SEMANTIC SEARCH IN SOCIAL MEDIA ANALYSIS

Liviu-Adrian COTFAS
Bucharest University of Economic Studies, Romania
liviu.cotfas@ase.ro
Ioan ROXIN
University of Franche-Comté, Montbéliard, France
ioan-roxin@univ-fcomte.fr
Camelia DELCEA
Bucharest University of Economic Studies, Romania
camelia.delcea@csie.ase.ro

Abstract. *Many of the messages posted nowadays on social media networks have been shown to contain rich sentiment indicators regarding almost any topic. Making sense of this huge amount of data could provide invaluable benefits to companies trying to better understand how the public perceives their products and services. Thus, the present paper proposes an approach that combines Semantic Web technologies and Natural Language Processing in order to both overcome the limitations of traditional search approaches and to provide a near-real time image of the public's perception. Using Named Entity Recognition, the social media messages are annotated with the corresponding concepts in the proposed ontology, while a bag-of-words Sentiment Analysis algorithm uncovers the sentiment polarity of each tweet. The results of the analysis are stored as semantically structured data, thus making it possible to fully exploit the possibilities offered by semantic web technologies, such as inference and accessing the vast amount of knowledge in Linked Open Data, for further analysis.*

Keywords: social media, semantic web, ontology, semantic search
JEL classification: L86, H12, H84
DOI: 10.12948/ie2019.01.07

1. Introduction

The last few years have witnessed an amazingly fast-paced growth in the usage of social media networks. Platforms such as Twitter, Facebook and Instagram allow users to publish messages on a wide variety of topics [1], [2]. Many of these messages have been shown to contain sentiment indicators regarding companies and their products and services [3], [4], turning the social media networks into a reach data source for understanding the public's opinion. Among the most popular social media networks, Twitter possesses a series of characteristics, such as the possibility of writing concise messages, which makes it especially appealing to the users who aim to express their feelings [5]. The messages posted on this social media network are called tweets and have a maximum length of 280 characters.

Accurately understanding the social media users' opinions could provide invaluable information to companies willing to better comprehend how their products and services are perceived. Compared to traditional marketing studies, which usually take time and involve high costs, social media analysis offers the promise of obtaining almost real-time opinions from a huge number of actual or potential customers.

Discovering the products and service mentioned in social media messages, as well as the expressed perception can be achieved using techniques such as Named Entity Recognition – NER and Sentiment Analysis – SA, both of them growing areas of the Natural Language

Processing – NLP field, that involve a multidisciplinary approach which combines elements from various fields such as artificial intelligence, psychology and linguistics. Given the fact that many social media messages are written using a casual language, the NER approach must take into account the fact that that users frequently employ various hashtags, abbreviations and spellings when referring to the same product or service. Sentiment analysis, a technique used in order to determine whether a text express a positive, negative or neutral polarity, has already been successfully applied to a variety of tasks, including analysing customer reviews [6], stock market prediction [7] and predicting the results of political elections [8].

In this paper, an approach that combines Semantic Web technologies and Natural Language Processing is proposed, in order to both surpass the limitations of traditional search approaches and to provide a near-real time image of the public's perception regarding a company's products and service. By storing the collected data using semantic web technologies, it becomes possible to discover not only the messages related to the currently analysed topic, but also to other more specific topics. Thus, the public's opinion can easily be analysed at different levels of granularity: concerning a specific product, concerning the products in a certain category, concerning all the products of the company.

The paper is organized as follows. The next section focuses on the semantic web ontologies that will be used to discover all the social media messages relevant for a specific topic. The third section includes the steps taken in order to identify the concepts mentioned in the social media messages. The fourth section presents the proposed semantic search approach and how it can be used to better uncover the public's opining regarding a certain topic using sentiment analysis. The last section summarizes the paper and introduces possible future research directions.

2. Social Media Analysis Ontologies

The Semantic Web, also known as the web of data, can be thought of as an evolution of the nowadays web of documents. By focusing on standard formats and exchange protocols, data can more easily be shared between organizations using machine-readable formats [9]. It is governed by rules and guidelines established by the World Wide Web Consortium - W3C and uses the Resource Description Framework – RDF [10] and the Ontology Web Language – OWL [11] as a foundation for publishing and linking data.

The primary means of representing data within the semantic web are ontologies, defined according to Borst [12] as a "formal specification of a shared conceptualization". Ontologies organize concepts in hierarchies, using a shared vocabulary that provides a common understanding, in order to denote the concepts, their properties and the relations between them. The information stored in ontologies can easily be retrieved using a specialized query language, known as SPARQL and new relationships inside the data can be discovered through inference using semantic reasoning engines.

Ontologies have already been successfully used in many social media analysis tasks, including detecting trending news and topics [13], modelling of extreme financial events [14], understanding people behaviour in an earthquake evacuation scenario [15], extracting user preferences regarding the characteristics of a product [4] and analysing the emotions expressed in social media messages [16].

The concepts required in order to semantically search information in previously collected social media messages can be grouped in the following three categories:

- concepts that describe the social media specific knowledge;
- concepts that represent the analysed entities, such as products or service;
- concepts that provide a connection between the social media messages, the analysed entities, as well as with any other additional data obtained using NLP techniques, such as

sentiment or emotion analysis.
Since large monolithic ontologies are considered more difficult to understand, handle and extend, a modular approach has been used [17].

2.1. Social Media Ontology

For representing the social media concepts and their properties, we have chosen to use the ontology that we have proposed in [16], which extends well recognized ontologies such as SIOC and FOAF with the concepts specific to Twitter and follows the recommended ontology modelling best practices. The *tw* prefix is used in the following to denote classes or properties belonging to this ontology.

2.2. Social Media Analysis Ontology

The Social Media Analysis ontology provide a connection between the social media message, the identified topic and the expressed sentiment, as shown in Figure 3. The *sma* prefix is used in the rest of the paper to denote classes or properties belonging to this ontology.

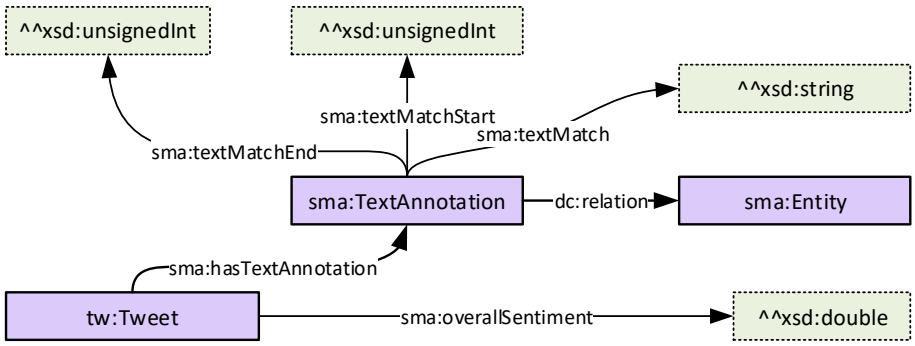

Figure 3. Main concepts in Social Media Analysis Ontology

Starting from this ontology, the classes corresponding to the actual types of products and services that will be analysed can be declared as shown in Figure 4.

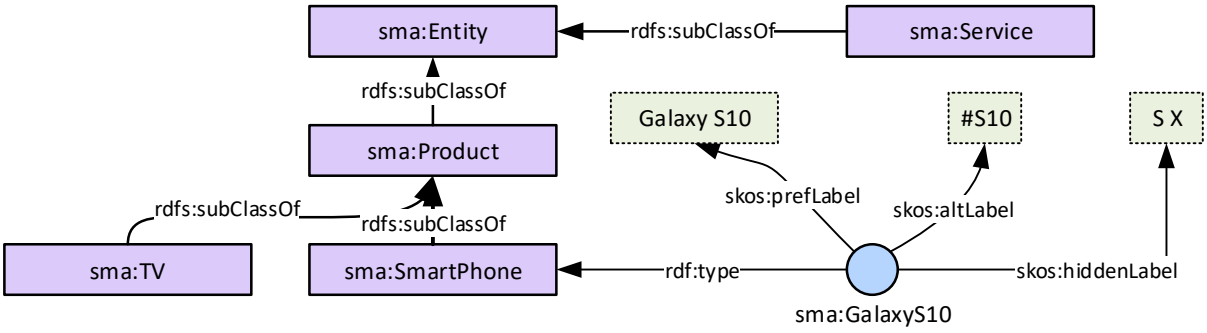

Figure 4. Declaring the analyzed classes and instances

3. Social Media Analysis

Unlike carefully authored texts, tweets pose multiple challenges, mainly due to their short, noisy and context-dependent nature. The steps taken in order to analyse the social media messages are described in the subsections below.

3.1. Tweet Retrieval

The tweets can be collected using one of the available Twitter APIs, such as the Twitter Public Stream API, which offers the possibility to filter the huge number of tweets published every second using specific keywords or hashtags, such as "#samsung".

3.2. Pre-processing

A pre-processing step is needed before analysing a tweet, given the fact that many social media users write their messages in a casual language. During this step, the text of the tweet is split into individual tokens and a normalization procedure is applied, during which all-caps words are converted to lowercase and duplicated letters are removed [18].

3.3. Sentiment Analysis

While many sentiment analysis algorithms have been proposed in the scientific literature, we have chosen to use the bag of words model described in [16], given its low complexity and adequate results.

3.1. Named Entity Recognition

Named Entity Recognition is a Natural Language Processing technique that tries to identify named entity mentions in unstructured text [19]. The tokens resulting from the pre-processing step are first stemmed, and afterwards, compared with the stemmed versions of the tokens associated using *skos:prefLabel*, *skos:altLabel* and *skos:hiddenLabel* with the instances of *sma:Entity* class. The results of the analysis process are stored using semantic web technologies as shown in Figure 5.

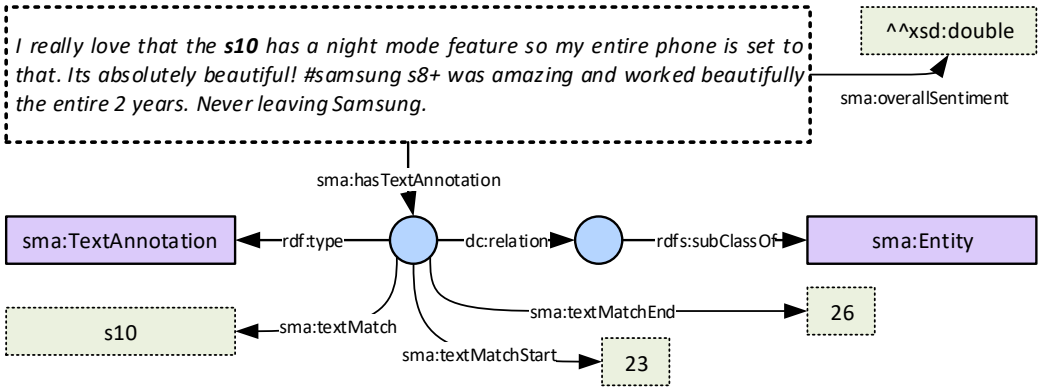

Figure 5. Annotated Tweet

4. Semantic Search

The collected data can easily be explored using SPARQL (SPARQL Protocol and RDF Query Language as shown in Figure 6 that includes a query that will return the tweets relevant for a certain smart phone, together with the associated sentiment score.

```
SELECT ?tweetContent ?sentiment
WHERE
{
       ?tweet sioc:content ?tweetContent ;
              sma:overAllSentiment ?sentiment ;
              sma:hasTextAnnotation ?textAnnotation .
       ?textAnnotation dc:relation ?relation .
       FILTER(?relation = sma:GalaxyS10)
}
```

Figure 6. SPARQL query for the tweets associated with a certain smart phone

Given the fact that the *sma:GalaxyS10* is an instance of the *sma:SmartPhone* class, Figure 7 includes the query that will return the tweets that are relevant for any instance of the *sma:SmartPhone* class, including the ones for *sma:GalaxyS10*.

```
SELECT ?tweetContent ?sentiment
WHERE
{
       ?tweet sioc:content ?tweetContent .
       ?tweet sma:overAllSentiment ?sentiment .
       ?tweet sma:hasTextAnnotation ?textAnnotation .
       ?textAnnotation dc:relation ?relation .
       ?relation rdf:type sma:SmartPhone .
}
```

Figure 7. SPARQL query for the tweets relevant for any smart phone

The query above, can be adjusted to return the tweets for all the products, by replacing *sma:SmartPhone* with the class *sma:Product*.

5. Conclusions

The present paper proposes a novel approach for uncovering from the vast amount of social media messages only the ones relevant for a specific topic. By organizing the analysed topics as a hierarchy with the help of semantic web technologies, it becomes possible to leave behind the limitations of traditional search approaches, in order to discover not only messages related to the currently analysed topic, but also to other highly related topics. Once all the social media messages on a particular topic have been discovered, it is possible to use sentiment analysis algorithms in order to analyse the public's opinion at different level of granularity. Moreover, the proposed approach can easily be incorporated as a part of a larger semantic web-based social media analysis platform.

Among the possible further research directions, we consider investigating more advanced deep learning approaches for both Named Entity Recognition and Sentiment Analysis. Also, while knowing the sentiments expressed by social media users is definitely important, analysing the categories of emotions contained in social media messages using emotion analysis can provide much deeper insights, by putting the focus on the actual feelings, such as happiness, sadness, surprise or anger.

Acknowledgment

This work was supported by a grant by UEFISCDI ("Unitatea Executivă pentru Finanţarea Învăţământului Superior, a Cercetării, Dezvoltării şi Inovării"), project FutureWeb ("Modelarea empirică şi dezvoltarea experimentală a instrumentelor asociate tehnologiilor emergente din domeniul reţelelor sociale online"), project number: PN-III-P1-1.2-PCCDI-2017-0800, 86PCCDI/2018.

References

[1] A. N. Smith, E. Fischer, and C. Yongjian, "How Does Brand-related User-generated Content Differ across YouTube, Facebook, and Twitter?," Journal of Interactive Marketing, vol. 26, no. 2, pp. 102–113, May 2012.

[2] E. Uzunoğlu, S. Türkel, and B. Yaman Akyar, "Engaging consumers through corporate social responsibility messages on social media: An experimental study," Public Relations Review, vol. 43, no. 5, pp. 989–997, Dec. 2017.

[3] A. Pak and P. Paroubek, “Twitter as a Corpus for Sentiment Analysis and Opinion Mining,” in Proceedings of the Seventh International Conference on Language Resources and Evalua-tion, Valletta, 2010, pp. 1320–1326.
[4] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, “Ontology-based sentiment analysis of twitter posts,” Expert Systems with Applications, vol. 40, no. 10, pp. 4065–4074, Aug. 2013.
[5] J. H. Parmelee and S. L. Bichard, Politics and the Twitter revolution: how tweets influence the relationship between political leaders and the public. Lanham, Md: Lexington Books, 2012.
[6] B. Liu and L. Zhang, “A Survey of Opinion Mining and Sentiment Analysis,” in Mining Text Data, C. C. Aggarwal and C. Zhai, Eds. Boston, MA: Springer US, 2012, pp. 415–463.
[7] A. Khadjeh Nassirtoussi, S. Aghabozorgi, T. Ying Wah, and D. C. L. Ngo, “Text mining for market prediction: A systematic review,” Expert Systems with Applications, vol. 41, no. 16, pp. 7653–7670, Nov. 2014.
[8] S. Rill, D. Reinel, J. Scheidt, and R. V. Zicari, “PoliTwi: Early detection of emerging political topics on twitter and the impact on concept-level sentiment analysis,” Knowledge-Based Systems, vol. 69, pp. 24–33, Oct. 2014.
[9] T. BERNERS-LEE, J. HENDLER, and O. LASSILA, “THE SEMANTIC WEB,” Scientific American, vol. 284, no. 5, pp. 34–43, 2001.
[10] “RDF - Semantic Web Standards.” [Online]. Available: https://www.w3.org/RDF/. [Accessed: 18-Nov-2018].
[11] “OWL 2 Web Ontology Language Document Overview (Second Edition).” [Online]. Available: https://www.w3.org/TR/owl2-overview/. [Accessed: 18-Nov-2018].
[12] W. N. Borst, Construction of engineering ontologies for knowledge sharing and reuse. Universiteit Twente, 1997.
[13] A. Ejaz, S. K. Fatima, Q. N. Rajput, and S. A. Khoja, “Analyzing News from Electronic Media and Topics Discussed on Social Media Using Ontology,” in 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS), 2018, pp. 349–354.
[14] H. Qu, M. Sardelich Nascimento, N. N. Qomariyah, and D. L. Kazakov, “Integrating Time Series with Social Media Data in an Ontology for the Modelling of Extreme Financial Events,” LREC 2016 Proceedings, 23-May-2016. [Online]. Available: http://eprints.whiterose.ac.uk/128500/. [Accessed: 11-Jan-2019].
[15] I. S. M. Iwanaga, T. Nguyen, T. Kawamura, H. Nakagawa, Y. Tahara, and A. Ohsuga, “Building an earthquake evacuation ontology from twitter,” in 2011 IEEE International Conference on Granular Computing, 2011, pp. 306–311.
[16] L.-A. Cotfas, C. Delcea, A. Segault, and I. Roxin, “Semantic Web-Based Social Media Analysis,” in Transactions on Computational Collective Intelligence XXII, vol. 9655, N. T. Nguyen and R. Kowalczyk, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 147–166.
[17] S. Ben Abbès, A. Scheuermann, T. Meilender, and M. D’Aquin, “Characterizing Modular Ontologies,” in 7th International Conference on Formal Ontologies in Information Systems - FOIS 2012, Graz, Austria, 2012, pp. 13–25.
[18] Y. Bao, C. Quan, L. Wang, and F. Ren, “The Role of Pre-processing in Twitter Sentiment Analysis,” in Intelligent Computing Methodologies, Springer, 2014, pp. 615–624.
[19] L. Derczynski et al., “Analysis of named entity recognition and linking for tweets,” Information Processing & Management, vol. 51, no. 2, pp. 32–49, Mar. 2015.