"The use of Pearl is frustrating because its lack of precision"*) and Unfulfillment (saying *"I felt left aside because of their absolute lack of Customer Care Service"*). These comments are classified and labeled. Others, on the other hand, might point to users feeling *"Thrilled"* (*"You can easily set up things to perform a full mix of your favorite songs using your mobile. And you can share it with your friends. Ain't it thrilling?"*), while several others express the emotion *"Fulfillment"* (*"It's all I need. It fulfills my expectations"*). These latter results indicate that the product will be recommended in the environment where it has influence. From the Product Manager Feedback component, these positive emotions might be driven towards what is called a *"lead"* in Customer Relationship Management (CRM) terminology, namely a commercial opportunity which can be conveyed through the form of a recommendation. As described in the previous section, where the breakthroughs of using semantics have been carefully discussed from an architectural viewpoint, including the underlying logical formalisms used by the Product Feedback Manager, semantic benefits are now also addressed from a Use Case perspective.

Fundamentally, both *"Frustration"*, *"Fulfillment"*, *"Thrilled"* and any other emotions are instances of the Customer Emotion Ontology (CEO), particularly at the Emotion concept instance level. The same applies for *"DJPhone"* as a Product concept instance and the potential set of customers providing

opinions and feedback being Customer concept instances from CRM ontology. Having such a complete knowledge-intensive structure as the backbone of the SEMO dynamics, it is easy to manage, extract and analyze a number of customer feedback management strategies. For example, customers greatly frustrated because of DJPhone expectations can be tracked, assisted and encouraged to test a different range of products or benefit from commercial discounts, all as part of a commercial strategy to maximize the efficiency of a number of CRM techniques, as pointed out previously, by optimizing customer satisfaction, which results in significant business turnover.

To sum up, semantics are the very backbone of the SEMO approach from a twofold standpoint. First, as the perfect Knowledge Representation structure and technology, and, last but not least, as the underlying logic-based formal mechanism to knowledge-wise added value.

## **5. EVALUATION**

### *5.1. Research Design*

With the aim of getting feedback concerning the work performed, an evaluation was carried out by means of the application of a questionnaire. The questionnaire was provided after the subjects had completed four differentiated steps. In the first place, as part of student assignments in one of the subjects in their last year of the Computer Science degree program, "Software Engineering III", the students were asked to use the ESACAKE tool (Colomo-Palacios *et al.*, 2008) as support for requirement management in a software development project. In the subsequent step, they were asked to post commentary in Orkut in relation to the use of ESACAKE. Third, the information obtained by SEMO was processed. Lastly, the users were required to review the semantic classification of their comment and afterwards immediately fill out a questionnaire regarding the processing of information of information realized by SEMO, with the questionnaire designed specifically for that purpose.

The aim of the questionnaire was to show whether the annotation and semantic categorization based on textual content performed by SEMO was correct. The questionnaire was composed of two different parts. Firstly, the subject had to provide identification data: age and gender. Secondly, the subject was required to categorize his emotions based on his commentary using Laros and Steenkamp's taxonomy (Laros and Steenkamp, 2005). Once the element in the taxonomy which



represented his emotion was established, the user was asked to compare the result with the annotation realized by SEMO. The comparison, which could yield distinct results, was pointed in the questionnaire by means of closed questions. The comparison values assigned were in the following categories:

- Agreement
- Agreement with the basic emotions in the second level but not with the individual emotion of the third level.
- Agreement in the Valence of emotions (positive and negative) of the first level
- Zero agreement

### 5.2. Sample

The sample was composed of students in their last year of the Computer Science degree program at Carlos III University. These students use the ESACAKE tool to carry out the drawing up of user requirements in the course "Software Engineering III". The sample was composed of 17 women (32%) and 35 men (68%), with an average age of 25.6. Although this population might not completely reflect future users, most studies in the literature have used academics to provide queries and judge the relevance (Morrison, 2008).

### 5.3. Results

The results of the surveys, which were performed using printed copies, were subsequently coded in the SPSS statistical analysis tool. Users identified 91 emotions in their texts; SEMO found 73. The distribution of emotions found by users were twenty-two with negative valence and sixty-nine positive. The distribution of third level emotions identified by subjects and the level of agreement reached by SEMO can be observed in Table 2 and Figure 4. Table 2 presents third level emotions identified by subjects; in the columns scores of concordance are presented including third level emotion (Agree), second level- basic emotion (Basic Emotion), first level or valence (valence) and no matching (disagree). In addition, Figure 4 shows these data in graphical form, using referred levels of agreement (from Agree to Disagree). In this figure it can be seen that Basic Emotion agreement is reached in most cases, and even full agreement can be reached in many cases by SEMO.

**INSERT TABLE 2 HERE**

INSERT FIGURE 4 HERE

SEMO presents promising results. Specifically, detection of the "basic emotion" is the largest group of tests performed, with 36% of the total, followed by "agree" with 27% of cases. In particular, the emotion "content" has a very high number of occurrences (35%), with the number of matches in the different bands also being very high. Likewise, of the 91 emotions that have been identified, SEMO can find a total of 73, with this figure representing a very interesting ratio from a NLP tracking point of view.

To evaluate the performance of annotation of SEMO, we used the standard recall, precision and  $F_1$  measures. Recall and precision measures reflect the different aspects of annotation performance. These measures were first used to measure an Information retrieval system by Cleverdon et al. (1966).  $F_1$  measure was later introduced by van Rijsbergen (1979) in order to combine precision and recall measures, with equal importance, into a single parameter for optimization. The use of these measures is not new in sentiment classification effectiveness (Miao, Li and Dai, 2009; Strapparava and Mihalcea, 2008; Tan and Zhang, 2008; van Atteveldt et al., 2008). Precision, Recall and  $F_1$  measures are defined as follows:

$$\text{Precision} = \frac{\text{Categories found and correct}}{\text{Total Categories Found}}$$

$$\text{Recall} = \frac{\text{Categories found and correct}}{\text{Total Categories Correct}}$$

$$F_1 = \frac{(2 * \text{Precision} * \text{Recall})}{(\text{Precision} + \text{Recall})}$$

Table 3 shows the experimental results of our system, applying precision, recall and  $F_1$  measures to three scenarios, from total coincidence of emotion to just valence coincidence:

INSERT TABLE 3 HERE

#### 5.4. Discussion.

Taking the results into account, if we assume a minimum of basic emotion concordance, performance results are satisfactory. It is true that total agreement must be reached, but identifying basic emotions is indeed an important result. Compared to previous works, these results are promising. Miao et al. (2009) presented values of

precision, recall and F1 similar to SEMO, and since they only use the valence or orientation (positive and negative) of the customer reviews, SEMO results can be considered on the same level, with SEMO being as feasible and effective as the approach of Miao et al. (2009). With respect to the work of van Atteveldt et al. (2008), which only distinguishes between positive and negative relations, SEMO also presents similar scores. In the empirical study of sentiment analysis for Chinese documents by Tan and Zhang (2008), the use of Information gain for sentimental terms selection and Support vector machines for sentiment classification presented and F1 of 0.9043, 4 points more than SEMO joint F1 measure. However, the results of SEMO in finding basic emotions (or a predefined set of emotions) are very promising if these results are compared to similar efforts (e.g. Danisman and Alpkocak, 2008; Strapparava and Mihalcea, 2008). However, the empirical test of SEMO also has room for improvement. A more in-depth analysis regarding the lack of precision within basic emotions analysis reveals that this can be a result of an incomplete definition of the vocabularies used in NLP, which must be enhanced. On the other hand, analyzing disagreement scores are as follows. Eighteen emotions were not identified and two more were misclassified. According to these two, authors read comments and discovered that subjects use irony in their opinions. Since all other comments were made with frankness, we assume that SEMO must improve its NLP features in order to analyze language taking into account all shades of opinion.

## **6. CONCLUSIONS AND FUTURE WORK**

The invention and subsequent adoption of CRM has initiated a change in business practices with respect to customers. The interaction philosophy of CRM complements the exploitation of new Social Network channels provided by Web 2.0. In this environment of constant communication between clients and organizations, but also clients among themselves, a platform which permits the automatic analysis of customer opinions and their emotional implications can have a profound impact on a Social Web environment. This scenario was the basis for the development of SEMO. SEMO is an analyzer of emotions expressed by users within Social Networks. Based on Natural Language Processing and the application of semantics for the categorization of opinions, the results of the application are promising, particularly from the viewpoint of applicability to marketing and new product development in co-creative environments. This affirmation is based on the fact

that the evaluation of the results in the experimental setup are very positive, as measured by Precision and Recall rates. Taking into account the possibilities initiated by the current research effort, three separate lines of future research may be considered. First of all, extending the capabilities of the framework to cover other aspects relative to Web 2.0, such as Chat services or corporate blogs; Secondly, extending the use of SEMO to also deal with conventional CRM systems which extract textual information from Call Centers or which are support for Sales Force Automation. Lastly, extending the functionalities of SEMO for the attention and annotation of emotions of customers in telephone customer service environments and by means of a video call center. This latter extension of the model would involve dealing with semantic emotional voice synthesis and the synthesis of gestures and body language in the interaction between client and the elements of attention.

## References

- Baader, F., Calvanese, D., Mcguinness, D.L., Nardi, D. and Patel-Schneider, P.F. (2003). *The description logic handbook: theory, implementation, and applications*, Cambridge, UK: Cambridge University Press.
- Bagozzi, R.P., Gopinath, M. and Nyer, P.U. (1999). The role of emotions in marketing, *Journal of the Academy of Marketing Science* 27(2): 184-206.
- Band, W. (2008). *The Forrester Wave™: Enterprise CRM Suites, Q3 2008*. <http://www.forrester.com/Research/Document/Excerpt/0,7211,44968,00.html> (accessed 08th February 2010).
- Berners-Lee, T., Hendler, J. and Lassila, O. (2001). The Semantic Web, *Scientific American* 284 (5): 35-40.
- Bueren, A., Schierholz, R., Kolbe, L.N. and Brenner, W. (2005). Improving performance of customer-processes with knowledge management, *Business Process Management Journal* 11(5): 573-588.
- Chen, I.J. and Popovich, K. (2003). Understanding customer relationship management (CRM), *Business Process Management Journal* 9(5): 672-688.
- Cleverdon, C.W., Mills, J. and Keen, E.M. (1966). *Factors determining the performance of indexing systems*, Cranfield, U.K.: College of Aeronautics.

- Colomo-Palacios, R., Gómez-Berbís, J.M., García-Crespo, A. and Puebla-Sánchez, I. (2008). Social Global Repository: using semantics and social web in software projects, *International Journal of Knowledge and Learning* 4(5): 452-464.
- Danisman, T. and Alpkocak, A. (2008). Feeler: Emotion Classification of Text Using Vector Space Model, in *Proceedings of the AISB 2008 Convention, Communication, Interaction and Social Intelligence* (Aberdeen, UK 2008). 53-59.
- Davenport, T.H., Harris, J.G. and Kohli, A.K. (2001). How do they know their customers so well?, *Sloan Management Review* 42(2): 63-73.
- Deerwester, S., Dumais, S.T., Landauer, T.K., Furnas, G.W. and Harshman, R.A. (1990). Indexing by latent semantic analysis, *Journal of the Society for Information Science* 41(6): 391-407.
- Eikelmann, S., Hajj, J. and Peterson, M. (2008). Opinion piece: Web 2.0: Profiting from the threat, *Journal of Direct, Data and Digital Marketing Practice* 9(3): 293-295.
- Eckerson, W.W. (1995). Three Tier Client/Server Architecture: Achieving Scalability, Performance, and Efficiency in Client Server Applications, *Open Information Systems* 10(1):1-12.
- Fensel, D. (2002). *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, Berlin / Heidelberg: Springer.
- Francisco, V., Gervás, P. and Peinado, F. (2007). Ontological reasoning to configure emotional voice synthesis, in *Proceedings of the First International Conference of Web Reasoning and Rule Systems* (Innsbruck, Austria, 2007); Berlin / Heidelberg: Springer. 88-102.
- García-Crespo, A., Colomo-Palacios, R., Mencke, M. and Gómez-Berbís, J.M. (2008). CUSENT: Social sentiment analysis using semantics for customer feedback, in M.D. Lytras and P. Ordóñez (Eds.) *Social web evolution: integrating semantic applications and web 2.0 technologies*. Hershey, PA: IGI Global.
- Giunchiglia, F., Marchese, M. and Zaihrayeu, I. (2007). Encoding classifications into lightweight ontologies, *Journal on Data Semantics* 8: 57-81.
- Gountas, J., and Gountas, S. (2007). Personality orientations, emotional states, customer satisfaction, and intention to repurchase, *Journal of Business Research* 60(1):

72-75.

Heyman, P. and Garcia-Molina, H. (2006). *Collaborative Creation of Communal Hierarchical Taxonomies in Social Tagging Systems*, Palo Alto, California: Stanford University.

Huang, M.H. (2001). The Theory of Emotions in Marketing, *Journal of Business and Psychology* 16(2): 239-247.

Izard, C.E. (1977). *Human Emotions*, New York: Plenum Press.

Jarrar, M. (2008). Towards Effectiveness And Transparency In E-Business Transactions, An Ontology For Customer Complaint Management, in R. García (Ed.) *Semantic Web for Business: Cases and Applications*. Hershey, PA: IGI Global.

Kazmer, M.M., Burnett, G. and Dickey, M.H. (2007). Identity in customer service chat interaction: Implications for virtual reference, *Library and Information Science Research* 29 (1): 5-29.

King, S. and Burgess, T.F. (2008). Understanding success and failure in customer relationship management, *Industrial Marketing Management* 37 (4): 421-431.

Laros, F.J.M. and Steenkamp, J.B.E.M. (2005). Emotions in consumer behavior: A hierarchical approach, *Journal of Business Research* 58(10): 1437-1445.

Levitt, T. (1983). After the sale is over,..., *Harvard Business Review* 61(5): 87-94.

Linoff, G.S. and Berry, M.J. (2002). *Mining the web, transforming customer data into customer value*, New York: John Wiley.

López, J.M, Gil, R., García, R., Cearreta I. and Garay, N. (2008). Towards an Ontology for Describing Emotions, in *Proceedings of the 1st World Summit on the Knowledge Society*. (Athens, Greece, 2008); Berlin / Heidelberg: Springer. 96-104.

Magro, D. and Goy, A. (2008). The Business Knowledge for Customer Relationship Management: an Ontological Perspective, in *Proceedings of 1st International Workshop on Ontology-supported Business Intelligence* (Karlsruhe, Germany, 2008); New York, NY: ACM International Conference Proceeding Series. Article No.4.

Mahdavi, I., Cho, N., Shirazi, B. and Sahebjamnia, N. (2008). Designing evolving user profile in e-CRM with dynamic clustering of Web documents, *Data & Knowledge Engineering* 65 (2): 355-372.

Maoz, M. (2008). Magic Quadrant for CRM Customer Service

- Contact Centers, 2008  
<http://www.gartner.com/DisplayDocument?id=626908> (accessed 08th February 2010).
- Marston, P. (2008). The Forrester Wave™: Midmarket CRM Suites, Q3 2008. Available online: <http://www.forrester.com/Research/Document/Excerpt/0,7211,44969,00.html> (accessed 08th February 2010).
- Mathieu, Y. (2005). Annotation of emotions and Feelings in texts. In *Proceedings of the First International Conference ASCII 2005* (Beijing, China, 2005 ); Berlin / Heidelberg: Springer. 350-357.
- McKinsey (2007) How businesses are using Web 2.0: A McKinsey Global Survey. *The McKinsey Quarterly*, March 2007. <http://www.mckinseyquarterly.com/PDFDownload.aspx?L2=16&L3=16&ar=1913&gp=0> (accessed 08th February 2010).
- Miao, Q., Li, Q. and Dai, R. (2009). AMAZING: A sentiment mining and retrieval system, *Expert Systems with Applications*, 36(3): 7192-7198.
- Morrison, P.J. (2008). Tagging and searching: Search retrieval effectiveness of folksonomies on the World Wide Web, *Information Processing and Management*, 44 (4): 1562-1579.
- O'Reilly, T. (2005). *What Is Web 2.0. Design Patterns And Business Models For The Next Generation Of Software*, 30 September 2005  
<http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html> (accessed 08th February 2010).
- OpenSocial. (2008). *Google Code official web site*  
<http://code.google.com/apis/opensocial/> (accessed 08th February 2010).
- O'Reilly, T. (2007). What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software, *Communications & Strategies* 1: 17-27.
- Rivera, I., Mencke, M., Gómez, J.M., Alor-Hernández, G. and García-Crespo, A. (2008). A Collaborative Open Social Network Dataset based on Email Ranking and Filtering, in *Proceedings of the 3rd IEEE International Conference on Systems* (Cancún, Mexico, 2008); Washington, DC: IEEE Computer Society Press. 13-18.
- Salton, G. Wong, A. and Yang, C. S. (1975). A Vector Space Model for Automatic Indexing, *Communications of the ACM* 18(11): 613-620.

- Schmitz, P. (2006). Inducing Ontology from Flickr Tags. Collaborative Web Tagging Workshop, in *Proceedings of the 15th WWW Conference (Edinburgh, UK, 2006)*; New York: ACM Press. 63-72.
- Schoefer, K., and Diamantopoulos, A. (2008). Measuring experienced emotions during service recovery encounters: construction and assessment of the ESRE scale, *Service Business* 2(1): 65-81.
- Shadbolt, N., Hall, W. and Berners-Lee, T. (2006). The Semantic Web Revisited, *IEEE Intelligent Systems* 21(3): 96-101.
- Shani, D. and Chalasani, S. (1992). Exploiting niches using relationship marketing, *The Journal of Consumer Marketing* 9(3): 33-42.
- Stone, M. (2009). Staying customer-focused and trusted: Web 2.0 and Customer 2.0 in financial services, *The Journal of Database Marketing & Customer Strategy Management* 16(2): 101-131.
- Strapparava, C. and Mihalcea, R. (2008). Learning to Identify Emotions in Text, in *Proceedings of the 2008 ACM symposium on Applied computing* (Fortaleza, Ceara, Brazil); Rochester, Illinois: ACM Publishing. 1556-1560.
- Takashi and Manabu (2006). A Survey of Sentiment Analysis, *Journal of Natural Language Processing* 13(3): 201-241.
- Tan, S. and Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents, *Expert Systems with Applications*, 34(4): 2622-2629.
- van Atteveldt, W., Kleinnijenhuis, J., Ruigrok, N. and Stefan Schlobach, S. (2008). Good News or Bad News? Conducting Sentiment Analysis on Dutch Text to Distinguish Between Positive and Negative Relations, *Journal of Information Technology & Politics*, 5(1): 73-94.
- Van Damme, C., Christiaens, S. and Vandijck, E. (2007). Building an employee-driven CRM ontology, In *Proceedings of the IADIS Multi Conference on Computer Science and Information Systems*, (Lisbon, Portugal, 2007); Lisbon, Portugal: IADIS Press.330-334.
- Van Dolen, W., de Ruyter, K. and Lemmink, J. (2004). An empirical assessment of the influence of customer emotions and contact employee performance on encounter and relationship satisfaction, *Journal of Business Research* 57(4): 437-444.



Van Rijsbergen, C. J. (1979). *Information Retrieval*, Newton, MA: Butterworth-Heinemann.

Zeelenberg, M. and Pieters, R. (2004). Beyond valence in customer dissatisfaction A review and new findings on behavioral responses to regret and disappointment in failed services, *Journal of Business Research* 57(4): 445-455.

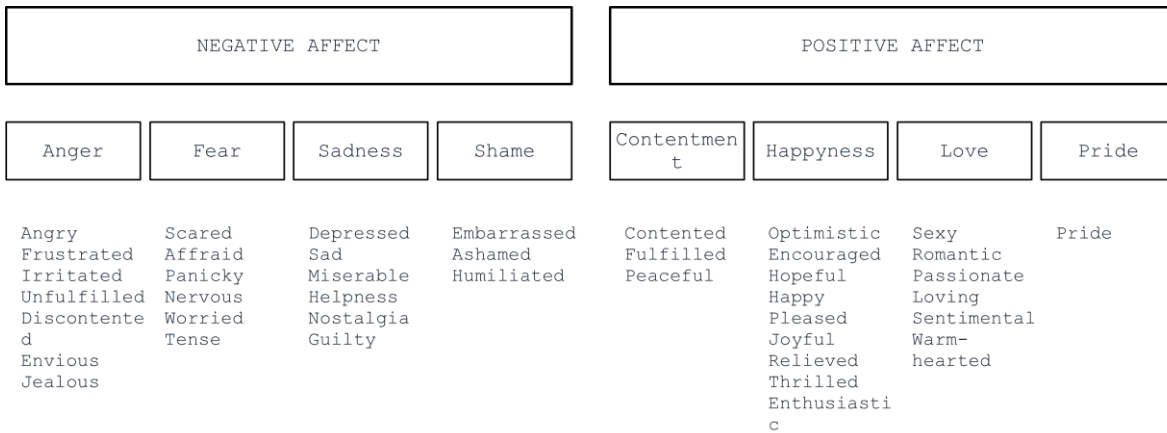


Fig. 1. Hierarchy of consumer emotions

Table 1. RDF Storages Features

Tool	Scale (mil. of state m.)	Inference	Load Speed (1000 st./s ec .)	Hardw are (GB of RAM)	Comment
KAON2	~10	OWL DL + rules	20	0.5	Backward- chaini ng; concrete figure s missin g
RacerPro	1	OWL DL	-	0.5	Backward- chaini ng
Minerva (IBM)	2	OWL Lite +/-	>1	0.5	
Triple20	40	OWL Lite +/-	6	2	
SwiftOWL IM	10-80	OWL Lite +/-	20-60	1-16	
Sesame 2.0 NS	70	RDFS +	6	0.8	Named graph suppor t
ORACLE 10R2	100	RDFS +	>1	2	Named graph suppor t
Jena v2.1/2.3	7-200	-	?-6	2-?	Speed and RAM not report ed for

---

					200M st.
KOWARI	235	None	4	?	
RDF	262	OWL	>1	?	Backwa rd- chaini ng
Gateway		Lite +/-			
AlegroGr aph	1 000	RDFS -	20	2	
OpenLink	1 060	None	12	8	
BigOWLIM	1 060	OWL Lite +/-	4	12	

---

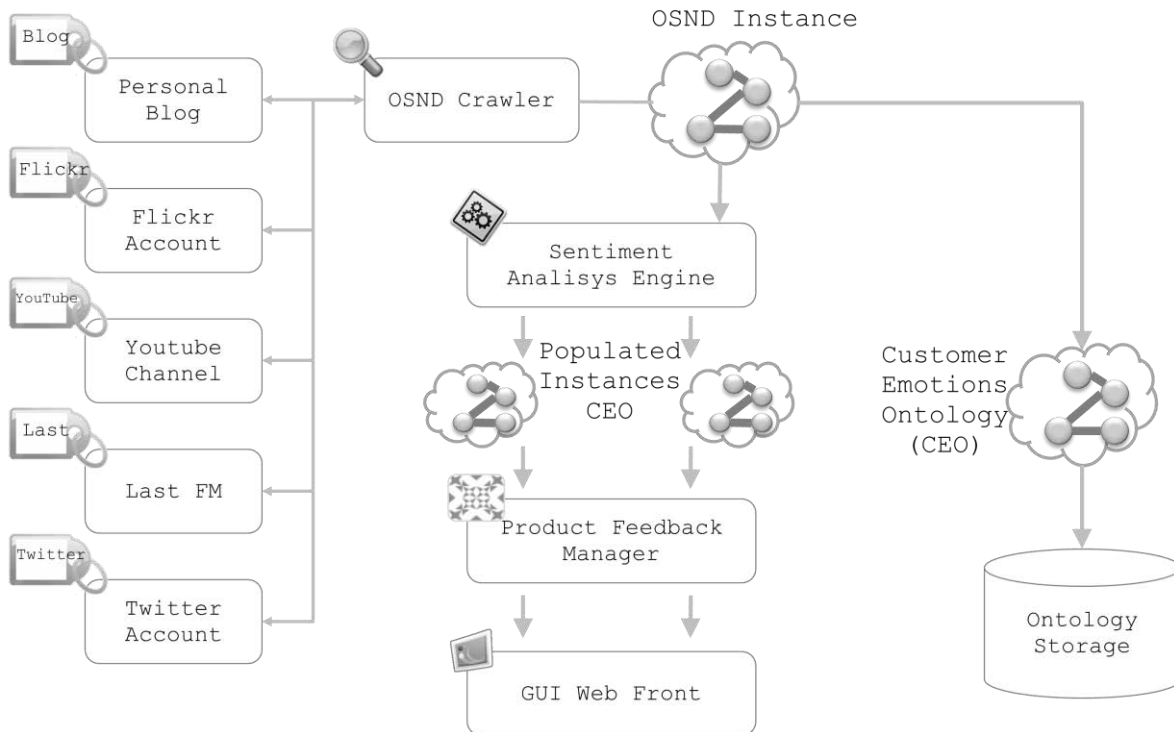


Fig. 2. SEMO Dynamics

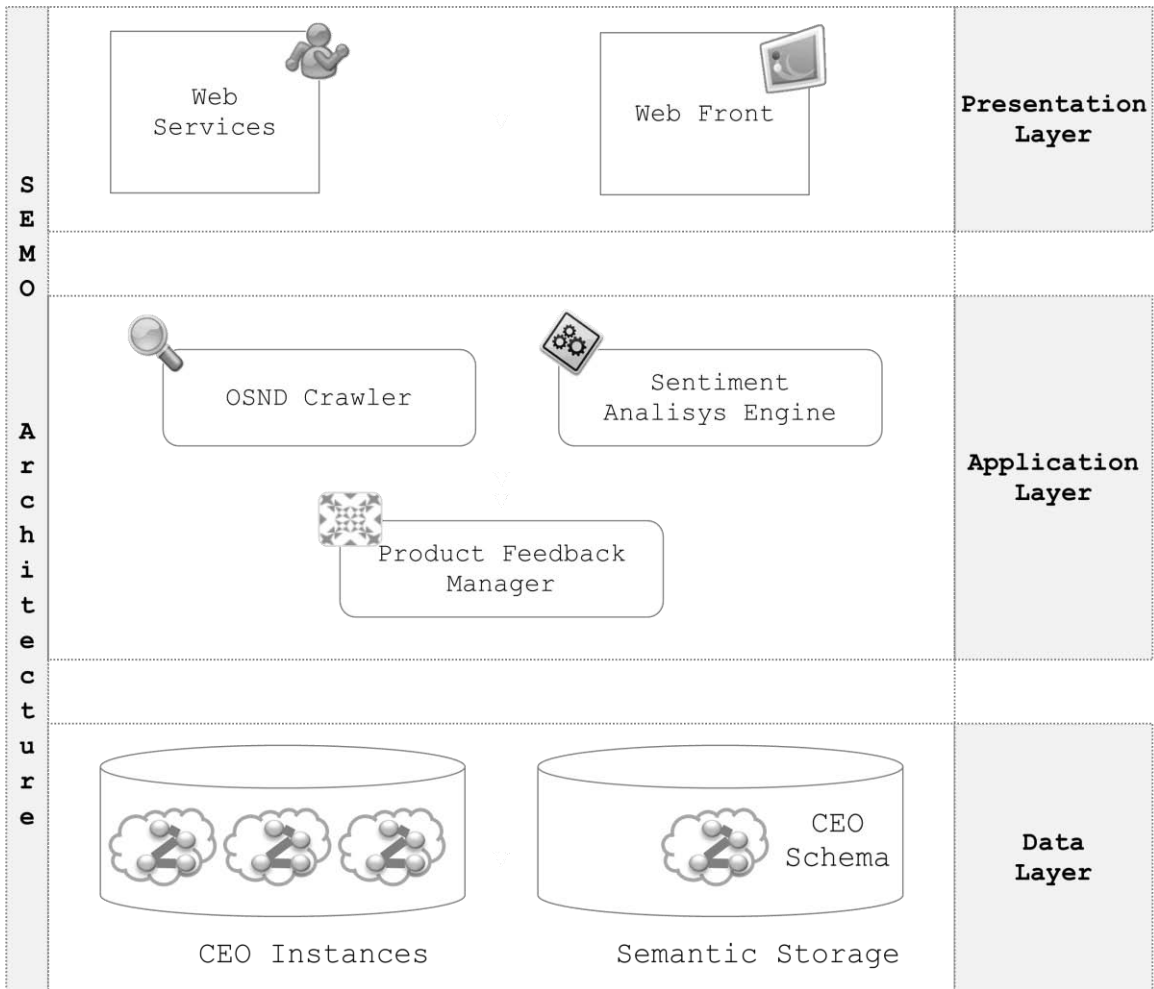


Fig. 3. SEMO Layer Architecture View

Table 2. Emotions identified by subjects

	Agree	Basic Emotion	Valence	Disagree
Frustrated	4	2	0	0
Irritated	1	2	0	1
Unfulfilled	1	2	0	0
Discontented	2	6	0	1
Contented	7	7	10	8
Fulfilled	1	3	1	1
Optimistic	1	3	1	0
Happy	5	5	0	4
Pleased	3	3	1	5
TOTAL	25	33	13	20

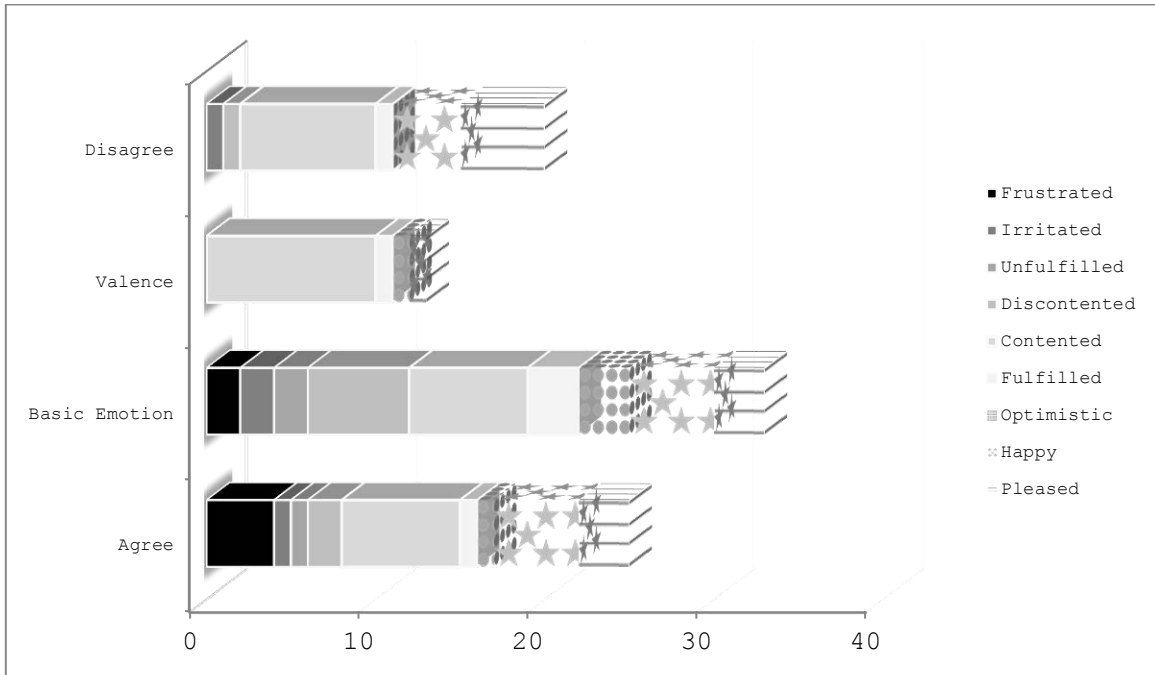


Fig. 4. Emotions identified by subjects in a graphic way grouped by matching category



Table 3. Precision, Recall and F1 Measures in different scenarios

	Precision	Recall	F1
Agree	0.274725275	0.342465753	0.304878049
Basic Emotion +Agree	0.637362637	0.794520548	0.707317073
Valence+ Basic Emotion +Agree	0.78021978	0.97260274	0.865853659