An Ontology-Based Opinion Mining Approach for the Financial Domain

Juana María Ruiz-Martínez, Rafael Valencia-García,
and Francisco García-Sánchez[(✉)]

Facultad de Informática, Universidad de Murcia,
Campus de Espinardo, 30100 Espinardo (Murcia), Spain
{jmruymar, valencia, frgarcia}@um.es

Abstract. Opinion mining is a sub-discipline of computational linguistics that uses information retrieval techniques in order to determine whether a piece of text expresses a positive, negative or neutral opinion. In this paper, we present an approach for the opinion mining of financial news through the process of identifying their semantic polarity. Our approach relies on an algorithm that combines several gazetteer lists and leverages an existing financial ontology. The financial-related news are obtained from RSS feeds and then automatically annotated with positive or negative markers. The outcome of the process is a set of news organized by their degree of positivity and negativity. The preliminary experimental results seem promising as compared against traditional approaches.

Keywords: Opinion mining · Sentiment analysis · Financial news · Ontologies · Semantic web

1 Introduction

User participation is the primary driver of value in Web 2.0 applications. Such a simple idea has had a tremendous impact on the way in which users interact with the Web. While in traditional Web 1.0 sites, companies published content and users were just mere information consumers, in the Web 2.0 era users play a more active role in Web interactions, becoming not only consumers but also producers of information and media. In this new context, namely the Social Web, a key challenge is to understand the opinions and sentiments, not only of the general public and consumers, but also of companies, banks, and politics. Opinion mining is a recent sub-discipline at the crossroads of information retrieval and computational linguistics. The focus of opinion mining does not concern what topic a text is about, but rather what opinion that text expresses [1]. It determines whether the news and the comments in online forums, blogs or comments relating to a particular topic (financial asset, product, book, movie, etc.) are positive, negative or neutral.

Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online texts [2]. In its conception, sentiment analysis referred to the processing of products' attributes from product reviews [3–5]. Nowadays, sentiment polarity analysis methods are employed in a variety of application domains including politics, sports, media, and finances [6–8]. A vast amount of

financial news is published regularly on the Web and the financial domain is constantly changing and evolving. Under these circumstances many authors emphasize the need for mechanisms to automatically extract sentiments from news rather than relying on the intuition of analyst as to what is good or bad news [6]. However, when dealing with automatic sentiment polarity analysis, existing techniques that involve checking the similarity between a text and a seed list of words are not sufficient. In this context, we believe that the already mature Semantic Web technology may be a valuable addition to traditional approaches.

The Semantic Web can be seen as an extension of the current Web, in which information is given a well-defined meaning, thus better enabling computers and people to work in cooperation [9]. The Semantic Web vision is based on the idea of explicitly providing the knowledge behind each Web resource in a manner that is machine processable. Ontologies [10] constitute the standard knowledge representation mechanism for the Semantic Web and can be used to structure information. The formal semantics underlying ontology languages enables the automatic processing of the information in ontologies and allows the use of semantic reasoners to infer new knowledge. In this work, we propose a semantically-enhanced mechanism for opinion extraction from natural language texts. The proposed algorithm has been extensively validated through thorough test-bed experiments in the financial domain. The methodology presented here is supported by natural language processing techniques capable of semantically annotating financial news texts complying with a financial ontology. The annotated financial news items are then further analyzed in order to group them into two separate sets, one with positive financial news and the other with negative financial news.

The rest of paper is organized as follows. Section 2 presents the technological background necessary for the development of the methodology. In Sect. 3, the platform and the way it works is described in detail. In Sect. 4, the experimental results of the evaluation are shown. Related work and the discussion are included in Sect. 5. Finally, conclusions and future work are put forward in Sect. 6.

2 Technological Background

The methodology proposed here is based on two main elements, namely, ontologies and natural language processing tools. In this section, the key features of these technologies are pointed out.

2.1 Ontologies and the Semantic Web

Ontologies constitute the standard knowledge representation mechanism for the Semantic Web [10]. The formal semantics underlying ontology languages enables the automatic processing of the information and allows the use of semantic reasoners to infer new knowledge. In this work, an ontology is seen as "a formal and explicit specification of a shared conceptualization" [10]. Ontologies provide a formal, structured knowledge representation, and have the advantage of being reusable and shareable. They also provide a common vocabulary for a domain and define, with different

levels of formality, the meaning of the terms and the relations between them. Knowledge in ontologies is mainly formalized using five kinds of components: classes, relations, functions, axioms and instances [11].

Ontologies are thus the key for the success of the Semantic Web vision. The use of ontologies can overcome the limitations of traditional natural language processing methods and they are also relevant in the scope of the mechanisms related, for instance, with Information Retrieval [12], Semantic Search [13], Service Discovery [14] or Question Answering [15].

Next, the financial ontology that has been developed for the purposes of this work is described.

2.1.1 Financial Ontology

The financial domain is becoming a knowledge intensive domain, where a huge number of businesses and companies hinge on, with a tremendous economic impact in our society. Consequently, there is a need for more accurate and powerful strategies for storing data and knowledge in the financial domain. In the last few years, several finances-related ontologies have been developed. The BORO (Business Object Reference Ontology) ontology is intended to be suitable as a basis for facilitating, among other things, the semantic interoperability of enterprises' operational systems [16]. On the other hand, the TOVE ontology (Toronto Virtual Enterprise) [17], developed by the Enterprise Integration Laboratory from the Toronto University, describes a standard organization company as their processes. A further example is the financial ontology developed by the DIP (Data Information and Process Integration) consortium, which is mainly focused on describing semantic web services in the stock market domain [18]. Finally, the XBRL Ontology Specification Group, developed a set of ontologies for describing financial and economic data in RDF for sharing and interchanging data. This ontology is becoming an open standard means of electronically communicating information among businesses, banks, and regulators [19].

As part of this work, a financial ontology has been developed on the basis of the above referred ontologies, with the focus set on the stock exchange domain. The ontology, created from scratch, has been defined in OWL 2. This ontology covers three main financial concepts (see Fig. 1):

– A financial market is a mechanism that allows people to easily buy and sell financial assets such us stocks, commodities and currencies, among others. The main stock markets such as New York Stock Exchange, NASDAQ or London Stock Exchange have been modelled in the ontology as subclasses of the Stock_market class.
– The Financial Intermediary class represents the entities that typically invest on the financial markets. Examples of such entities are banks, insurance companies, brokers and financial advisers.
– The Asset class represents everything of value on which an Intermediary can invest, such as stock market indexes, commodities, companies, currencies, to mention a few. So, for instance, enterprises such as Apple Inc., General Electric or Microsoft belong to the Company concept and currencies such as US dollar or Euro are included as individuals of the Currency concept.

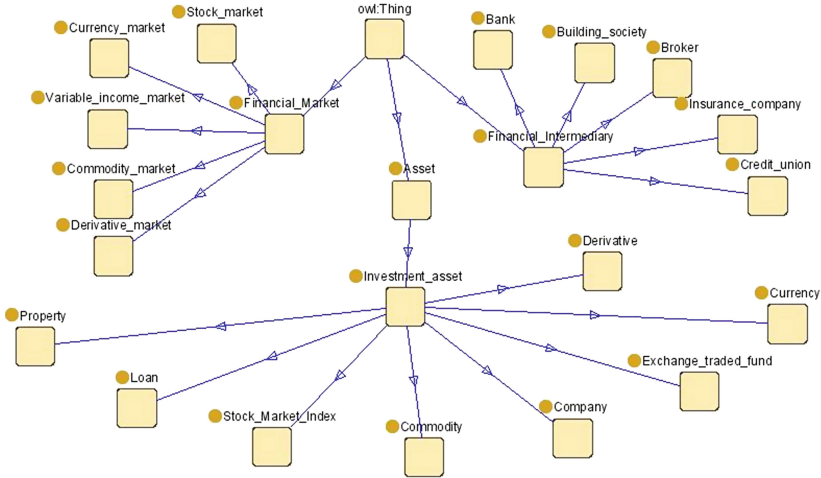**Fig. 1.** An excerpt of the financial ontology

2.2 Natural Language Processing and Sentiment Analysis

Sentiment annotation can be seen as the task of assign positive, negative or neutral sentiment values to texts, sentences, and other linguistic units [20]. In this work, the values positive, negative and neutral have been assigned to general terms, which express some kind of sentiment (e.g. '*benefit*', '*positive*', '*danger*') and to financial terms (e.g. '*risk capital*', '*rising stock*', '*bankruptcy*'). Moreover, terms pertaining to the financial domain have been semantically annotated as '*risk premium*', '*capital market*' or '*Ibex35*' for example.

The open source software GATE[1] carries out sentiment and semantic annotation by means of gazetteers lists. GATE is an infrastructure for developing and deploying software components that process human language. One of the GATE's key components is gazetteer lists. A gazetteer list is a plain text file with one entry (a term, a number a name, etc.), which permits to identify these entries in the text. In this work, the lists have been developed using BWP Gazetteer.[2] This plugin provides an approximate gazetteer for GATE, based on Levenshtein's Edit Distance for strings. Its goal is to handle texts with noise and errors, in which GATE's default gazetteers may have difficulties. The implemented lists are based on the linguistic particularities of the financial domain.

Grishan and Kittredge [21] define a sublanguage as the specialized form of a natural language that is used within a particular domain or subject matter. A sublanguage is characterized by a specialized vocabulary, semantic relationships, and in many cases specialized syntax [22]. The boundaries of financial news domain are not very sharply defined [22]. For example, "*Euribor rates rise after ECB interest warnings*" or "*Portugal needs the luck of Irish*" are both headline of financial news, although the

---

[1] http://gate.ac.uk/.

[2] http://gate.ac.uk/gate/doc/plugins.html#bwp.

second one does not contain any financial term or a particular syntactic structure. Nevertheless, it is possible to define a wide set of financial specialized vocabulary (e.g. '*Euribor*', '*Ibex35*', '*investors*') which coexists with frequently used non-specialized terms (e.g. '*to rise*', '*unemployed*', '*construction*').

In this work, the semantic and sentiment gazetteers developed are employed to mark up all sentiment words and associated entities in our ontology. Six different kinds of gazetteers have been developed on the basis of the common characteristics and vocabulary of financial domain. The lists are used by the system in order to create three different types of annotations, that is, semantic annotations, sentiment annotations and modifier annotations. Semantic annotation refers to financial terms that are present in the financial ontology. Sentiment annotation indicates the polarity of selected terms. Modifiers annotation refers to elements that can invert or increase the polarity of the previously annotated terms. For each kind of annotation a gazetteer category has been created. Thus, semantic, sentiment and modifiers gazetteers have been developed. Each gazetteer category consists of one or more gazetteer lists, as explained below.

1. **Semantic gazetteer**
   a. Financial domain vocabulary gazetteer. This gazetteer contains the most relevant domain terms and entities. It has been directly mapped onto the ontology classes and individuals and their corresponding labels including synonyms. Examples in this category are '*Annual Percentage Rate*' (APR), '*Compound Interest*', '*Dividend*', '*Income Tax*', '*Apple*' and '*BBVA*'. This list is used for the semantic annotation and it does not contain any information related with opinions.

2. **Sentiment gazetteer**
   b. Positive sentiment gazetteer. It contains general terms that imply a positive opinion such as, for example, '*growth*', '*trust*', '*positive*' or '*rising*'.
   c. Negative sentiment gazetteer. It contains general terms that imply a negative opinion such as, for example, '*danger*', '*doubts*' or '*to cut*'.
   d. Financial positive sentiment gazetteer. It contains terms related to the financial domain that imply a positive opinion. For example, '*earning*', '*profitability*' or '*appreciating asset*'.
   e. Financial negative sentiment gazetteer. It contains terms related to financial domain that imply a negative opinion. For example, '*depreciation*', '*Insufficient Funds*' or '*creditor*'.

3. **Modifier gazetteer**
   f. Intensifier gazetteer. It contains terms that are used to change the degree to which a term is positive or negative such as, for example, '*very*', '*most*' or '*extremely*'.
   g. Negation gazetteer. It contains negation expressions such as, for example, '*no*', '*never*' or '*deny*'.
   h. Temporal sentiment gazetteers. They contain temporal expressions that imply a modification in the whole news. These expressions appear in conjunction with positive or negative linguistic expressions modifying their meaning. They usually increase or decrease negative or positive sentiment. There are two temporal gazetteers, one with long-term expressions and the other with short-term expressions. "*Last year*", "*trimester*" or "*several weeks*" are examples of the first

type, while "*this morning*", "*today*" "*this week*" are examples of the second type. The following sentences show an example of the modification capacity of temporal terms in the financial domain:

(1) Apple shares have risen around 17 % in the last month.

(2) Apple shares have fallen 4.5 % this morning.

Here, "*last month*" and "*this morning*" can relativize the weight of the global meaning. In general, long-term positive or negative opinions are more reliable than short-term opinions. That is, if the user searches for the general status of Apple shares and the system retrieves these two entries, then the general opinion should be positive.

3   Platform Architecture

The architecture of the platform is shown in Fig. 2. The architecture is composed of four main components: the financial news extraction module, the semantic annotation module, the opinion-mining module and the search engine. Next, these components are described in detail.
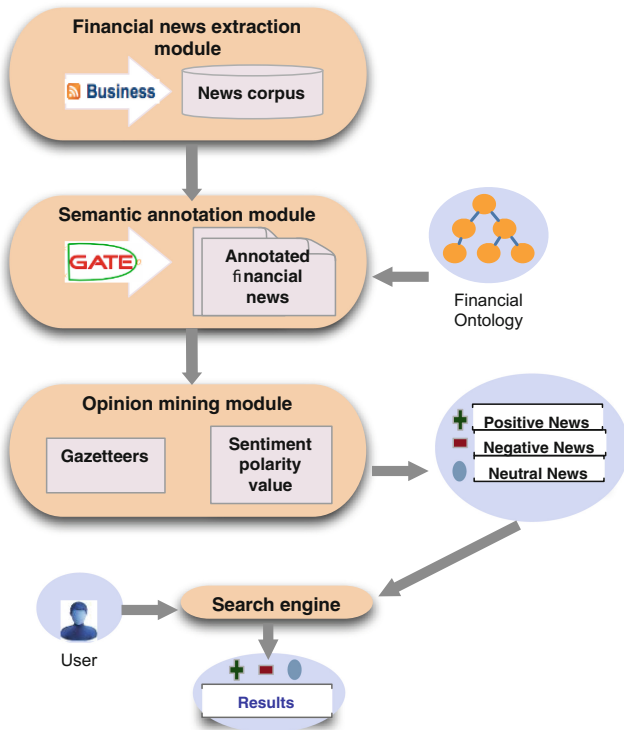
Fig. 2. Architecture of the system.

3.1 Financial News Extraction Module

This module manages the list of RSS feeds. RSS is a family of Web feed formats used for syndicating content from blogs or Web pages and is commonly used by newspapers. RSS is an XML file that summarizes information items and links to the information sources [23]. Once the resources have been selected, this module generates a set of abstracts, which will be used as input for the system. An example list of financial news-related RSS feeds is shown in Table 1.

Table 1. Example of RSS feeds

http://money.cnn.com/services/rss/
http://europe.wsj.com/xml/rss/3_7481.xml
http://www.ft.com/rss/home/europe
http://www.ibtimes.com/rss/

For each RSS source the last news are obtained and stored in a database. The information that is retrieved from each news is the date of publication, the information source, the url and the abstract. Abstracts constitute the corpus from which the system extracts the information. We only consider the abstract and the headline because they usually condense the polarity of news. Indeed, the analysis of the whole text can induce to error, since the sentiment polarity of an entire document is not necessarily the sum of its parts.

3.2 Semantic Annotation Module

This module identifies the most important linguistic expressions in the financial domain using the previously described semantic gazetteer. For each linguistic expression, the system tries to determine whether the expression under question is an individual of any of the classes of the domain ontology. Next, the system retrieves all the annotated knowledge that is situated next to the current linguistic expression in the text, and tries to create fully-filled annotations with this knowledge.

Each class in the ontology is defined by means of a set of relations and datatype properties. Then, when an annotated term is mapped onto an ontological individual, its datatype and relationships constitute the potential information which is possible to obtain for that individual. For example, a company has associate relationships such as '*Moody'sRate*', '*tradeMarket*' or '*isLegalRepresentativeFor*'. In Fig. 3, an example of the annotation process of financial news using the semantic gazetteer in GATE is depicted.

3.3 Opinion Mining Module

The main objective of this module is to classify the set of news obtained in the previous module according to their polarity: positive, negative or neutral. For any retrieved news which has been annotated, the sentiment orientation or sentiment polarity value is computed. For this, the module makes use of the previously described gazetteer lists.
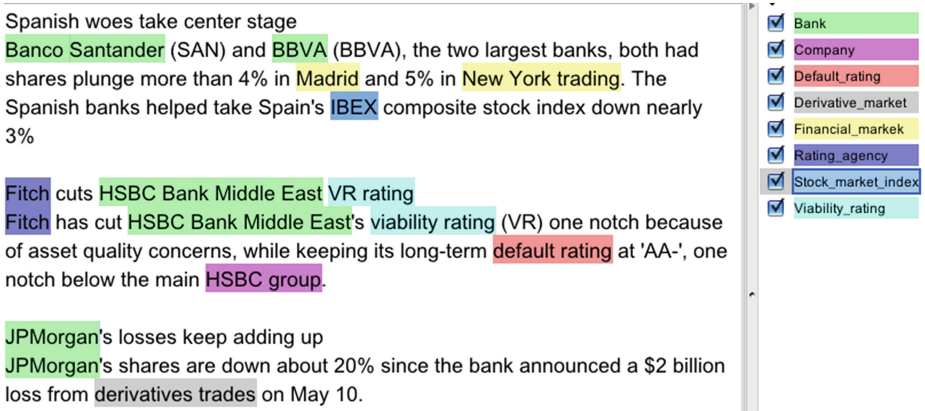
Fig. 3. Example of knowledge entities identified in financial news.

The sentiment polarity (SP) value for each news item is calculated by summing the polarity values of all annotated terms in the news. In this process, the system must consider both the terms polarity included in the positive and negative gazetteers and the contextual valence shifters included in the negation and intensifier gazetteers.

For any annotated term (**at**) in a sentence s $\in$ S, its SP value (**SP(at)**) is computed as follows:

1. If $at \in$ GeneralPositive^k, SP(at) = Positive1
2. If $at \in$ DomainPositive^k, SP(at) = Positive2
3. If $at \in$ GeneralNegative^k, SP(at) = Negative1
4. If $at \in$ DomainNegative^k, SP(at) = Negative2
5. If within the relevant cotext of at, there is a term $at' \in$ Negation, SP(at) = −SP(at)
6. If within the relevant cotext of at, there is a term $at' \in$ Intensifier, SP(at) = 2 × SP(at)
7. When within the relevant cotext of at, there is a term $at' \in$ Temporal, if …
7.1. $at' \in$ LongTerm, SP(at) = 2 × SP(at)
7.2. $at' \in$ ShortTerm + Negative(SP), SP(at) = 2 × SP(at)
7.3. $at' \in$ ShortTerm + Positive(SP), SP(at) = 1 × SP(at)

Then the polarity of each news item is represented as the sum of all SP(at) present in such news item (n):

$$f^k SP(n)^k = \sum_{at \in n} SP(at) \qquad (1)$$

In the above algorithm, the term '*cotext*' refers to the linguistic set that surrounds an annotated term within the limit of a sentence, i.e. the rest of annotated terms present before and after it and pertaining to the same sentence. '*Positive1*' and '*Positive2*' refer to the degree of positivity of an annotated term, while '*Negative1*' and '*Negative2*' refer to the degree of negativity of an annotated term.

When a long-term temporal expression is found, its value is calculated taking into account the *at* pertaining to its cotext. If a positive *at* is found, then its value is 2. On the contrary, if a negative *at* is found its value is −2. Sort- term temporal expressions are calculated in the same way for negative value, i.e. adding −2. However, for positive value the system only adds 1positive. This is because we consider that financial short-term positive values change too frequently to consider them at the same level as long-term values.

Next, if the semantic polarity value of a news is less than 0, the news is labelled as negative. In contrast, if the value is higher than 0, the news is labelled as positive. Finally, if the sum of all values is 0 the news is labelled as neutral. An example of how the algorithm works is shown in Fig. 4.

Fig. 4. Semantic polarity annotation example

Let us suppose that a user searches for the company '*Adidas*'. In the example depicted in Fig. 4, four different news items are retrieved. In the figure, semantic annotations are the elements surrounded by a rectangle, which have been mapped onto ontology instances. GeneralPositive are indicated with one '+' sign and DomainPositive with two, '++'. On the other hand, GeneralNegative are indicated with one '−' sign and DomainNegative with two, '−'. The modifiers Negative, Temporal and Intensifier are indicated with 'N', 'T', 'I' respectively, together with the corresponding positive or negative symbol.

The outcome of the process is three positive and one negative news items. In this particular example, the presence of long-term temporal expressions, such as '*2012*' or '*year*', in conjunction with positive annotated terms, gives to the news a high positive value. The user can organize the final results in accordance with their degree of positivity and negativity.

3.4   Semantic Search Engine

In OWL-based ontologies, '*rdfs:label*' is an instance of '*rdf:property*' that may be used to provide a human readable version of a resource name. In this work, all the resources in the ontology have been annotated with the '*rdfs:label*' descriptor. By considering that, the main objective of this module is to identify the financial news items that are related to the query issued by a user. Besides, this module is responsible for classifying and sorting the results in accordance with the sentiment classification that was described in the previous section.

The system is constantly crawling news information from RSS feeds and creating semantic annotations for the news pages. If no annotations are created for a news item, then such news item is not stored in the database. On the other hand, the news items that have been successfully annotated are processed to obtain their sentiment classification, which is also stored in the database. For example, let us suppose that the ontology contains the taxonomy presented in Fig. 3. There are two kinds of companies, namely, "Energy company" and "ICT company". Each of these classes contains a set of individuals such as "Microsoft" and "GE energy", respectively. If the user is searching for news about "Microsoft", the system will certainly return all the news annotated with the individual Microsoft. Moreover, news related to other ICT companies could be relevant to the user, so the system also shows other news about companies such as Google, Apple and Nokia. If the user queries the system for "Energy companies", then the result will include all the news that contains the concept "Energy company" and therefore the news related to the "GE Energy", "Texaco" and "Shell" companies will be retrieved. Furthermore, if the query is such a general word as "Company", the user is given the possibility of filtering the results according to the subclasses of "Company", namely, "Energy company" and "ICT company".

4   Evaluation

In this section, the experimental results obtained by the proposed method in the financial news domain are presented. The corpus of the experiment contains 57.210 words and comprises 900 abstracts of financial news (512 negative and 388 positive). This corpus has been extracted from the RSS feeds shown in Table 1 and each news item has been manually labelled, either as a positive news or a negative one, by two different annotators. This constitutes the baseline for the evaluation, which works as follows: if the result displayed by the system fits in with the manually annotated news, the result is considered correct, otherwise, incorrect. In the sentiment analysis field, it is agreed that human-based annotations are around 70–80 % precise (i.e. 2 different humans can disagree in 20–30 % of cases). However, for the purposes of this experiment, the news items that have been source of disagreement between annotators have been removed.

In the experiment, a total of five queries are issued to the system to find information in the financial domain. The results of the experiment are shown in Table 2. It is possible to observe that the sentimental analysis accuracy results are very promising, with an aggregate accuracy mean of 87 %. These results take into account the system's final decision (positive or negative) and not the process that the system carries out to produce such decision.

Table 2. Hits results in information retrieval.

Query		Baseline	Our approach	Accuracy
1	Pos	33	28	84.85 %
	Neg	11	9	81.82 %
	Total	44	37	84.09 %
2	Pos	13	13	100 %
	Neg	36	34	94.44 %
	Total	49	47	95.92 %
3	Pos	15	14	93.33 %
	Neg	29	24	82.76 %
	Total	44	38	86.36 %
4	Pos	25	21	84 %
	Neg	97	86	88.66 %
	Total	122	107	87.70 %
5	Pos	66	55	83.33 %
	Neg	14	12	85.71 %
	Total	80	67	83.75 %
Total		678	592	87.32 %

5 Related Works and Discussion

In the literature, a number of methods for the automatic sentiment analysis from financial news streams have been described. The proposal of [7] uses theories of lexical cohesion in order to create a computable metric to identify the sentiment polarity of financial news texts. This metric is readapted in [6] to Chinese and Arabic financial news. The analysis of financial news is a particularly relevant topic in the prediction of the behavior of stock markets. For example, in [8] the authors use some simple computational linguistic techniques, such as bag of words or named entities, together with support vector machine and machine learning techniques to assist in making stock market predictions. In fact, in real life, stock market analysts' predictions are usually based on the opinions expressed in the news. Our approach does not focus solely on the stock market. It covers a wider range of financial news.

Semantic technologies have been around for a while, offering a wide range of benefits in the knowledge management field. They have revolutionized the way that systems integrate and share data, enabling computational agents to reason about information and infer new knowledge [10]. The accuracy results of opinion mining and sentiment polarity analysis can be improved with the addition of semantic techniques, as shown in [24]. In that work, some semantic lexicons are created in order to identify sentiment words in blog and news corpora. Then, a polarity value is attached to each word in the lexicon and such polarity is revised when a modifier appears in the text. The main problem of corpora-based methods is the cost of the annotation process. A further problem, namely, the obsolescence of resources, is present on constantly changing domains such as the financial domain. This challenge is partially overcome by using bag of words or gazetteer lists which are easy to update if required.

The FIRST project[3] provides an information extraction, information integration and decision making infrastructure for information management in the financial domain. The decision making infrastructure includes a module responsible for the sentiment annotation from financial news and blog posts. Its main aim is to classify the polarity of sentiment with respect to a sentiment object of interest [25]. These sentiment objects are classified by means of an ontology-guided and rule-based information extraction approach. Even though the ontology contains the financial-domain related relevant objects, the classification process is carried out entirely using JAPE rules. This work is the closest to the one presented here, but their approach does not leverage the reasoning capabilities of the ontology. In fact, none of the previous works considers the ontological capabilities in the task of sentiment annotation. Combining the usage of semantic gazetteers in a textual level with the ontology in a conceptual level makes our system adaptable to other languages with minimum cost.

6   Conclusions

The boom of the Social Web has had a tremendous impact on a number of different research topics. In particular, the possibility of extracting various kinds of added-value, informational elements from users' opinions has attracted researchers from the information retrieval and computational linguistics fields. This process is called opinion mining and is currently one of the most challenging research topics. More specifically, opinion mining is concerned with analyzing the opinions of a particular matter expressed by users in the form of natural language that appear in a series of texts. The opinion mining process makes it possible to figure out whether a user's opinion is positive, negative or neutral, and how strong it is [26]. Similarly, sentiment analysis deals with the computational treatment of opinions expressed in written texts.

The addition of the already mature semantic technologies to this field has proven to increase the results accuracy. In this work, a semantically-enhanced methodology for the annotation of sentiment polarity in financial news-related natural language texts is presented. The proposed methodology is based on an algorithm that combines several gazetteer lists and leverages an existing financial ontology. The sentiment algorithm assigns different degrees of positivity or negativity to relevant annotated terms and calculates what the polarity of the news is. The financial-related news items in our experiment are obtained from RSS feeds and then automatically annotated with positive or negative markers. The outcome of the process is a set of financial news texts organized by their degree of positivity and negativity.

This approach contributes to the research on financial sentiment annotation, and the development of decision support systems (1) by proposing a novel approach for financial sentiment determination in news which combines ontological resources with natural language processing resources, (2) by describing an algorithm for assigning different degrees of positivity or negativity to classify results into various categories

---

[3] http://project-first.eu/.

identified by the classifier, and (3) by proposing a set of resources, i.e. gazetteer lists and an ontology, for sentiment annotation.

References

1. Esuli, A., Sebastiani, F.: Sentiwordnet: a publicly available lexical resource for opinion mining. In: Proceedings of 5th Conference on Language Resources and Evaluation, LREC 2006 (2006)
2. Chen, H., Zimbra, D.: AI and opinion mining. Intell. Syst. IEEE **25**(3), 74–80 (2010)
3. Popescu, A.M., Etzioni, O.: Extracting product features and opinions from reviews. In: Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing, Association for Computational Linguistics, pp. 339–346 (2005)
4. Ding, X., Liu, B.: The utility of linguistic rules in opinion mining. In Proceedings of 30th Annual International ACM Special Interest Group on Information Retrieval Conference (SIGIR 2007), Amsterdam, The Netherlands (2007)
5. Balahur, A., Montoyo, A.: Determining the semantic orientation of opinions of products-a comparative analysis. Procesamiento del lenguaje natural **41**, 201–208 (2008)
6. Ahmad, K., Cheng, D., Almas, Y.: Multi-lingual sentiment analysis of financial news streams. In: Proceedings of the Second Workshop on Computational Approaches to Arabic Script-based Languages, Linguistic Society of America, Linguistic Institute, Stanford University, pp. 1–12 (2007)
7. Devitt, A., Ahmad, K.: Sentiment analysis in financial news: a cohesionbased approach. In Proceedings of the Association for Computational Linguistics (ACL), pp. 984–991 (2007)
8. Schumaker, R.P., Chen, H.: Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Trans. Inf. Syst. **27**, 1–19 (2009)
9. Berners-Lee, T., Hendler, J.: Publishing on the semantic web. Nature **410**, 1023–1025 (2001)
10. Studer, R., Benjamins, V.R., Fensel, D.: Knowledge engineering: principles and methods. Data Knowl. Eng. **25**(1–2), 161–197 (1998)
11. Gruber, T.R.: A translation approach to portable ontology specifications. Knowl. Acquis. **5**(2), 199–220 (1993)
12. Valencia-García, R., Fernández-Breis, J.T., Ruiz-Martínez, J.M., García-Sánchez, F., Martínez-Béjar, R.: A knowledge acquisition methodology to ontology construction for information retrieval from medical documents. Expert Syst. Knowl. Eng. J. **25**(3), 314–334 (2008)
13. Lupiani-Ruiz, E., García-Manotas, I., Valencia-García, R., García-Sánchez, F., Castellanos-Nieves, D., Fernández-Breis, J.T., Camón-Herrero, J.B.: Financial news semantic search engine. Expert Syst. Appl. **38**(12), 15565–15572 (2011)
14. García-Sánchez, F., Valencia-García, R., Martínez-Béjar, R., Fernández-Breis, J.T.: An ontology, intelligent agent-based framework for the provision of semantic web services. Expert Syst. Appl. Part 2 **36**(2), 3167–3187 (2009)
15. Valencia-García, R., García-Sánchez, F., Castellanos-Nieves, D., Fernández-Breis, J.T.: OWLPath: an OWL ontology-guided query editor: IEEE Transactions on Systems. Man Cybern. Part A **41**(1), 121–136 (2011)

16. Partridge, C.: The role of ontology in integrating semantically heterogeneous databases. Report No.: LADSEB-CNR Technical Report 05/2002 (2002)
17. Fox, M.S., Gruninger, M.: Enterprise modeling. AI Mag. **19**(3), 109 (1998)
18. Corcho, O., Losada, S., Martínez Montes, M., Bas, J.L., Bellido, S.: Financial Ontology. DIP deliverable D10.3 (2004)
19. Bonsón, E., Cortijo, V., Escobar, T.: Towards the global adoption of XBRL using international financial reporting standards (IFRS). Int. J. Account. Inf. Syst. **10**(1), 46–60 (2009)
20. Andreevskaia, A., Bergler, S.: When specialists and generalists work together: overcoming domain dependence in sentiment tagging. In: Proceedings of ACL 2008: HLT, pp. 290–298 (2008)
21. Grishman, R., Kittredge, R.: Analyzing Language in Restricted Domains: Sublanguage Description and Processing. Lawrence Erlbaum, Hillsdale (1986)
22. Grishman, R.: Adaptive information extraction and sublanguage analysis. In: Kushmeric, N. (ed.) Proceedings of Workshop on Adaptive Text Extraction and Mining at Seventeenth International Joint Conference on Artificial Intelligence, Seattle, WA (2001). http://nlp.cs.nyu.edu/pubs/papers/grishman-ijcai01.pdf
23. Murugesan, S.: Understanding web 2.0. IT Prof. **9**(4), 34–410 (2007)
24. Godbole, N., Srinivasaiah, M., Skiena, S.: Largescale sentiment analysis for news and blogs: In: Proceedings of the International Conference on Weblogs and Social Media (ICWSM) (2007)
25. Klein, A., Häusser, T., Altuntas, O., Grcar, M.: Large scale information extraction and integration infrastructure for supporting financial decision making. Deliverable: D4.1 First semantic information extraction prototype (2012). http://project-first.eu/content/d41-first-semantic-information-extraction-prototype
26. Zhao, L., Li, C.: Ontology Based Opinion Mining for Movie Reviews. In: Karagiannis, D., Jin, Z. (eds.) KSEM 2009. LNCS, vol. 5914, pp. 204–214. Springer, Heidelberg (2009)