



Ontology-based sentiment analysis of twitter posts

Efstratios Kontopoulos^{a,*}, Christos Berberidis^{a,1}, Theologos Dergiades^{a,2}, Nick Bassiliades^{a,b,3}

^a School of Science and Technology, International Hellenic University, Thessaloniki, Greece

^b Department of Informatics, Aristotle University of Thessaloniki, Thessaloniki, Greece

ARTICLE INFO

Keywords:

Micro-blogging
Twitter
Tweet
Sentiment analysis
Ontology

ABSTRACT

The emergence of Web 2.0 has drastically altered the way users perceive the Internet, by improving information sharing, collaboration and interoperability. Micro-blogging is one of the most popular Web 2.0 applications and related services, like Twitter, have evolved into a practical means for sharing opinions on almost all aspects of everyday life. Consequently, micro-blogging web sites have since become rich data sources for opinion mining and sentiment analysis. Towards this direction, text-based sentiment classifiers often prove inefficient, since tweets typically do not consist of representative and syntactically consistent words, due to the imposed character limit. This paper proposes the deployment of original ontology-based techniques towards a more efficient sentiment analysis of Twitter posts. The novelty of the proposed approach is that posts are not simply characterized by a sentiment score, as is the case with machine learning-based classifiers, but instead receive a sentiment grade for each distinct notion in the post. Overall, our proposed architecture results in a more detailed analysis of post opinions regarding a specific topic.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

The emergence of Web 2.0 has drastically altered the way users perceive the Internet, by improving information sharing, collaboration and interoperability. Contrary to the first generation of websites, where users could passively view content, Web 2.0 users are encouraged to participate and collaborate, forming virtual on-line communities. Additional Web 2.0 traits include data openness and metadata, dynamic content, rich user experience and scalability tolerance (Skiba, 2006). Among the most popular Web 2.0 applications, one can come across social-networking sites (e.g. Facebook, MySpace), wikis, blogs, multi-media sharing sites (e.g. YouTube, Flickr), mash-ups and rich web applications. One of these activities is *micro-blogging*, which initially attracted comparatively less attention, but gradually became a highly popular communication tool for a considerable percentage of users.

Micro-blogging is in principle based on blogs (i.e. Web logs), where users can post opinions, experiences and queries on any chosen topic. The main difference between micro- and traditional blogs is the strict constraint in content size (Kaplan & Haenlein,

2011). Currently, the most popular on-line micro-blogging service is *Twitter*, which enables its users to send and receive text-based posts, known as “*tweets*”, consisting of up to 140 characters. Twitter was created in 2006 and currently records over 140 million active users that generate over 340 million tweets per day. With their rapidly increasing popularity, micro-blogging services, like Twitter, have evolved into a practical means for sharing opinions on almost all aspects of everyday life. The strict character limit of tweets forces users to be concise and eventually more expressive than with social networks and blogs. Thus, micro-blogging posts are imbued with emotional information and are considered as rich *opinion mining* data sources (Pak & Paroubek, 2010). Additionally, tweets can be processed more effectively than lengthy blog posts and articles.

Opinion mining, also known as *sentiment analysis*, is the process aiming to determine whether the **polarity of a textual corpus** (document, sentence, paragraph etc.) tends towards **positive, negative or neutral**. Sentiment analysis constituted a popular research area even before Twitter and micro-blogging, since it can offer advantages to a variety of domains, from sales predictions (Liu, Huang, An, & Yu, 2007) to politics (Park, Ko, Kim, Liu, & Song, 2011) and investors’ choices (Dergiades, 2012). Furthermore, the automatic detection of sentiment on textual corpora has comprised a research topic for many approaches. Examples include, among others, product and services reviews (Kang, Yoo, & Han, 2012; Prabowo & Thelwall, 2009), articles on the Web (Melville, Gryc, & Lawrence, 2009) and news feeds (Moreo, Romero, Castro, & Zurita, 2012; Wanner, Rohrdantz, Mansmann, Oelke, & Keim, 2009). However, the

* Corresponding author. Tel.: +30 2310 807538.

E-mail addresses: e.kontopoulos@ihu.edu.gr (E. Kontopoulos), c.berberidis@ihu.edu.gr (C. Berberidis), t.dergiades@ihu.edu.gr (T. Dergiades), nbassili@csd.auth.gr (N. Bassiliades).

¹ Tel.: +30 2310 807534.

² Tel.: +30 2310 807501.

³ Tel.: +30 2310 998913.

expressive power and immediacy of Twitter in particular has led researchers to try utilizing it in further approaches related to politics (Tumasjan, Sprenger, Sandner, & Welpe, 2010), tourism (Claster, Cooper, & Sallis, 2010) and many more.

One of the most promising applications of sentiment analysis of tweets could be in the field of economic and financial modeling. Econometric specifications that do not incorporate variables, such as the investors' or consumers' sentiment, prove to be inefficient. However, the real-time discovery of individuals' sentiment is a challenging task that often involves analysts digging manually through virtually infinite articles and news feeds. Twitter analysis offers one of the best solutions towards the automatic discovery of sentiment. Machine learning-based sentiment classifiers constitute a prevailing sentiment analysis tool. However, they can often prove less efficient in the case of tweets (Barbosa & Feng, 2010; He & Alani, 2011), since the latter do not typically consist of representative and syntactically consistent words, due to the character restriction. An additional limitation is that classifiers usually distinguish sentiment into classes (positive, negative and neutral), assigning a corresponding score to the post as a whole, regardless the fact that many aspects of the same "notion" may be discussed in a single post. Consider, for example, the sample tweet T_{ex} : "The screenplay was wonderful, although the acting was rather bad". The machine-learning based approaches would return a single quantitative (sentiment score) or qualitative (positive, negative or neutral) result. In this paper, we propose the deployment of ontology-based techniques towards a more fine-grained sentiment analysis of Twitter posts. According to the proposed approach, tweets are not simply characterized by a sentiment score, but instead receive a sentiment grade for each distinct notion in the post. This results overall in a more elaborate analysis of post opinions regarding a specific topic. More specifically, regarding the sample tweet T_{ex} above, our proposed ontology-based approach distinguishes the features of the domain (in this case, screenplay and acting) and assigns respective scores, resulting in a more detailed sentiment analysis of the given statement.

The rest of the paper consists of the following: Section 2 reports on the related research efforts focusing on sentiment analysis on micro-blogging data, followed by a section elaborating on the deployment of ontologies in the micro-blogging domain. Section 4 explains the proposed sentiment analysis approach, outlining the distinct phases of the methodology and also includes a baseline scenario that better illustrates the whole process. Finally, the paper concludes with some final remarks as well as directions for future work.

2. Sentiment analysis in micro-blogging data

Researchers in their effort to perform sentiment analysis on micro-blogging posts initially applied mainstream methodologies, used for analyzing the sentiment of "normal" textual corpora, like e.g. product reviews. More specifically, two main approaches were commonly used, in order to conclude, whether a piece of text (sentence, paragraph or document) expresses a positive or negative sentiment: the *lexicon-based* and the *machine learning-based* approach. The former approach (Kaji & Kitsuregawa, 2007; Neviarouskaya, Prendinger, & Ishizuka, 2009; Taboada, Brooke, Tofiloski, Voll, & Stede, 2011) is based on *opinion words*, namely, words that are commonly used in expressing positive or negative sentiment. Opinion words are typically contained in a dictionary called *opinion lexicon*. However, tweets are not considered "normal" pieces of text, since the 140-character threshold imposes limitations in the length of words and phrases. A further peculiarity is the extensive usage of "every-day" (i.e. jargon) expressions, abbreviations and emoticons (sequences of symbols representing an emotion). One could suggest adding these words and symbols to

the lexicon, but this would nevertheless generate dubious results. These expressions are usually of a dynamic nature, changing constantly and being replaced frequently, following each time the popular trends on the Web. An additional disadvantage is the fact that jargon expressions are often domain dependent. These factors lead to low recall, when the lexicon-based method is applied on "informal" corpora of text, like posts from micro-blogs.

An alternative approach is the application of machine learning methodologies (Pang & Lee, 2008). According to this approach, a sentiment classifier is trained, in order to distinguish positive, negative and neutral sentiments in textual corpora. Typical features used in training the classifiers are unigrams or bigrams (n -grams of size 1 and 2, respectively (De Kok & Brouwer, 2012)). The drawback of the machine learning-based methods is mainly focused on the manual labeling required over (usually) massive sets of tweets. Additionally, the labeling has to be performed on each distinct domain of interest, in order to achieve satisfactory training levels for the classifier on the given domain (Aue, 2005).

As outlined in (Saif, He, & Alani, 2012), in relation to sentiment analysis which is derived from Twitter posts, the following approaches can be distinguished:

1. Working with noisy labels or distant supervision (Read, 2005), by creating, for example, emoticon vocabularies for representing sentiment and for training supervised sentiment classifiers, such as Naïve Bayes (NB), Maximum Entropy (MaxEnt) and Support Vector Machines (SVMs) (Barbosa & Feng, 2010; Pak & Paroubek, 2010).
2. Applying a combination of feature engineering (e.g. using the feature-based model and the tree kernel-based model for sentiment classification, as well as n -gram and lexicon features) with machine learning methods to improve sentiment classification accuracy on tweets (Agarwal, Xie, Vovsha, Rambow, & Passonneau, 2011; Kouloumpis, Wilson, & Moore, 2011).

As described in the next section, there already exist sentiment analysis methods that deploy ontology-based techniques. Nevertheless, none of the existing approaches applies ontologies similarly to the way we propose in this work that results in specifically evaluating the sentiment for the various distinct notions (i.e. properties) at hand.

3. Micro-blogging and ontologies

An ontology can be defined as an "explicit, machine-readable specification of a shared conceptualization" (Studer, Benjamins, & Fensel, 1998). Ontologies are used for modeling the terms in a domain of interest as well as the relations among these terms and are now applied in various fields, like agent and knowledge management systems and e-commerce platforms (Gómez-Pérez & Corcho, 2002). Other applications include natural language generation, intelligent information integration, semantic-based access to the Internet and extracting information from texts. However, the most important contribution of ontologies is the key role they play in the development of the *Semantic Web*.

The Semantic Web is an extension of the current Web, where information is given a well-defined meaning, encouraging cooperation among human users and computers (Berners-Lee, Hendler, & Lassila, 2001). Ontologies serve as the primary means of knowledge representation in the Semantic Web. Although various ontology languages have emerged, the currently dominant standards are RDF/S (*Resource Description Framework Schema*) and OWL (*Web Ontology Language*). Additional points of motivation for preferring the use of ontologies in an application include: (a) analyzing domain knowledge and separating the latter from operational

knowledge, (b) Enabling the reuse of domain knowledge, (c) Making domain assumptions explicit, and (d) Sharing a common understanding of the information structure among people and/or software agents.

Regarding the deployment of ontologies in the micro-blogging domain, to the best of our knowledge, the most prominent effort belonging to this category is the recent work by Iwanaga, Nguyen, Nakagawa, and Ohsuga (2011). In their approach, the authors present a methodology for populating an existing earthquake evacuation ontology with instances based on tweets. The proposed approach extracts related information like evacuation center names, products offered at the centers and the timestamp of each tweet. Additional information retrieved from the Web is appended, including the evacuation center address (retrieved via Google Maps), the center's latitude and longitude (via Geocoding) and Japanese-to-English translation of all the above (via Google Translation). This information is obtained in real time, even though it is not included in the initial tweet.

Other directions of research include the development of ontologies for representing micro-blog posts and relationships between social-network users as for example FOAF (see Brickley & Miller, 2010), SIOC (see Breslin, Harth, Bojars, & Decker, 2005), OPO (see Stankovic, Passant, & Laublet, 2009), SMOB2 (see Passant, Bojars, Hastrup, & Laublet, 2010), or ontologies for representing levels of emotions (e.g. Baldoni, Baroglio, Patti, & Rena, 2012; Francisco, Germans, & Peinado, 2007). These topics, however, are unrelated to the line of research presented in this paper. Our approach assumes a different direction as far as the usage of ontologies is concerned. While Iwanaga et al. (2011) deploy the ontology as a means for modeling the application domain, we extend the usability of the ontology in our approach. The domain of reference is semantically divided and sentiment scores are assigned not to whole statements (i.e. tweets), but to the various aspects of the topic at hand. For example, according to “traditional” sentiment analysis techniques, the sample tweet T_{ex} presented in the introduction would probably acquire a positive score, since “wonderful” is rather stronger than “bad”. However, according to our proposed ontology-based approach, the subject (movie) would have two distinct features (screenplay and acting) with respective scores 8 and –3, providing higher granularity in the sentiment analysis of the given statement.

4. Description of the proposed approach

As already explained, the basic idea behind the proposed approach is to take advantage of a domain ontology for providing more elaborate sentiment scores regarding the notions contained in a tweet. The aim is to have a system that accepts as input a tweet (or a set of tweets) regarding a specific subject and provides sentiment scores for every aspect/feature of this subject. The architecture of the system we developed is presented in detail in a following section.

The proposed methodology is divided in two phases: (a) creation of the domain ontology (Section 4.1), and (b) sentiment analysis on a set of tweets, based on the concepts and properties included in the ontology (Section 4.2). These two phases are further described next.

4.1. Creating the domain ontology

In order to create a domain ontology, one can adopt various methods, like extending existing ontologies or developing the ontology from the ground up. In this work, we examine two alternative approaches, which are presented subsequently: (a) Formal Concept Analysis, and (b) Ontology Learning.

4.1.1. Formal Concept Analysis

Formal Concept Analysis (FCA) is a mathematical data analysis theory, typically used in Knowledge Representation and Information Management (Ganter & Wille, 1999). Its main characteristic is that it applies a user-driven step-by-step methodology for creating domain models. With the recent emergence of the Semantic Web and the establishing of ontologies as its principal means for knowledge representation (see Section 3), FCA has been accounted as a valuable engineering tool for deriving an ontology from a collection of objects and their properties (Obitko, Snasel, Smid, & Snasel, 2004). Towards this affair, FCA has recently been applied in various occasions (e.g. Fu & Cohn, 2008; Ning, Guanyu, & Li, 2010; Zhang & Xu, 2011) and has been preferred in this work, since it offers the following advantages:

1. *Appropriate ontology size*: The domain ontology is gradually developed, depending on the given data set (i.e. in this case tweets). Therefore, it does not contain unnecessary concepts and/or properties, which would result in redundancy and potential illegibility on behalf of the user. On the other hand, based on the same principle, the output ontology does not lack essential concepts, either.
2. *Better ontology design*: Concepts and concept hierarchies are not explicitly defined, but are dynamically designated via the detected properties. This leads to better ontology design and a more distinct classification of concepts.
3. *Domain specific ontology*: Towards creating a domain ontology, one would reasonably consider utilizing existing ontologies that describe the given domain in more detail. However, the aim is not to exhaustively describe the application domain, but to develop an ontology that corresponds to the given set of tweets, namely, the notions that currently “matter the most” to the users. Otherwise, the result would be an ontology that contains numerous classes and attributes that never appear in the data set and for which no sentiment assessment score can be derived.

4.1.1.1. *FCA basic elements*. The main building block in FCA is the concept, which is described via two sets: The extension, which is a set of objects and the intension, which is a set of attributes (Ganter & Wille, 1999). Every object that belongs to the concept has all the attributes in the intension and every attribute that belongs to the concept is shared by all objects of the extension.

The relationships between the set of objects and the set of attributes are represented by a formal context. A formal context $\mathcal{K}(\mathcal{O}, \mathcal{A}, \mathcal{I})$ is a triple where:

- \mathcal{O} is a set of formal objects,
- \mathcal{A} is a set of attributes, and,
- \mathcal{I} is a binary incidence relation between the objects and the attributes; $\mathcal{I} \subseteq \mathcal{O} \times \mathcal{A}$, where $(o, a) \in \mathcal{I}$ is read as “object o has attribute a ”.

A formal context can be represented as a cross-table, where the rows represent \mathcal{O} , the columns represent \mathcal{A} and the incidence relation \mathcal{I} is represented by a series of crosses as shown in Table 1. Concepts can be organized into a concept lattice, which is based on the mathematical notion of lattices (Davey & Priestley, 2002). Concept lattices are visualized via Hasse diagrams (or line diagrams – see Fig. 5). The nodes in the diagram represent concepts, while attributes and objects are denoted above and below the nodes, respectively. By traversing all the paths leading down from a node, one can retrieve the concept's extension, while the opposite path leading up from the node retrieves the intension.

Table 1
Smartphone ontology.

Model series	Properties								
	Android	Camera	Battery	4G	microUSB	Processor	Windows	Display	iOS
Apple iPhone		+	+			+		+	+
Samsung galaxy	+	+	+	+	+	+		+	
HTC one	+	+	+	+	+	+		+	
Nokia lumia		+	+		+	+	+	+	

Algorithm createOntology

Input: A default concept (c)
Variables: An initially empty set of tweets (W)
 An initially empty set of objects (O)
 An initially empty set of attributes (A)
Output: A cross-table, filled with objects and attributes (T)

1. $W \leftarrow \text{retrieveTweets}(c)$; // automatically choose the first n tweets
2. **foreach** $w \in W$ **do**
3. $o \leftarrow \text{retrieveObject}(w)$;
4. **if** $o \neq \text{NULL}$ **then**
5. $O := O \cup \{o\}$;
6. $A' \leftarrow \text{retrieveAttributes}(w)$;
7. **foreach** $a \in A'$ **such that** $(o, a) \neq \emptyset$ **do**
8. $A := A \cup \{a\}$;
9. $T \leftarrow \text{populateTable}(O, A)$;
10. **return** T

Fig. 1. Ontology creation algorithm via FCA.

4.1.1.2. Populating the concept cross-table. According to FCA, the aim is to create the concept cross-table that corresponds to the domain ontology. As mentioned previously, the table corresponds to the incidence relation $\mathcal{I} \subseteq \mathcal{O} \times \mathcal{A}$; thus, it is essential to determine sets \mathcal{O} and \mathcal{A} . These sets can be derived via the algorithm illustrated in Fig. 1.

1. The algorithm accepts as initial input a concept parameter (e.g. “#smartphone”), determined manually by the user. The *retrieveTweets* method retrieves the first N tweets (default value for N is 100) that belong to the result set corresponding to the concept parameter. Towards this affair, *Twitter4j*⁴ was utilized, a Java library that gives access to the Twitter API and assists in integrating the Twitter service into any Java application. The *retrieveObject* method subsumes that every tweet is inspected for references to objects and, if any such reference is detected, the corresponding attributes are retrieved via method *retrieveAttributes*. For every attribute associated to the object, the attribute is **appended** to the existing set of attributes. The latter two methods are currently performed manually by an ontology engineer. Fig. 2(A) displays two sample tweets used for the creation of the ontology. The detected objects and attributes are represented in boldface fonts.

Eventually, the final concept table is populated with the detected objects and attributes. Table 1 displays the resulting cross-table as well as the corresponding concept lattice, after analyzing 100 retrieved tweets for the concept “smartphone”.

4.1.2. Ontology learning

An alternative to the manual FCA methodology presented above is offered by the various existing semi-automatic *ontology learning*

techniques (e.g. Cimiano & Völker, 2005; Hazman, El-Beltagy, & Rafea, 2009). Ontology learning, also known as *ontology extraction*, *ontology generation* or *ontology acquisition*, refers to the task of automatically creating an ontology, via extracting concepts and relations from a given data set. Nevertheless, the task of **creating an ontology in a fully automated manner** still remains **elusive** to a great degree.

In this work, we resorted to *OntoGen* (Fortuna, Grobelnik, & Mladenic, 2007), a semi-automatic, data-driven ontology editor. The software deploys text-mining techniques via an efficient user interface that **reduces development time and complexity**. Overall, the tool attempts to **bridge the gap between complex ontology editors and domain experts**, who do not necessarily possess ontology engineering skills.

OntoGen interactively offers assistance during the development of domain ontologies, by suggesting concepts and relations and by automatically assigning instances to the concepts. The user can accept or reject the suggestions or perform manual adjustments. Most of the aid provided by the system is based on the provided underlying data. In our case, the role of this data set is played by the initial set of retrieved tweets (see previous subsection), which is fed to OntoGen as a set of named line-documents (i.e. each tweet is stored as a separate text file, with the first word in the line serving as the document title/ID). Fig. 3 illustrates the resulting ontology visualization via OntoGen, after examining the same set of 100 retrieved tweets for the concept “smartphone”.

4.1.3. Augmenting the semantics

The ontology created via FCA (Section 4.1.1) or Ontology Learning (Section 4.1.2) is in essence a simple taxonomy of concepts and attributes. In order to augment the underlying semantics, the ontology is enriched with synonyms and hyponyms (subordinate notions) of the detected attributes. For example, in the

⁴ Twitter4j Java library: <http://twitter4j.org/en/index.jsp>

A	iPhone or Android? Seems the only downside to the iPhone is battery life - but droids can't be much better can they? #Smartphone	
	Samsung Galaxy S3 with 4,8 inch Display ? #Samsung #GalaxyS3 #Smartphone	
B	Best phone camera i have used so far in my life is of Samsung Galaxy S2 fact	2
	So the HTC one has got crap battery life but its brill if you want to take photos while falling out of a plane ? I give up	-8

Fig. 2. (A) Sample tweets used for the creation of the ontology, and (B) Sample tweets used for sentiment analysis, accompanied by the corresponding sentiment scores.

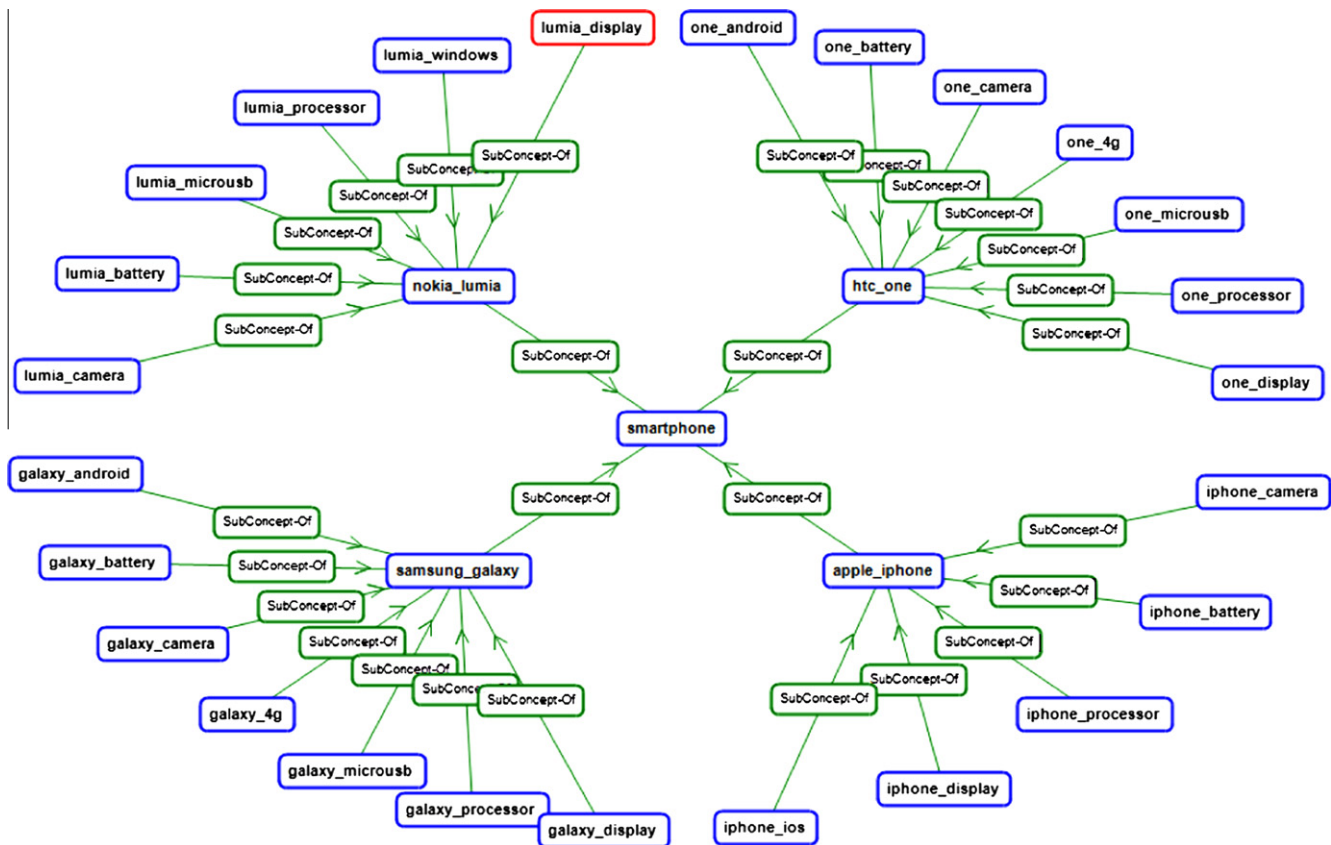


Fig. 3. Ontology visualization via OntoGen.

“smartphone” universe used throughout this paper, the “display” attribute could also be expressed as “monitor” or “screen”, which are synonymous words.

For appending the sets of synonyms and hyponyms to the ontology, we used the popular *WordNet* lexical database (Miller, 1995), which retrieves *synsets* (groups of synonymous words or collocations) of the synonyms and hyponyms of every given word. Each synonym and hyponym is then added to the ontology and associated with the initial attribute. Syntactically, in the OWL DL representation of the ontology, these associations are expressed via the owl:equivalentProperty and rdfs:subPropertyOf constructs, respectively.

4.2. Sentiment analysis on tweets

The previously described process (Section 4.1) results in a formulated and populated domain ontology. The second phase of the proposed methodology constitutes the main effort of this work

and performs the automatic sentiment analysis on a set of tweets. Fig. 4 displays the overall architecture of the approach we propose in this paper.

As can be seen from Fig. 4, the overall process involves retrieving a set of tweets that correspond to entities in the ontology and performing sentiment analysis on each of the retrieved tweets. There are three distinct steps in the procedure: (1) querying the ontology for the corresponding attributes of each object, (2) retrieving the relevant tweets, and (3) performing the sentiment analysis.

4.2.1. Step#1: Taking advantage of the ontology

In order to take advantage of the domain ontology created during the previous steps, the retrieved tweets have to contain information regarding the objects and attributes of reference. This is achieved via JENA (Rajagopal, 2005), a Java API for processing RDF/S and OWL ontologies. Having an ontology-based structured hierarchy of classes and properties, JENA assists in retrieving ob-



Fig. 4. Architecture of the proposed approach.

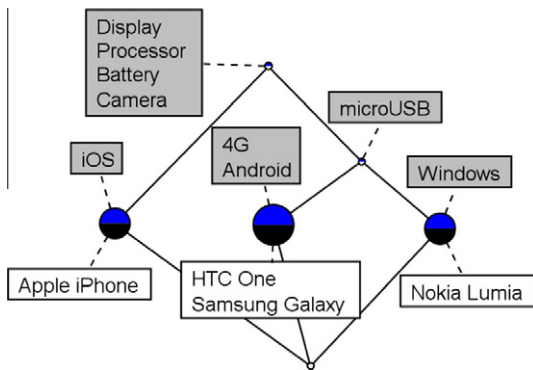


Fig. 5. Ontology visualization via a Hasse diagram.

ject-attribute pairs (o_i, a_{ij}) – see Fig. 4. More specifically, for every object/class o_i , all attributes/properties a_{ij} are retrieved via processing RDF/S triples of the form: $\langle a_{ij} \text{ rdfs:domain } o_i \rangle$.

4.2.2. Step #2: Retrieving the relevant tweets

For every property a_{ij} of an object o_i a relevant query is submitted to Twitter via the Twitter4J library described previously. The query has the form “ $o_i a_{ij}$ ”, where different terms are separated by whitespaces, resulting in an intersection query. Alternatively, one could execute a hashtag intersection query, like e.g. “ $\#o_i \#a_{ij}$ ”, which would nevertheless drastically reduce the result set, without necessarily increasing the precision.

A predefined number of tweets t_{1i}, \dots, t_{1n} is retrieved (default number is 100) that contain the relevant keywords. A secondary phase of preprocessing takes place on the retrieved set of tweets. The preprocessing phase involves removing characters or sequences of characters that cannot assist during the subsequent sentiment analysis phase, in order to reduce the noise in the data set. More specifically, for each retrieved tweet, the following items of text constitute representative examples to be removed:

1. Replies to other users' tweets, represented by strings starting with '@'.
2. URLs (i.e. strings starting with 'http://').
3. Hashtags, which are strings starting with '#' used for categorizing messages are not removed as a whole. Instead, only the '#' character is removed, since the rest of the string often forms a legible word that contributes to better understanding the tweet.

The remainder of each tweet is added into a collection of sentences.

4.2.3. Step #3: Sentiment analysis

After going through the preprocessing phase during the previous step, the retrieved tweets are submitted to OpenDover for sentiment analysis. OpenDover⁵ is a web service that tags the opinions

and sentiments detected in a textual corpus, based on the subject domain, as well as the intensity of the sentiment expression. A sentiment score s is assigned to each tweet, where $s \in [-10, 10]$, depending on the appreciation level of the submitted sentence. Fig. 2(B) displays two sample preprocessed tweets retrieved during this phase, as well as their corresponding sentiment score generated from OpenDover.

OpenDover was considered an appropriate choice for the proposed approach, since it is suitable for extracting sentiment from isolated sentences. An additional advantage is OpenDover's ontology-based architecture that offers the capability of detecting each time the domain of reference, adjusting the sentiment scores accordingly. It should be noted, however, that OpenDover could be substituted in the proposed architecture (Fig. 4) by any other equally efficient sentiment analysis tool and approach; the novelty in this work lies in the ontology-based analysis of tweets preceding the sentiment analysis phase. This analysis provides a more fine-grained sentiment evaluation regarding the distinct topics of a specific subject, discussed in each tweet.

On the other hand, deploying a third-party sentiment analysis service like OpenDover may be considered as a drawback, since the exact process of extracting the sentiment from a sentence cannot be verified – the source code and methodology behind OpenDover are not publicly available. Thus, an imminent goal for the future is to integrate a custom sentiment analysis methodology in our approach and compare the resulting sentiment scores.

4.3. Baseline scenario

The current subsection describes a baseline scenario that better illustrates the usability of the proposed approach. Suppose that a user wishes to perform a market research regarding smartphones and wants to determine other users' opinions on Twitter regarding the most popular smartphone models.

As the process described in this paper outlines, the first step is the creation of the domain ontology. In the scenario, we are going to adopt the FCA approach for creating the ontology (see Section 4.1.1). Thus, according to the ontology creation algorithm (Fig. 1), a default number of tweets is retrieved, based on the initial concept “#smartphone”. After processing the tweet set as the algorithm suggests, the resulting ontology takes the form displayed in Table 1. As mentioned earlier, the ontology can be visualized via a Hasse diagram, illustrated in Fig. 5; the diagram was created with ConExp (Yevtushenko, 2000), a software tool for analyzing formal contexts in FCA, drawing the corresponding concept lattices and exploring dependencies between attributes.

Alternatively, one could resort to ontology learning techniques for semi-automatically creating the ontology. Fig. 3 displays the resulting ontology visualization after using the OntoGen software tool.

In essence, Table 1 depicts the formal context $\mathcal{K}(\mathcal{O}, \mathcal{A}, \mathcal{I})$ of the scenario, where:

- \mathcal{O} is the set of retrieved objects, namely, the detected smartphone model series in the tweets, where $\mathcal{O} = \{\text{Apple iPhone, Samsung Galaxy, HTC One, Nokia Lumia}\}$,

⁵ OpenDover sentiment tagging web service: <http://opendover.nl/>.

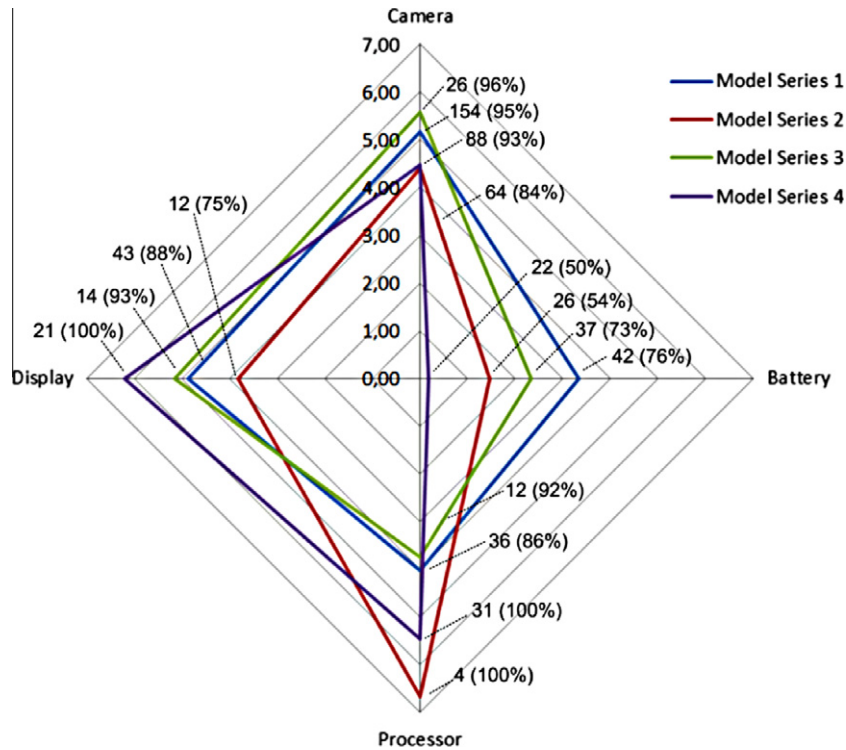


Fig. 6. Sentiment values corresponding to the smart phone attributes in the scenario.

- \mathcal{A} is the set of properties, where $\mathcal{O} = \{\text{Android, Camera, Battery, 4G, microUSB, Processor, Windows, Display, iOS, and,}$
- The incidence relation \mathcal{I} is represented by a series of crosses as shown in the table, where a '+' in cell (i, j) indicates that object o_i has attribute a_j .

In order for the model series comparison to make sense, we remove the attributes that are not common for every smartphone. This results in attribute set $\mathcal{A}' = \{a_j | \forall o_i \in \mathcal{O}, (o_i, a_j) \in \mathcal{I}\}$. More specifically, $\mathcal{A}' = \{\text{Camera, Battery, Processor, Display}\}$ and, obviously, $\mathcal{A}' \subset \mathcal{A}$. However, this modification in the attribute set is taking place only for fairness of comparison among the different smartphone models and does not affect in any sense the methodology proposed in this work.

The ontology is integrated to the proposed system illustrated in Fig. 4. The system automatically collects tweets and submits them to the OpenDover web service, according to the previously described sequence of phases. The tweets retrieved for the scenario spanned over a 1-week period and are available as a comma-separated file at: <http://goo.gl/UQvdx>.

The resulting sentiment values for each object-attribute pair are stored and the final results are depicted in Fig. 6. More specifically, the graph illustrates the average sentiment scores per attribute and model series. For each score, the total number of retrieved tweets is also displayed and, in parentheses, the corresponding positive-to-total tweet ratio. For reasons of objectiveness, the actual names of the smart phone model series in both tables have been substituted with generic tags.

4.4. Evaluation

The purpose of the current subsection is twofold: (a) to estimate the recall ratios⁶ for the two versions of our proposed architecture (see Fig. 4) as well as for the custom-built system (which has been introduced for evaluation purposes only) and (b) to evaluate whether the observed differences in the way the selections are performed by each method can be characterized as qualitatively analogous or not. The two versions of our proposed architecture are: (a) the full-fledged ontology-based semantically-enabled system (SEM), and, (b) the same system without the synonym/hyponym augmentation (see Section 4.1.3), but still with ontology support (ONT). The custom-built system is stripped of any ontology-based domain representation and, thus, cannot retrieve tweets referring to specific object-attribute pairs; it is limited to retrieving tweets regarding the superclass of the domain, which is associated to the "#smartphone" tag⁷ (CUS). The introduction of the CUS method is attributed to the fact that our approach adopts an utterly novel path and therefore it is not feasible to identify other methods that can be used as a comparison base. Additionally, there is no point in evaluating the returned sentiment results, since the OpenDover sentiment classifier used in this work does not constitute a contribution of ours.

The recall ratios for the three examined selection methods are estimated by using 10 randomly taken samples, each comprised of 100 observations. The estimation results, for all taken samples, reveal that SEM achieves steadily higher recall ratios from ONT and both present steadily higher recall ratios from CUS (See Panel A in Table 2). Solely based on the recall ratio as a comparison criterion a first round conclusion is that SEM performs better than ONT and both outperform CUS. However, we cannot draw any

⁶ Given a random sample of T observations, the recall ratio for a particular methodology is defined by the ratio of the total number of relevant selected tweets over the sample size.

⁷ The domain of reference still remains the world of smartphones, mentioned previously in the baseline scenario.

Table 2
Concordance statistics.

Sample number	Panel A			Panel B		
	Recall ratios			Concordance statistics		
	SEM	CUM	ONT	SEM–CUS	SEM–ONT	CUS–ONT
Sample1	0.94	0.72	0.26	0.30	0.78**	0.40
Sample2	0.86	0.72	0.40	0.36	0.86***	0.48
Sample3	0.89	0.60	0.26	0.29	0.71**	0.46
Sample4	0.90	0.67	0.27	0.37	0.77***	0.40
Sample5	0.79	0.60	0.32	0.41	0.81***	0.48
Sample6	0.90	0.69	0.29	0.33	0.77***	0.42
Sample7	0.85	0.65	0.26	0.29	0.80***	0.39
Sample8	0.80	0.53	0.26	0.40	0.73***	0.53
Sample9	0.88	0.60	0.36	0.44	0.72***	0.50
Sample10	0.89	0.61	0.28	0.33	0.72***	0.41

Notes: The number of observations for all the examined samples above is equal to 100. ***, ** and * indicate statistical significance at the 0.1 significance level, respectively.

solid conclusion if we do not first examine the significance of the degree of synchronization between all the investigated methods. Statistically significant synchronization between two methods implies no qualitative difference in the way the selections are performed by the methods under consideration while insignificant synchronization suggests the opposite. For this purpose we employ the non-parametric Concordance Statistic (CS hereafter), as suggested by [Harding and Pagan \(2002\)](#), which evaluates the degree of synchronization among the three alternative selection methods (m_j with $j = 1, 2, 3$).⁸

Assuming for every selection method (m_j) that the finally produced outcome can be typified by two mutually exclusive states (relevant or irrelevant tweet selection), then the CS signifies (for two compared methods) the fraction of observations that indicate simultaneously the same state. To facilitate the computation of the CS we introduce a binary random variable $S_{m_1,i}$ which receives for observation i the value of one if the m_1 method selects a relevant tweet and zero otherwise. For the remaining selection methods (m_2 and m_3) analogously are introduced the binary random variables $S_{m_2,i}$ and $S_{m_3,i}$. The CS between m_1 and m_2 (CS_{m_1,m_2}) mathematically is defined as:

$$CS_{m_1,m_2} = T^{-1} \left\{ \sum_{i=1}^T (S_{m_1,i} \cdot S_{m_2,i}) + \sum_{i=1}^T (1 - S_{m_1,i}) \cdot (1 - S_{m_2,i}) \right\} \quad (1)$$

where, T is the sample size and $S_{m_1,i}$ and $S_{m_2,i}$ are dichotomous random vectors defined as previously. A noticeable drawback of the above statistic is that it does not allow us to establish whether the identified degree of synchronization is statistically significant. To surmount the said drawback a *Generalized Method of Moments (GMM)* estimator is conscripted. In particular, the significance level of the CS is identified through the magnitude of the t -Statistic that corresponds to the coefficient a in the following moment condition:

$$E \left((S_{m_1,i} - \bar{S}_{m_1,i}) \cdot (S_{m_2,i} - \bar{S}_{m_2,i}) - a \right) = 0 \quad (2)$$

where, $\bar{S}_{m_1,i}$ and $\bar{S}_{m_2,i}$ denote the sample means for methods m_1 and m_2 , respectively.

Equation (2) is estimated via the GMM technique using the Marquardt optimization algorithm, the Bartlett kernel and a fixed bandwidth equal to five. The estimated CS for the three resulting pairs between the alternative methods, along with their significance level are analytically illustrated in Panel B of [Table 2](#). The results reveal that all the estimated degrees of synchronization for the first

(SEM and CUS) and the third pairs (CUS and ONT) are consistently non-significant and lie in a range between 0.29 to 0.44 and 0.39 to 0.53, respectively. On the contrary, the estimated degrees of synchronization for the second pair (SEM and ONT) are all statistically significant, mainly at the 0.01 significance level, with values that vary between 0.71 and 0.86. Overall, it can be argued that there is reasonable statistical evidence to support significant synchronization between the two versions of our architecture, while there is no evidence toward the direction of synchronization between the custom-built system and each one of the two versions of our architecture. Therefore, our proposed architecture, given the higher observed recall ratios, appears to perform evidently better than the custom-built system method.

4.5. Difficulties

During our work on the given field, a number of difficulties emerged, the most important of which are briefly outlined in this subsection, including suggestions for addressing each issue. A significant difficulty involves the unpleasantly high ratio of advertising tweets. The latter are not necessarily negative for Twitter users, but unfortunately distort the derived results to a degree. Advertisements can either be misunderstood by our system as positive user tweets or are rejected by OpenDover, since the vocabulary they contain conveys neither positive nor negative sentiments. A recommended solution to this problem might involve integrating into the architecture a subjectivity classifier, like the one proposed in ([Barbosa & Feng, 2010](#)). Such a classifier can distinguish subjective from objective tweets, isolating the advertisements and offering an additional filter to the set of tweets to be processed.

Another critical downside when working with Twitter and micro-blogging services in general involves the extensive use of jargon in the published posts. This is an issue that unavoidably occurs partly due to the imposed character limit to posts, but can also be attributed to the every-day communication nature of micro-blogs. The latter are not suitable for posting long pieces of text or fully justified opinions on a matter. A further difficulty was outlined in a previous section and involves the use of OpenDover as a third-party sentiment analysis service. The fact that the system is not open-source constitutes a drawback, since it is not possible to analyze the software, or even attempt to improve its efficiency. Therefore, one has to “blindly” rely on the derived sentiment scores. Nevertheless, an imminent solution to this issue involves integrating a custom-built sentiment analysis tool to the architecture of the proposed system, which constitutes one of our goals for future improvements.

5. Conclusions and future work

The paper argued that sentiment analysis constitutes a rapidly evolving research area, especially since the emergence of Web 2.0 and its related technologies (social networks, blogs, wikis etc.) The recent explosion in the usage of micro-blogging services, and particularly Twitter, has shifted attention to sentiment analysis of micro-blogging posts and tweets. There exist various machine learning-based approaches that perform sentiment analysis on tweets, with the drawback that they treat each tweet as one uniform statement and assigning a sentiment score to the post as a whole. This paper proposes the deployment of ontology-based techniques for determining the subjects discussed in tweets and breaking down each tweet into a set of aspects relevant to the subject. The result is the assignment of a sentiment score to each distinct aspect. A baseline scenario is also presented that deals with the domain of a popular product (smartphones) and results in comparatively evaluating the distinct features of each model series.

⁸ In that way we actually examine whether the observed differences in the estimated recall ratios are statistically significant.

Our goals for future improvements of the proposed approach initially involve the integration of a custom-built sentiment classifier that will substitute OpenDover in our architecture. A further aim is to integrate a fully automatic ontology-building functionality, potentially through a combination of ontology learning techniques. Nevertheless, keeping the manual and semi-automatic ontology creation approaches still remains useful, as they offer a more controlled means for building the domain vocabulary. Exploring various methods for visualizing the resulting sentiment is an additional direction for providing more thorough information to the user. Finally, provided that investors (or consumers) are prone to exogenous sentiment waves, an interesting research direction would be the development of time-series sentiment indexes for a range of investment (or consumer) goods. In case these developed indexes contain useful predictive power with respect each time to the future price movements of the investigated good, they may act as valuable tools in forming efficient strategies for all market participants.

Acknowledgements

The present scientific paper was executed in the context of the project entitled “International Hellenic University (Operation – Development)”, which is part of the Operational Programme “Education and Lifelong Learning” of the Ministry of Education, Lifelong Learning and Religious affairs and is funded by the European Commission (European Social Fund – ESF) and from national resources. Additionally, the authors would like to thank Dr. Ioannis Katakis for his valuable comments on the paper.

References

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). Sentiment analysis of twitter data. In *Proceedings of the ACL 2011 workshop on languages in social* (pp. 30–38). Media.
- Aue, A., & Gamon, M. (2005). Customizing sentiment classifiers to new domains: a case study. In *Proceedings of the international conference on recent advances in natural language processing (RANLP-05)* (pp. 207–218).
- Baldoni, M., Baroglio, C., Patti, V., & Rena, P. (2012). From tags to emotions: Ontology-driven sentiment analysis in the social semantic web. *Intelligenza Artificiale*, 6(1), 41–54.
- Barbosa, L., & Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd international conference on computational linguistics: Posters (COLING '10)* (pp. 36–44). Stroudsburg, PA, USA: Association for Computational Linguistics.
- Berners-Lee, T., Hender, J., & Lassila, O. (2001). The semantic web. *Scientific American*, 284(5), 34–43.
- Breslin, J. G., Harth, A., Bojars, U., & Decker, S. (2005). Towards semantically-interlinked online communities. In *2nd European semantic web conference (ESWC 2005)* (pp. 500–514). May 29–June 1.
- Brickley, D., & Miller, L. (2010). FOAF Vocabulary Specification 0.98. Namespace Document 9 August 2010 – Marco Polo Edition, FOAF Project. Available from: <<http://xmlns.com/foaf/0.1/>>, last access: November 2012.
- Cimiano, P., & Völker, J. (2005). Text2Onto – a framework for ontology learning and data-driven change discovery. In *10th International conference on applications of natural language to information systems (NLDB)* (pp. 227–238). LNCS.
- Claster, W. B., Cooper, M., & Sallis, P. (2010). Thailand – tourism and conflict: Modeling sentiment from twitter tweets using Naïve Bayes and unsupervised artificial neural nets. In *Proceedings of the CIMSIM'10* (pp. 89–94). Washington, DC, USA: IEEE Computer Society.
- Davey, B. A., & Priestley, H. A. (2002). *Introduction to lattices and order*. Cambridge University Press. ISBN 978-0-521-78451-1.
- De Kok, D., & Brouwer, H. (2012). Natural language processing for the working programmer. E-book available under the creative commons attribution 3.0 License (CC-BY). 2011, Available from: <<http://www.nlpwp.org/nlpwp.pdf>>, last access: November 2012.
- Dergiades, T. (2012). Do investors' sentiment dynamics affect stock returns? Evidence from the US economy. *Economics Letters*, 116(3), 404–407.
- Fortuna, B., Grobelnik, M., & Mladenic, D. (2007). OntoGen: Semi-automatic ontology editor. In M. J. Smith & G. Salvendy (Eds.), *Proceedings of the 2007 conference on human interface* (pp. 309–318). Berlin, Heidelberg: Springer-Verlag.
- Francisco, V., Germans, P., & Peinado, F. (2007). Ontological reasoning to configure emotional voice synthesis. In *Proceedings of the web reasoning and rule systems (RR 2007)* (pp. 88–102). Springer, LNCS 4524.
- Fu, G., & Cohn, A. G. (2008). Utility ontology development with formal concept analysis. In C. Eschenbach & M. Grüninger (Eds.), *Proceedings of the 2008 conference on formal ontology in information systems (FOIS 2008)* (pp. 297–310). Amsterdam, The Netherlands: IOS Press.
- Ganter, B. B., & Wille, R. (1999). *Formal concept analysis, mathematical foundation*. Berlin: Springer Verlag.
- Gómez-Pérez, A., & Corcho, O. (2002). Ontology languages for the semantic web. *IEEE Intelligent Systems*, 17(1), 54–60.
- Harding, D., & Pagan, A. (2002). Dissecting the Cycle: A Methodological Investigation. *Journal of Monetary Economics*, 49(2), 365–381.
- Hazman, M., El-Beltagy, S. R., & Rafea, A. (2009). Ontology learning from domain specific web documents. *International Journal of Metadata, Semantics and Ontologies*, 4(1/2), 24–33.
- He, Y., & Alani, H. (2011). Semantic smoothing for twitter sentiment analysis. In *Proceedings of the 10th international semantic web conference (ISWC)*. demo/poster session.
- Iwanaga, I., Nguyen, T. M., Kawamura, T., Nakagawa, H., Tahara, Y., & Ohsuga, A. (2011). Building an earthquake evacuation ontology from twitter. In *Proceedings of the IEEE international conference on granular computing (GrC)* (pp. 306–311). 8–10 November.
- Kaji, N., & Kitsuregawa, M. (2007). Building lexicon for sentiment analysis from massive collection of HTML documents. In *Proceedings of the joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL)* (pp. 1075–1083).
- Kang, H., Yoo, S., & Han, D. (2012). Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews. *Expert Systems with Applications*, 39(5), 6000–6010.
- Kaplan, A. M., & Haenlein, M. (2011). The early bird catches the news: Nine things you should know about micro-blogging. *Business Horizons*, 54(2), 105–113.
- Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the OMG!. In *Proceedings of the ICWSM*.
- Liu, Y., Huang, X., An, A., & Yu, X. (2007). ARSA: A sentiment-aware model for predicting sales performance using blogs. In *Proceedings of the 30th annual international ACM SIGIR conference on research and development in information retrieval* (pp. 607–614). Amsterdam, The Netherlands. July 23–27.
- Melville, P., Gryc, W., & Lawrence, R. D. (2009). Sentiment analysis of blogs by combining lexical knowledge with text classification. In *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD '09)* (pp. 1275–1284). New York, NY, USA: ACM.
- Miller, G. A. (1995). WordNet: A lexical database for English. *Communications of the ACM*, 38(11), 39–41.
- Moreo, A., Romero, M., Castro, J. L., & Zurita, J. M. (2012). Lexicon-based comments-oriented news sentiment analyzer system. *Expert Systems with Applications*, 39(10), 9166–9180.
- Neviarouskaya, A., Prendinger, H., & Ishizuka, M. (2009). SentiFul: Generating a reliable lexicon for sentiment analysis. In *Proceedings of the affective computing and intelligent interaction and workshops (ACII 2009)*, 3rd international conference on affective computing and intelligent interaction and workshops (pp. 10–12). IEEE, September 1–6.
- Ning, L., Guanyu, L., & Li, S. (2010). Using formal concept analysis for maritime ontology building. In *Proceedings of the 2010 international forum on information technology and applications (IFTA '10)* (pp. 159–162). New York, USA: Springer. Vol. 2.
- Obitko, M., Snasel, V., Smid, J., & Snasel, V. (2004). Ontology design with formal concept analysis. In V. Snasel & R. Belohlavek (Eds.), *Concept lattices and their applications* (pp. 111–119). Ostrava: Czech Republic.
- Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the 7th international conference language resources and evaluation (LREC '10)*. European Language Resources Association ELRA.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.
- Park, S., Ko, M., Kim, J., Liu, Y., & Song, J. (2011). The politics of comments: Predicting political orientation of news stories with commenters sentiment patterns. In *Proceedings of the ACM 2011 conference on computer supported cooperative work (CSCW '11)* (pp. 113–122). New York, NY, USA: ACM.
- Passant, A., Bojars, U., Breslin, J. G., Hastrup, T., Stankovic, M., & Laublet, P. (2010). An overview of SMOB 2: Open, semantic and distributed microblogging. In *4th international conference on weblogs and social media* (pp. 303–306). ICWSM.
- Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, 3(2), 143–157.
- Rajagopal, H. (2005). JENA: A Java API for ontology management. Colorado software summit: October 23–28, 2005 © Copyright 2005, IBM Corporation, Available from: <<http://goo.gl/QMaQC>>, last access: November 2012.
- Read, J. (2005). Using emoticons to reduce dependency in machine learning techniques for sentiment classification. In *Proceedings of the ACL-05, 43rd meeting of the association for computational linguistics*. Association for Computational Linguistics.
- Saif, H., He, Y., & Alani, H. (2012). Alleviating data sparsity for twitter sentiment analysis. In *2nd Workshop on making sense of microposts (#MSM2012): Big things come in small packages at World Wide Web (WWW) 2012* (pp. 2–9). Lyon, France.
- Skiba, D. J. (2006). Web 2.0: Next great thing or just marketing hype? *Nursing Education Perspectives*, 27(4), 212–214.
- Stankovic, M., Passant, A., & Laublet, P. (2009). Status messages for the right audience with an ontology-based approach. In *1st International workshop on collaborative social networks – collaboratesn 2009 – at the 5th international conference on collaborative* (pp. 1–7). Computing.

- Studer, R., Benjamins, R., & Fensel, D. (1998). Knowledge engineering: Principles and methods. *IEEE Trans on Data and Knowledge Eng.*, 25(1–2), 161–197.
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307.
- Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. In *Proceedings 4th international AAAI conference on weblogs and social* (pp. 178–185). Media.
- Wanner, F., Rohrdantz, C., Mansmann, F., Oelke, D., & Keim, D. A. (2009). Visual sentiment analysis of RSS news feeds featuring the US presidential election in 2008. In *Workshop on visual interfaces to the social and the semantic web (VISSW)* (pp. 1–8).
- Yevtushenko, S. A. (2000). System of data analysis concept explorer. In *Proceedings of the 7th national conference on artificial intelligence KII-2000* (pp. 127–134). Russia. (in Russian).
- Zhang, R., & Xu, H. (2011). Building the ontology system in semantic web based on formal concept analysis and rough set. *Journal of Convergence Information Technology*, 6(7), 56–62.