

# A Multidomain and Multilingual Conceptual Data Model for Online Reviews Representation

Marcirio Silveira Chaves<sup>1</sup> and Winnie Picoto<sup>2</sup>

<sup>1</sup>*Business and Information Technology Research Center (BITREC), Universidade Atlântica, Fábrica da Pólvora de Barcarena, 2730-036 Barcarena, Portugal*

<sup>2</sup>*Centro de Investigação Avançada em Gestão do ISEG (ADVANCE), Instituto Superior de Economia e Gestão – ISEG, Lisboa, Portugal*

Keywords: Model-Driven Engineering, Metamodel, UML, Online Reviews.

Abstract: User-Generated Content (UGC) such as online reviews are freely available in the web. This kind of data has been used to support clients' and managerial decision making in several industries, e.g. books, tourism or hospitality. However, the challenge is how to represent this information in a structured way in order to leverage on the information provided by the use of Web 2.0 applications. To deal with this challenge, models and metamodels have been used to support a set of concrete applications in several sub-domains into Computer Science and Information Systems body of knowledge (Karagiannis and Höfferer, 2006). This paper focuses on the model-driven engineering and introduces a new multidomain and multilingual conceptual data model to represent UGC. This model is based on a characterization of online reviews and aims to capture all the facets of these reviews. The characterization of the reviews' sentences extends previous models (such as Martin and White, 2007; Ding et al., 2008; Liu, 2010). Applications build on the model proposed in this paper may allow in-depth analysis of the fine-grained and disperse knowledge existent in the UGC. Furthermore, as this model is domain-independent it can be used to represent multiple types of reviews.

## 1 INTRODUCTION

The usage of Web 2.0 websites to express opinions has increased in the last years generating a huge amount of data which has been used to support decision making. The distillation of knowledge from this unstructured and dispersed information can be a key factor to managers improve their products and services. Commercial Web 2.0 tools such as Clarabridge (2012), Attensity (2012), SocialMetrics (2012) and Synthesio (2012) are available, but their cost still remains inaccessible to Small and Medium Enterprises (SME). Considering that SME represent 99% of the European businesses (EC, 2012), the development of a financially affordable solution is needed.

Moreover, commercial Web 2.0 tools hardly present the data model they use to represent the unstructured opinions' texts into structured ones. Such a model is omitted or really does not exist in the literature. This literary gap evidences the few

attention that model-driven engineering has given to Web 2.0 data.

On the other hand, from the academic point of view, Liu (2010) presents a characterization of the opinions expressed thought plain text data, as blogs, tweets and full-blown service/product review. This characterization may be extended and used to develop a model to represent online reviews.

A model is an abstraction of phenomena in the real world, and a metamodel is yet another abstraction, highlighting properties of the model itself (van Gigch, 1991). Models and metamodels have been used to support a set of concrete applications in several sub-domains into Computer Science and Information Systems, as described in the survey carried out by Karagiannis and Höfferer (2006). The model-driven approach defines relationships among concepts in a domain and precisely specifies the key semantics and constraints associated with these domain concepts (Schmidt, 2006).

A conceptual data model, sometimes called domain model, is typically used to explore domain concepts with project stakeholders. It is also used to map concepts and their relationships in a domain. We argue that it is well suited to represent knowledge from opinionated texts.

The main question addressed in the present study is how to capture the different facets of online reviews in an usable knowledge representation? This paper first describes the properties of opinionated texts which should be taken into account along the model's design. The solution approach outlines a new multidomain and multilingual conceptual data model for capturing and storing knowledge from opinionated texts.

The model proposed is within the scope of the framework described in Chaves, Trojahn and Pedron (2012). This framework aims to provide an environment to customer knowledge management which integrates information from online reviews, stores it in a knowledge base and presents it in a user interface. A knowledge base is built to implement the conceptual data model proposed in this paper.

This paper is structured as follows: Section 2 discusses some of the existing work about modelling. Section 3 presents a characterization of online reviews, detailing the main elements which should be explicit and represented in a model. Section 4 introduces the scenario problem, raising some questions which can be answered with the help of a model. Section 5 describes the solution approach which is based on a conceptual data model proposed to represent knowledge from opinionated texts. Section 6 describes manual and automatic approaches for recognizing facets in online reviews. In Section 7 the conclusions, final remarks and further research are discussed.

## 2 RELATED WORK

UGC modelling is a recent field of knowledge and related works seem to be rare making difficult a direct comparison with existing studies. To our knowledge, this is the first academic work that seeks to develop a metamodel and a class model to structure the unstructured UGC. In fact, metamodels has been widely used in the software engineering to solve problems from the real world (Karagiannis and Höfferer, 2006). Related works are concerned with opinion mining of online reviews and online reviews classification. Although there have been a stream of works aiming at capture knowledge from those

unstructured review sentences which stands for the relevance to address this matter, less attention have been given to represent that information in a structured way so reviews' could be used to support managerial decisions.

Regarding previous studies that developed a metamodel, the present work is most closely related to Chaves, Rodrigues and Silva (2007). They used a metamodel to represent geographical information in order to implement a knowledge base. They also used structured data as the model input. Their model stores administrative and physical geographic information which is exported as ontologies. The ontologies represent knowledge integrated from geo-administrative and geo-physical domains. Although, the classes Feature, Source and Feature-Relationship are reused by the model proposed in this paper, we deal with unstructured data, i.e. web texts, and do not export them as ontologies.

Looking at existing works that have dealt with online reviews, Turney (2002), Bai (2011), Khan et al. (2010) and Hu and Lu (2004) have developed different approaches to extract knowledge from that unstructured data.

Turney (2002) applied a semantic orientation to unsupervised classification of reviews. He developed an algorithm for classifying reviews as recommended or not recommended based on the average semantic orientation for each review sentence. The algorithm has an average accuracy of 74% evaluated based on reviews from four different sectors (automobiles, banks, movies and travel destinations). Although that work does not propose a class model to structure those reviews, it presents a set of attributes to classify reviews that we have also considered in our model.

In line with that work, Bai (2011) developed a method for sentiment analysis. Her results show that the proposed method is capable to identify a parsimonious set of predictive features suggesting that sentiments are captured by conditional dependencies among words.

Khan et al. (2010) presents an automated method for opinion extraction from customer reviews. The unstructured review sentences were represented in such a way that is possible to extract knowledge from it. They found that only a portion included opinion-oriented words that needed to be processed in order to get knowledge from those reviews and that the most important was the choice of feature set.

Hu and Lu's (2004) work aimed at mine and summarize customer reviews of a product which is argued by the authors to be different from the traditional ones since they only mined the features

on which the customers have expressed some opinion. Towards the accomplishment of such objective, they have taken three steps: (1) mining product features with comments; (2) opinion sentences identification and orientation; (3) results summary. The proposed method was validated by means of an experience which has demonstrated to be effective.

### 3 OPINIONS' CHARACTERIZATION

In order to understand what kind of knowledge to represent in the content underlying the online reviews, we build on the existing approaches of the opinion mining field. One of the most common approaches to deal with the opinion mining problem is the lexicon-based one. Lexicon-based approaches include the use of a list of nouns, verbs, adjectives and adverbs (Chesley et al., 2006) and a list of conjunctions and connectives (Liu, 2010). Chesley et al. (2006) use verbs and adjectives to classify English opinionated blog texts. Ding et al. (2008) use all of these parts of speech in a holistic lexicon-based approach. Khan et al. (2010) use auxiliary verbs to get features and opinion-oriented words about products from texts.

Processing textual data allows us to provide more fine-grained knowledge to the decision making. Online opinions can be expressed in several ways, using different words to express positive or negative feelings or even to be neutral. In order to capture as many information as possible from online reviews about a product or a service, first it is necessary to understand each sentence in the review.

We use and extend the definition proposed by (Ding et al., 2008; Liu, 2010; Martin and White, 2007) to analyse the reviews' sentences. Let the review be  $r$ . In the most general case,  $r$  is characterized as a set of the following elements  $\{O, F, SO, A, R, I, C, SG, H, S, PoS, CC\}$ , where: O: Object, F: Feature, SO: Semantic Orientation, A: Attitude, S: Source, SG: Suggestion; R: Recommendation, I: Intention, C: Complain, H: Holder, PoS: Part-of-Speech, and CC: Conjunction and Connective.

**1. Object (O):** An object is a product or a service, under review which is composed by a set of features. Objects are also called entities.

**2. Feature (F):** A feature is a component or part of an object. For example, *actor* and *photography* are features of a movie. Features are also called attributes or facets. According to Ding et al. (2008) a

feature can be explicitly or implicitly mentioned in a review.

**2.1 Explicit Feature:** If a feature  $f$  appears in review  $r$ , it is called an explicit feature in  $r$ . For example, in the sentence "*The hotel is located very near the city centre*", *location* is an explicit feature.

**2.2 Implicit Feature:** If  $f$  does not appear in  $r$  but is implied, it is called an implicit feature in  $r$ . In the sentence "*Hotel is far from public transportation*", *location* is an implicit feature.

**3. Sentence-Orientation (SO):** A review consists of a sequence of sentences  $r = \{s_1, s_2, \dots, s_m\}$  (Ding et al., 2008). A sentence can be evaluated from the following perspectives:

**3.1 Objectivity:** An objective sentence contains or mention facts, e.g. "This hotel is far from the airport, ca. 15km.", while a subjective sentence does not mention any fact, e.g. "The parking could be free".

**3.2 Positivity:** It describes the orientation present in a sentence, i.e. positive, negative, neutral and irrelevant. For example, the sentence "Free and fast wifi in the room", refers to a *positive* impression about the feature *room* in a hotel.

**3.3 Intensity:** It refers to the strength of the private state that is being expressed, in other words, how strong is an emotion or a conviction of belief (Wilson, 2008). It describes how intense it was the experience of using a product or service, i.e. very positive, positive, neutral, negative and very negative. For example, the sentence "Very kindly staff" refers to a very positive impression on the *staff* service.

**3.4 Negation:** It is marked if the sentence contains at least an occurrence of negation, e.g. not and never. This information is relevant to support algorithms in the task of recognizing several kinds of sentence-orientation (e.g. positivity and intensity).

**4. Attitude (A):** The evaluation of attitude expressions is a complex task. We use the Appraisal Theory (Martin and White, 2007), which is a systemic-functional approach to analysing how subjective language is used to express an attitude of some kind towards some target. Appraisal Theory specifies three attitude types: affect (personal emotional state), appreciation (evaluation of phenomena), and judgment (social or ethical appraisal of other's behaviour).

**4.1 Affect:** It refers to a personal emotional state, e.g. happiness, sadness or angry, and is the most explicitly subjective type of appraisal (Whitelaw et al. 2005). For example, the sentence "I was very happy for spend the holiday in this comfortable hotel" expresses a feeling of happiness.

**4.2 Appreciation:** It refers to the intrinsic object properties, e.g. dirty, cold or small.

**4.3 Judgment:** It refers to an opinion formed by judging other's behaviour, e.g. "attentive staff" and "friendly waiter". Both appreciation and judgment can be classified as positive or negative.

**5. Recommendation (R):** A recommendation is a positive or negative statement which explicitly mention in words, an action or advice as for example "I recommend!" or "if you plan to travel there, find another hotel closer to the main area, cause this one's not at all worth it!". In a recommendation it is also implicitly present the satisfaction of a reviewer. Reviews that state some intention can be also considered an implicit recommendation as for example, the sentence "If we get the chance to go to Roma again we will ensure that we stay at the X Hotel again".

**6. Suggestion (SG):** A hint given by the holder in order to improve the product or service and it could be an explicit or implicit suggestion:

**6.1 Explicit Suggestion:** It is a direct mention on a need of the product or service. For example, in a hotel service, an explicit suggestion is "The mattresses of the beds need to be exchanged".

**6.2 Implicit Suggestion:** It is an indirect mention on a need of the product or service, as for example, "the hotel does not offer Internet service".

**7. Intention:** it is a mention in which a holder explicitly intends (or not) to experiment again some product or service. It could be positive or negative.

**7.1 Positive Intention:** For example, "When I come back to Lisbon, I intend to stay there myself" and "I intend to come back there".

**7.2 Negative Intention:** For example, "Certainly, we won't repeat".

It is important to notice that an intention can be also considered as an implicit recommendation. The sentence "If we get the chance to go to Lisbon again we will ensure that we stay at the X Hotel again" is an example of this case.

**8. Complain (C):** It is a phrase expressing a feeling of dissatisfaction, or resentment. It is stronger than a suggestion.

**9. Opinion Holder (H):** The holder of a particular opinion is the person or the organization that holds the opinion (Ding et al., 2008). A holder is identified with demographic characteristics, e.g. name, city and country. For example, sites such as tripadvisor.com or booking.com classify holders according to types (including families with older children, families with young children, mature couples, groups of friends, solo travellers and young couples among others).

**10. Source (S):** An information source is a web site which contains a set of reviews. Examples include sites such as amazon.co.uk, tripadvisor.com and booking.com.

**11. Part-of-Speech (PoS):** In order to evaluate a sentence in a review, we should consider the parts-of-speech mentioned such as adjectives, adverbs and verbs. Adjectives are classified as positive (e.g. good, excellent and clean), negative (e.g. awful, boring and terrible), neutral (e.g. regular and indifferent) and dual, which can express positive and negative opinion (e.g. long). In this characterization, nouns are represented by concepts of a domain ontology and mapped as features (described in the item 2 - Feature).

**12. Conjunction and Connective (CC):** Connectives are words that help identifying additional adjective opinion words and their orientations. According to Liu (2010), one constraint is about conjunction (i.e. and), which says that conjoined adjectives usually have the same orientation. For example, in the sentence, "This room is beautiful and spacious.", if "beautiful" is known to be positive, it can be inferred that "spacious" is also positive, since people usually express the same opinion on both sides of a conjunction. Rules or constraints are also designed for other connectives (e.g. or, but, either-or, and neither-nor). For example, "This hotel is beautiful but difficult to get there", the occurrence after the connective *but* is an indicator of a negative opinion.

## 4 PROBLEM SCENARIO

The characterization described in Section 3 extends the definition proposed by Ding et al. (2008) and Liu (2010) and allows a better understanding of the complexity in processing online opinions written in natural language. An in-depth analysis need to be performed in order to show the relationships between each key text segment of this characterization. In that sense, a tool for supporting managerial decision making should be able to answer questions such as:

- What are the verbs more often used to describe positive opinions on the features A, B and C?
- Which online reviews make recommendations? Which of them are explicit?
- What types of holders give more (explicit and implicit) suggestions?
  - What is the most frequent semantic orientation in online reviews written in Italian?

- What is the co-occurrence between attitudes and features?

In order to answer these and other questions, it is necessary to design a conceptual data model to represent the complex characterization previously described. This model could be used to develop a decision support system to help managers to analyse the huge amount of information available on dispersed Web 2.0 applications.

However, finding relevant and useful information on the user generated content is a challenge in the Web 2.0 context. From the user point of view, it would be useful if reviews, comments and posts from multiple information sources could be summarized on a single view enabling him or her to make a decision about a specific product or service. On the other hand, from the managerial point of view, it is important to know specific features of a product or a service such as the best-selling gender of a book or a specific book, and the main negative reviews or comments about a hotel.

In both cases, the following requirements are addressed by the proposed model:

- To support multidomain and multilingual texts;
- To store multiple information sources;
- To support fast and simple generation of managerial reports;
- To allow a fine-grained storage of the online reviews' content.

## 5 SOLUTION APPROACH

This section presents a Unified Modeling Language (UML) class-based modelling approach to deal with the UGC featured in Section 3. It introduces the metamodel and details the main classes according to the characterization described in Section 3.

### 5.1 Opinion Conceptual Data Metamodel

Proliferation of social media applications and their increasingly complexity and sophistication claims for the development of a generic and robust metamodel to capture the tacit knowledge available in those applications' reviews. Figure 1 depicts the metamodel proposed to capture customers' opinions.

The class *Review* stores the identifier for the full content of each review, its date and the language in which the review is written. Each review is composed by one or more sentences or clauses which are stored in the class *Sentence*. The origin of each review is stored in the class *Source*. Each review has also a holder, who usually provides basic data such as name, city and country. The class *Holder* stores this data.

The class *Object* stores the name of the object, e.g. *hotel* Alpha, *movie* Beta and *book* Gama, that is reviewed. This class allows the metamodel to be multidomain as it is possible to store reviews in a wide range of topics. The class *Feature* stores the main terms in a domain. These terms can be provided by a domain ontology, such as *staff* or *room* for accommodations, *director* or *photographer* for a movie, and *author* or *title* for a bookstore.

The class *PoS* (Part-of-Speech) stores the parts-of-speech, i.e. adjectives, adverbs and verbs, contained in the sentences. The class *Semantic Orientation* captures the main perspectives in the context of an opinion. Some perspectives regards to the global object, i.e. *recommendation* or *intention*, while other express opinion on a specific feature, i.e. *positivity*, *complain*, *attitude*, *intensity* and *suggestion*.

The class *Feature-Sentence* makes the relationships among a feature, its sentences and the existing semantic orientation. The class *Co-oc PoS* stores the relationships between parts-of-speech and features. For example, the feature *staff* in the accommodation domain can be more often

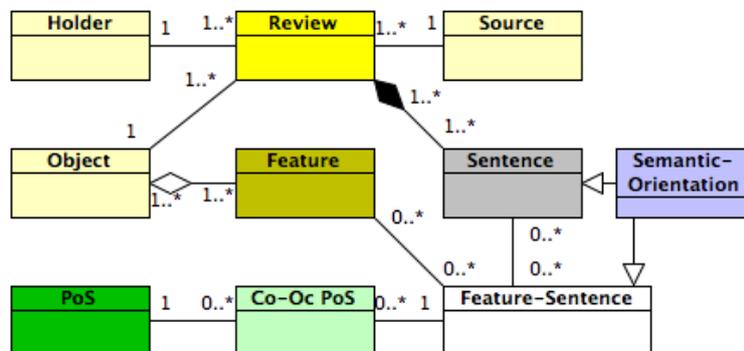


Figure 1: Opinion class metamodel.

mentioned with the adjective *helpful* in the reviews. The metamodel presented in Figure 1 is detailed in the next sections.

### 5.2 Representation of Reviews, Holders, Objects and Features

Figure 2 presents a more detailed part of *Review*, *Holder*, *Object* and *Feature* classes making explicit their main attributes.

The class *Review* stores the identifier of each review, date, language code in which the review is written and the rate. The class *Sentence* stores the reviews itself split in sentences. The class *Holder-Type* stores the types of customers, e.g. young couple, solo traveller and family with young children. The class *Object-Type* stores the types of objects, e.g. hotel, pension and apartment. The class *Object-Relationship* captures the relationships between objects, e.g. a hotel chain is composed by several individual hotels. The class *Feature-Relationship* captures the relationships between the features, e.g. *Swimming Pool* is part of *Pool*.

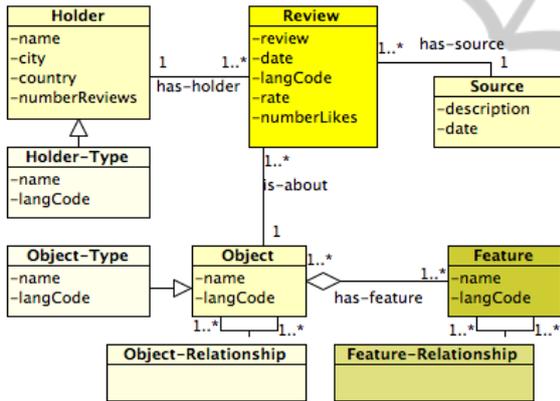


Figure 2: Review, Holder, Object and Feature Model.

### 5.3 Representation of Semantic Orientation of Features in the Sentences

This model intends to capture several perspectives regarding to the semantic orientation of a sentence. Some sentences present opinions about one or more features. In Figure 2, the classes *Sentence* and *Feature-Sentence* stores the existing indicators to each feature in each sentence.

The class *Sentence* is specialized in a *Recommendation* and an *Intention*. An opinion recognized as recommendation or intention is about the global object, e.g. hotel, movie and book, rather

than a specific feature. In addition, both can express a positive or negative opinion.

The semantic orientations of a sentence evolving specific features are stored in the class *Feature-Sentence*. A feature can be explicit or implicit in a sentence, which is captured in the attribute *explicit*. Sentences are also classified as the objectivity. An objective sentence contains or mentions facts, while a subjective sentence does not mention any fact. The class *Feature-Sentence* also captures a negative occurrence in each sentence.

Each feature in a sentence (*Feature-Sentence* class) can be also associated with a feeling related to:

- Positivity, e.g. positive, neutral and irrelevant;
- Attitude: which includes three types: affect, appreciation and judgment;
- Intensity, e.g. very positive, neutral and very negative;
- Complain, e.g. very serious, serious, not serious;
- Suggestion, e.g. “breakfast could be include more fruits”.

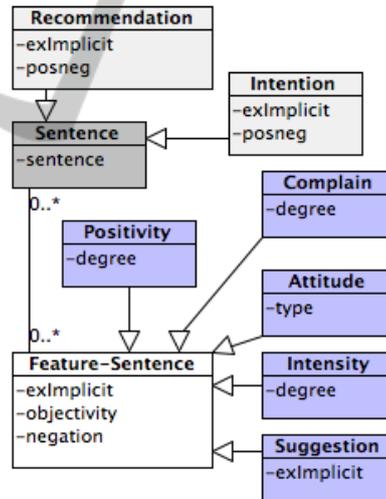


Figure 3: Representation of the semantic orientation of the features in the sentences.

Some of these opinions are about a specific feature of an object and other are generic ones. The classes *Positivity*, *Attitude*, *Intensity*, *Complain* and *Suggestion* store these kinds of opinion.

### 5.4 Representation of Co-Occurrences

The Opinion Model should also stores co-occurrences between features and syntactic categories (i.e. nouns, adjectives, adverbs, verbs and connectives) and features. Nouns are qualifiers, e.g.

price, design and accessibility, used in the comments. Connectives are usually conjunctions, e.g. or, but, either-or, and neither-nor. Co-occurrences allow managers to use the same words to advertise the product or service. For example, if *location* is more co-related with the noun or qualifier *metro*, this is an indicator that customers appreciate a hotel near a metro station. Figure 4 presents the part of the model that deal with co-occurrences.

The model also stores co-occurrences between features and syntactic categories, i.e. nouns, adjectives, adverbs and verbs, in the classes named *Co-Occurrence-X-Y*. The classes *Noun*, *Adjective*, *Adverb*, *Verb* and *Connective* store its own the list of terms.

It is important to notice that these classes can be very useful to Marketing managers, since the frequency of the co-occurrences indicates the association done by customers to each feature of the domain. For example, in hospitality industry, customers associate the adjectives *clean* and *spacious* to the feature room. With this information, a Marketing manager can explore these words in advertising campaigns.

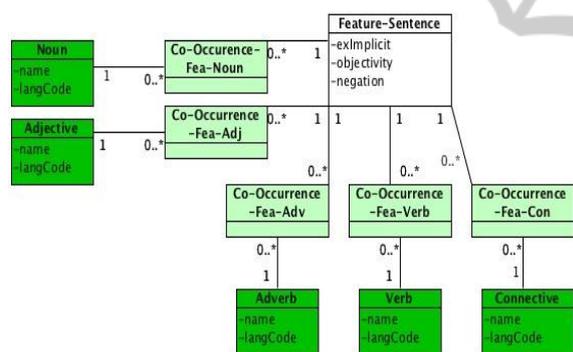


Figure 4: Representation of co-occurrences detected in online reviews.

### 5.5 Full Conceptual Data Model

Figure 5 presents the full conceptual data model which integrates all the parts previously described with four more classes. This model aims to be sufficiently generic to serve as building blocks for knowledge bases to store reviews and related data from different knowledge domains.

The multidomain aspects of this model can be explained based on the three classes: *Review*, *Object* and *Feature*. From this point, it is easier to understand the remain model and how to adapt it to represent knowledge in different domains.

Regarding the multilingual aspect, Figure 5

shows that each *review*, *object name* and *feature name* owns a multilingual representation as well as the specific classes *Holder Type* and *Object Type*. Each name in the syntactic category is also represented in its multilingual form using the attribute *langCode* whenever it is necessary to store text data.

Finally, using this model, managers can also explore historical data gathered. For example, it can be interesting for an accommodation manager or tourism agent to know how suggestions, recommendations and complains have evolved over time.

## 6 RECOGNIZING FACETS IN ONLINE REVIEWS

In order to show the relevance of the facets mentioned along this paper, we are working in manual and automatic tasks to foster the main classes presented in Figure 5. In both approaches features are mapped to concepts of an ontology which provides a common conceptualisation in a specific domain.

Considering the manual approach, Chaves, Gomes and Pedron (2012) analyzed a set of 1500 multilingual (English, Portuguese and Spanish) online reviews about small and medium hotels in the Lisbon region. They annotated each review according to the following criteria: types of holder, features, qualifiers (the most frequent terms used to mention features), positivity, intensity, country of origin of the holder and rate of review (numeric value assigned by the holder of the review). Their findings show that the features *room* and *service* were the concepts that guests pay more attention to in their review and ratings, and also points the main features which small and medium hotel managers should prioritize according to the profile of the guest (classes *Holder* and *Holder-Type* in Figure 5). Moreover, they show the co-relation between features and their intensity (i.e. strenght of polarity) in the reviews.

On the other hand, Tromp (2011) proposes a automatic four-step approach for multilingual sentiment analysis, which is composed by language identification, POS tagging, subjectivity detection and polarity detection. For polarity detection, he developed an algorithm which uses heuristic rules that stem from patterns (e.g. positive and negative).

The result of this step could be used to instantiate the classes *Sentence*, *Feature*, *Feature-Sentence* and

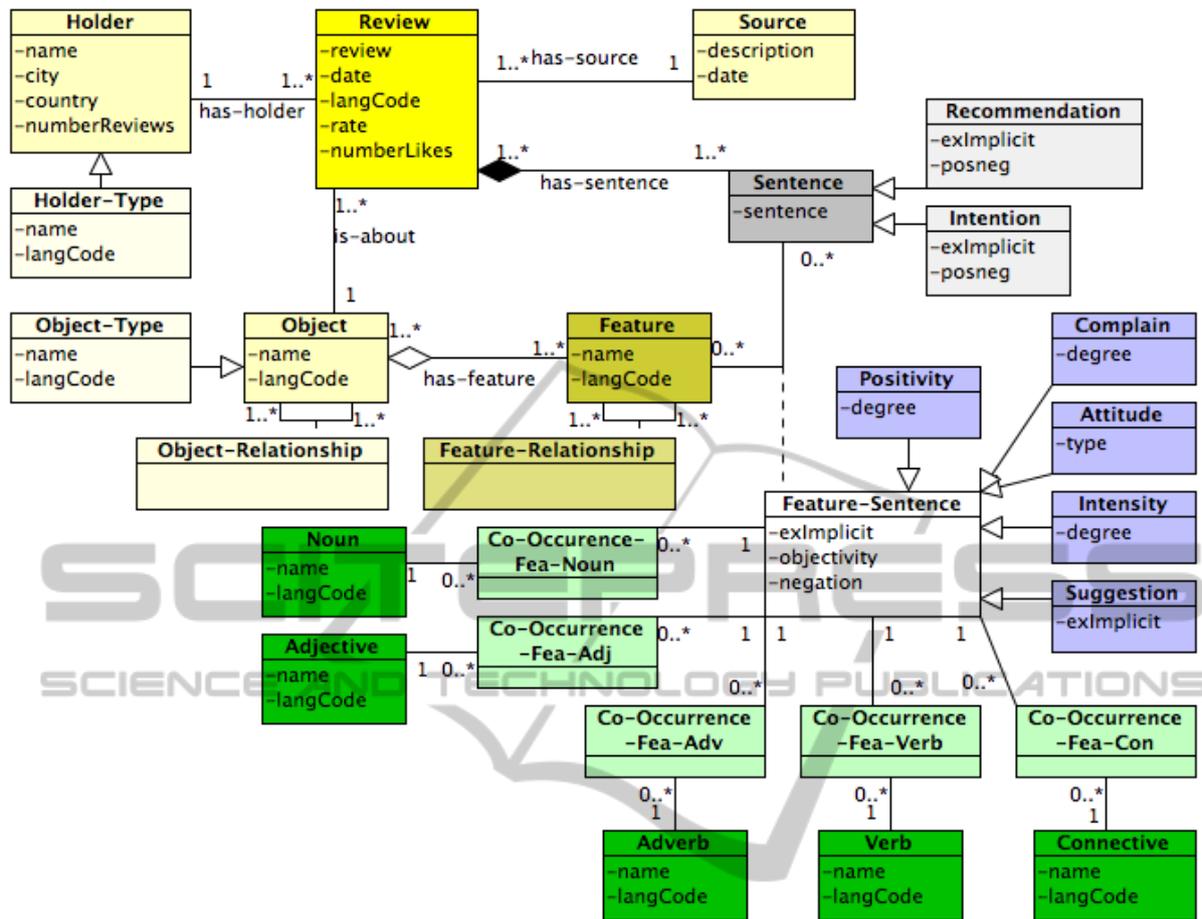


Figure 5: Full conceptual data model.

*Positivity* in the model proposed in Figure 5.

Chaves et al. (2012) developed an algorithm named PIRPO (Polarity Recognizer in Portuguese) which recognizes the polarity (class *Positivity* in Figure 5) expressed in each sentence of a review. PIRPO receives as input online reviews, a list of adjectives and concepts of an ontology. The output is a list of sentences with polarity recognized in each concept detected. This output is currently being used to instantiate the classes *Review*, *Sentence*, *Feature*, *Feature-Sentence* and *Positivity*. The extension of PIRPO will allow instantiate the remaining classes, mainly the semantic orientation of the features in the sentences.

## 7 CONCLUSIONS

The representation of the unstructured data in the web remains a challenge for the most of researchers and practioners in modelling field. This paper

introduces a multidomain and multilingual conceptual data model to capture the different facets of unstructured online reviews. It also presents a fine-grained characterization of opinionated texts, which evidences the main aspects of those opinions that could be relevant to managerial or user decision making.

It is important to mention that the level of complexity of this model varies according to the application requirements. The full conceptual data model can be partially or totally implemented depending on the context in which it is going to be used. Probably, the implementation of the semantic orientation of the features in the senteces is incremental due to the lack of algorithms to recognize all the aspects designed in the model.

The implementation of this model will also allow the storage of the content from multiple information sources, which will facilitate the fast and simple generation of managerial reports. Moreover, from the customer point of view, it will be possible to

search information about a specific object (e.g. book, movie or hotel) in a single place.

Applications built based on the proposed model may allow in-depth analysis of the fine-grained knowledge dispersed in the web. However, the success of the implementation of this conceptual data model is also dependent on an algorithm able to recognize the facets in the opinionated texts. The automatic identification of the semantic orientation of the features in the reviews remains a current challenge for Computer Science researchers.

To better test the model in the multidomain aspect, we should use instances from other domain than accommodations (e.g. books, cars or movies). Regarding to the multilingual representation, we should automatically load the reviews annotated by Chaves, Gomes and Pedron (2012). These annotations cover most of the facets in the conceptual data model.

Finally, the model proposed is in its first version and we know that there is room for improvements. As future work, the model will be also tested with an application for information visualization developed in Carvalho and Chaves (2012).

## ACKNOWLEDGEMENTS

This research was partially supported by the national funds of FCT – the Portuguese Science and Technology Foundation within the strategic project PEst-OE/EGE/UI4027/2011.

## REFERENCES

- Attensity, 2012. Available at <http://www.attensity.com>. Last access: January 6, 2012.
- Bai, X., 2011. Predicting Consumer Sentiments from Online Reviews. *Decision Support Systems* 50(4), March, *Elsevier Science*, p. 732-742.
- Carvalho, E.; Chaves, M. S., 2012. Exploring User Generated Data Visualization in the Accommodation Sector. *Proceedings of the 16<sup>th</sup> International Conference Information Visualisation, IEEE, Montpellier, France*, 10-13 July.
- Casey W., Navendu G., and Shlomo A., 2005. Using Appraisal Groups for Sentiment Analysis. In *Proceedings of the 14<sup>th</sup> ACM International Conference on Information and Knowledge management (CIKM '05)*. ACM, New York, NY, USA, 625-631.
- Chaves, M. S.; Gomes, R. and Pedron, C., 2012. Decision making based on Web 2.0 Data: The Small and Medium Hotel Management. *Proceedings of the 20<sup>th</sup> European Conference on Information Systems*, Barcelona, Spain, 10-13 June.
- Chaves, M. S.; Freitas, L. A.; Souza, M. and Vieira, R., 2012. PIRPO: An Algorithm to deal with Polarity in Portuguese Online Reviews from the Accommodation Sector. *Proceedings of the 17<sup>th</sup> International Conference on Applications of Natural Language Processing to Information Systems (NLDB)*, Groningen, The Netherlands, 26-28 June.
- Chaves, M. S.; Rodrigues, C. and Silva, M. J., 2007. Data Model for Geographic Ontologies Generation. XATA2007 - XML: Aplicações e Tecnologias Associadas. Ramalho, José Carlos; Lopes, João Correia and Carriço, Luís (Eds.). 15-16 February, Lisbon, Portugal.
- Chaves, M. S.; Trojahn, Cássia and Pedron, Cristiane Drebes, 2012. A Framework for Customer Knowledge Management based on Social Semantic Web: A Hotel Sector Approach. In: *Customer Relationship Management and the Social and Semantic Web: Enabling Clients Conexus*. Colomo-Palacios, Ricardo; Varajão, João and Soto-Acosta, Pedro (Eds.). p. 141-157, Hershey, PA: IGI Global. ISBN: 978-161-35-0044-6
- Chesley, P.; Vincent, B.; Xu, L. and Srihari R., 2006. Using Verbs and Adjectives to Automatically Classify Blog Sentiment. in *AAAI Symposium on Computational Approaches to Analysing Weblogs (AAAI-CAAW)*, 27–29.
- Clarabridge, 2012. Sentiment and Text Analytics Software - Clarabridge. Available at <http://clarabridge.com>. Last access: January 6, 2012.
- Consoli, D.; Diamantini, C. and Potena, D., 2009. Affective Algorithm to Polarize Customer Opinions. *Proceedings of the 11<sup>th</sup> International Conference on Enterprise Information Systems*, Volume HCI, ICEIS (5), Milan, Italy, May 6-10, 157-160.
- Ding, X., Liu, B., and Yu, P. S., 2008. A Holistic Lexicon-based Approach to Opinion Mining. *Proceedings of the Conference on Web Search and Web Data Mining (WSDM)* - ACM, Palo Alto, California, USA, p. 231-240.
- EC, (2012). European Commission: Enterprise and Industry. Small and Medium-sized Enterprises (SMEs) Fact and Figures about the EU's Small and Medium Enterprise. Available at [http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/index\\_en.htm](http://ec.europa.eu/enterprise/policies/sme/facts-figures-analysis/index_en.htm). Last access: January 8, 2012.
- Hu, M.; Liu, B., 2004. Mining and Summarizing Customer Reviews. *Proceedings of the 10<sup>th</sup> ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'04)*, August 22-25, Seattle, WA, USA, p. 168-177.
- Karagiannis, D.; Höfferer, P., 2006. Metamodels in Action: An overview. *Proceedings of the First International Conference on Software Paradigm Trends (ICSOFT)*, INSTICC Press, Setúbal Portugal, September 11-14. ISBN: 972-8865-69-4
- Khan, K., Baharudin, B. B., Khan, A. and Fazal\_e\_Malik, 2010. Automatic Extraction of Features and Opinion-Oriented Sentences from Customer Reviews. *World Academy of Science, Engineering and Technology*,

- Issue 62, February. ISSN:1307-6892.
- Liu, B., 2010. Sentiment Analysis and Subjectivity. *In Handbook of Natural Language Processing, Second Edition*, Eds: N. Indurkha and F. J. Damerou), CRC Press, Taylor and Francis Group, Boca Raton, FL. Chapter 28.
- Martin, J. R. and White, P. R. R., 2007, *The Language of Evaluation, Appraisal in English*. Palgrave Macmillan, First edition, London & New York, 256 pages.
- OMG, 1999. Unified Modeling Language Specification version 1.3. Technical Report, Object Management Group.
- Schmidt, D. C., 2006. Model-driven Engineering. *IEEE Computer* 39(2), February, p. 25–31.
- Synthesio, 2012. Synthesio. Available at <http://synthesio.com>. Last access: April 30, 2012.
- SocialMetricx, 2012. Socialmetrix - Social Media Analytics for serious decision making. Available at <http://www.socialmetrix.com>. Last access: January 6, 2012.
- Tromp, E., 2011. Multilingual Sentiment Analysis on Social Media. Master's Thesis. Department of Mathematics and Computer Science, *Eindhoven University of Technology*.
- Turney, P., 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews. *Proceedings of the 40<sup>th</sup> Annual Meeting of the ACL*, Philadelphia, July, p. 417-424.
- van Gigch, J. P., 1991. *System Design Modeling and Metamodeling*. Plenum, First edition. July, 453 pages. ISBN: 0306437406.
- Whitelaw, C.; Garg, N. e Argamon, S., 2005. Using Appraisal Groups for Sentiment Analysis. *In Proceedings of the 14<sup>th</sup> ACM International Conference on Information and Knowledge Management (CIKM '05)*. ACM, New York, NY, USA, p. 625–631.
- Wilson, T., 2008. Fine-Grained Subjectivity Analysis. PhD Dissertation, Intelligent Systems Program, *University of Pittsburgh*.