Developing Smart Cities Services through Semantic Analysis of Social Streams

Cataldo Musto, Giovanni Semeraro, Marco de Gemmis, Pasquale Lops
Department of Computer Science - University of Bari Aldo Moro, Italy
{name.surname}@uniba.it

ABSTRACT

This paper presents a domain-agnostic framework for intelligent processing of textual streams coming from social networks. The framework implements a pipeline of techniques for semantic representation, sentiment analysis, automatic content classification, and provides an analytics console to get some findings from the extracted data. The effectiveness of the platform has already been proved by deploying it in two smart cities-related scenarios: in the first it was exploited to monitor the recovering state of the *social capital* of L'Aquila's city after the dreadful earthquake of April 2009[1], while in the latter a semantic analysis of the content posted on social networks was performed to build a map of the most at-risk areas of the Italian territory.

In both scenarios, the outcomes resulting from the analysis confirmed the insight that the adoption of methodologies for intelligent and semantic analysis of textual content can provide interesting findings useful to improve the understanding of very complex phenomena.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: General

Keywords

Smart Cities, Big Data, Natural Language Processing, Semantics, Sentiment Analysis, Entity Linking

1. BACKGROUND AND CONTRIBUTIONS

The huge availability of data coming from social networks leads research scientists to new opportunities and challenges. As an example, mining micro-blogs content is drawing more and more attention since many latent information about people sentiment, preferences, thinking and opinions can be automatically extracted from textual streams, and this paves the way to the development of new innovative and intelligent services relying on content analytics of *human-generated* data.

However, the development of a framework for semantic analysis of content coming from social networks is not trivial, since several methodologies need to be combined to obtain a fine-grained semantic content representation. Indeed, a simple Natural Language Processing (NLP) pipeline is typically not enough to extract valuable findings from data, thus it is necessary to couple it with more complex content processing methodologies such as opinion mining, content classification, semantic processing, topic modeling and network analysis.

To this aim, in this paper we present a domain-agnostic framework for intelligent processing of social textual streams. Our framework can perform massive extraction and mining of textual content and implements state-of-the-art algorithms for most of the above mentioned methodologies. Moreover, it makes the output available through an interactive analytics console based on widespread and effective data visualization formalisms (maps, charts and tag clouds).

One of the distinguishing aspect of this work lies in the originality of the scenarios in which the framework has already been deployed: *L'Aquila Social Urban Network* and *the Italian Hate Map*. In both cases, our platform has been exploited to develop novel smart cities-related services based on the analysis of social streams. In the first case, the aggregation and the semantic analysis of micro-blogs posts has been performed to build a map of the most at-risk areas in Italy, while in the latter semantic processing has been coupled with sentiment analysis and text classification to obtain a snapshot of people feelings and opinions about the state of the city of L'Aquila after the earthquake of 2009. The rest of the article is organized as follows: Section 2 describes the architecture of the framework, while the use cases and some example of the outcomes provided by both scenarios are depicted in Section 3. Next, Section 4 presents Related Work in the area and Section 5 defines future directions and concludes our work.

2. DESCRIPTION OF THE FRAMEWORK

Our framework is based on the concept of *analysis*. Each analysis is run by defining a *source*, a set of *extraction heuristics* and some *processing steps*. In a typical pipeline, a user interacts with the framework by defining the social networks which act as the source of the framework and some heuristics to drive the extraction process. Next, the user defines what kind of processing she wants to perform on the content and what kind of data visualization she needs. The goal of the platform is to extract, analyze and aggregate data in order

[1] http://en.wikipedia.org/wiki/2009_L'Aquila_earthquake

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05. http://dx.doi.org/10.1145/2740908.2742132.

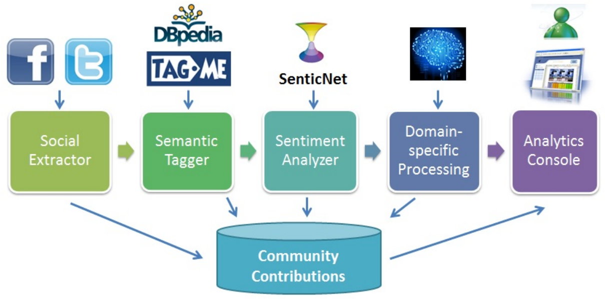


Figure 1: The architecture of the framework

to produce some analytics which is valuable for the users. It is important to underline that the framework is totally domain-independent, thus it can aggregate and extract any kind of content the user wants to analyze.

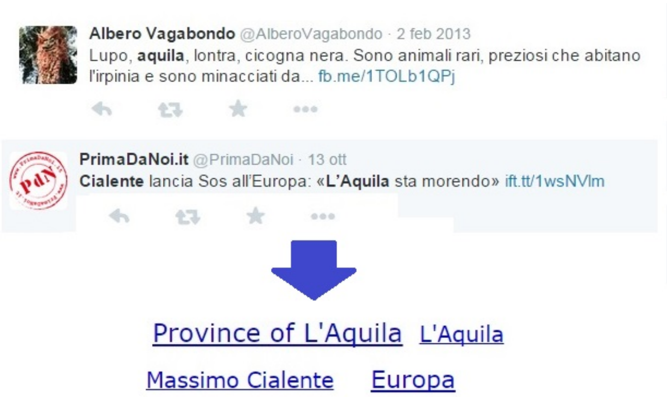
The general architecture of the framework is showed in Figure 1. Hereafter, a brief description of each component is provided.

Social Extractor: is the essential component of the whole pipeline. Given some *extraction heuristics*, the component allows the framework to connect to a social network and to extract some content that matches the heuristics. This framework is able to extract content from both Facebook and Twitter by exploiting the official APIs. As extraction heuristics, six different alternatives have been implemented:

- **Content:** extracts all Tweets which match a specific term;
- **User:** extracts all Tweets posted by a specific user, given its user name;
- **Geo:** extracts all geolocalized Tweets, given latitude, longitude and radius;
- **Content+Geo:** extracts all geolocalized Tweets which match a specific term;
- **Page:** extracts all Facebook posts coming from a specific page (the main post as well as the replies);
- **Group:** extracts all Facebook posts coming from a specific group (the main posts as well as the replies);

Even if Facebook APIs let extract more complex content (e.g. all the *likes* of a specific user, all the discussions in a specific timeline and so on), we only took into account the content labeled as *public*, since the goal of the platform is a large-scale massive extraction of content, with no need of an explicit users' authorization. Regardless the specific scenario, all extracted information is then anonymyzed and locally stored in a database of *contributions*. Data are continuously extracted and stored, for the whole time interval an analysis is running. MongoDB was chosen as storage solution, since its document-based storage model perfectly fits the requirements of the framework. Hereafter, we will refer to all the contributions as *social content*, regardless the source they come from.

Semantic Tagger: this component further processes the content gathered by the *Social Extractor* before aggregating, filtering and presenting it in the Analytics Console. This requirement is due to the fact that the extraction process is carried out through a simple keyword-based matching. As a consequence, a lot of irrelevant content is extracted, especially when *polysemous* terms are used in the extraction heuristics. The next example will clarify this aspect.



Albero Vagabondo @AlberoVagabondo · 2 feb 2013
Lupo, **aquila**, lontra, cicogna nera. Sono animali rari, preziosi che abitano l'irpinia e sono minacciati da... fb.me/1TOLb1QPj

PrimaDaNoi.it @PrimaDaNoi · 13 ott
Cialente lancia Sos all'Europa: «**L'Aquila** sta morendo» ift.tt/1wsNVlm

Figure 2: Example of ambiguous Tweets and its resulting semantic representation

Let us suppose the term ”L'Aquila” is used to extract all Tweets where people talk about the city hit by the earthquake in 2009. Unfortunately, as shown in Figure 2, L'Aquila is a polysemous term, since in Italian it is the translation of the term *eagle*, as well. As a consequence, the second tweet actually discusses about the problems of the city after the earthquake, while the first is about the risk of extinction of the eagle[2].

This issue is typically tackled relying on NLP techniques which improve the representation of the content gathered by the *Social Extractor* by filtering out some noise and by introducing some semantics. Moreover, given that each analysis is supposed to extract and process the huge amount of content (just think about how many Tweets about a certain topic are posted every day), it is useful to further improve the organization of the information by aggregating the content and extracting some high-level concepts which can provide the user with a more general and abstract overview of the data. To this aim, in the *Semantic Tagger* a pipeline of *entity linking* algorithms has been implemented to produce a transparent, richer and fine-grained semantic content representation. Generally speaking, the goal of entity linking techniques is to identify the *entities* which are mentioned in a piece of text. This is done by relying on statistical approaches which scan through the content by exploiting large corpora of entities as Wikipedia. The output of the process is a set of entities, each of which is mapped to a univocal reference (usually, the Wikipedia page it refers to).

In our approach, each *social content* has been processed through a pipeline of state-of-the-art entity linking algorithms. We chose DBpedia Spotlight[3], Wikipedia Miner[4]

[2] For the sake of simplicity, both Tweets are reported in Italian. The translation of the first one is 'Cialente sends out an SOS to Europe, L'Aquila is going to die' while the translation of the latter is 'Wolf, eagle, otter, black stork they are rare and precious animals that live in Irpinia and are threatened by...'

[3] http://dbpedia-spotlight.github.io/demo/

[4] http://wikipedia-miner.cms.waikato.ac.nz/

and Tag.me[5]. Figure 2 provides the output coming from the processing of the first Tweet. As shown in the Figure, thanks to entity linking algorithms our framework is able to understand that a certain Tweet is about L'Aquila and Massimo Cialente (Mayor of the city). This representation allows to filter out most of the noise, since the content which is not relevant for the goal of a specific analysis is automatically removed and is not taken into account.

It immediately emerges that such processing incorporates stop words removal, bigrams recognition and entities identification (and disambiguation). Furthermore, since each entity is mapped to a Wikipedia page, we can further enrich content representation by introducing the most relevant ancestor categories of that page. By considering the previous example, given the concept *Massimo Cialente*, the representation is enriched by adding as extra features concepts such as *Democrats Politics*, *L'Aquila Mayors* and so on, thus extending the representation with other relevant features that may be of interest to understand what the content is about. It is important to further emphasize that entity linking algorithms can also enrich the representation by introducing features which not explicitly occur in the text, and this is tremendously important to obtain a more transparent and richer representation of social content.

To sum up, the goal of the Semantic Tagger is to process all the previously extracted content by exploiting entity linking algorithms, in order to obtain a richer and fine-grained semantic representation which is very valuable and useful for final users, in virtue of its transparency and readability.

Sentiment Analyzer: The goal of this component is to further enrich the comprehension of the content by analyzing the *opinion* conveyed by each social content. Going back to the previous example, it can be important to understand the opinion of people of L'Aquila about their mayor.

To this aim, we implemented a lexicon-based algorithm for sentiment analysis. Lexicon-based approaches [9] infer the sentiment conveyed by a piece of text by relying on (external) lexical resources which map each term to a categorical *(positive, negative, neutral)* or numerical sentiment score. As an example, terms as *happy* and *love* have a positive sentiment score, while terms as *sad*, *die* and so on have a negative score. Our algorithm is based on the insight that the polarity of a piece of text can be obtained on the ground of the polarity of the micro-phrases which compose it. A new *micro-phrase* is built whenever a *splitting cue* is found in the text. Punctuations, adverbs and conjunctions were used as *splitting cues*. If we take into account the first Tweet reported in Figure 2 *"Cialente sends out an SOS to Europe, L'Aquila is going to die"* it is split in two micro-phrases, delimited by comma.

Next, the polarity of each *micro-phrase* depends on the sentiment score of each term in the micro-phrase, which is obtained from an external vocabulary. Clearly, our algorithm is able to deal also with negation since the polarity of a sentence is inverted when a negation is found in the text. As lexical resource we exploited SenticNet [4], a lexical resource for *concept-level* sentiment analysis. SenticNet is able to associate polarity and affective information also to complex concepts, such as *accomplishing goal*, *celebrate special occasion* and so on. At present, SenticNet provides sentiment scores (in a range between -1 and 1) for 14,000 common sense concepts. By referring to the previous example again, the Tweet in Figure 2 contains only a term with a clear polarity (*die*), which influences negatively the overall score (its SenticNet score[6] is -0.235).

To sum up, thanks to this technique, it is possible associate an opinion (if any) to each social content which was previously extracted, and this can be tremendously important for several scenarios.

Domain-Specific Processing: beyond Semantic Tagging and Sentiment Anaylsis, each analysis carried out by the framework may require some further domain-specific processing steps. They could range from the application of Machine Learning-based techniques (as text classification or mining social networks) to the simple applications of heuristics to filter out or enrich the previously extracted data. Section 3 will describe what kind of processing we implemented for the scenarios we carried out.

Analytics Console: the goal of this component is to let the user visualize and interact with the aggregated results of the analysis. Three visualization widgets (see Figure 3) have been implemented: MAPS, TAG CLOUDS and CHARTS, which are used to describe the output of the Social Extractor, Semantic Tagger and Sentiment Analyzer component, respectively.

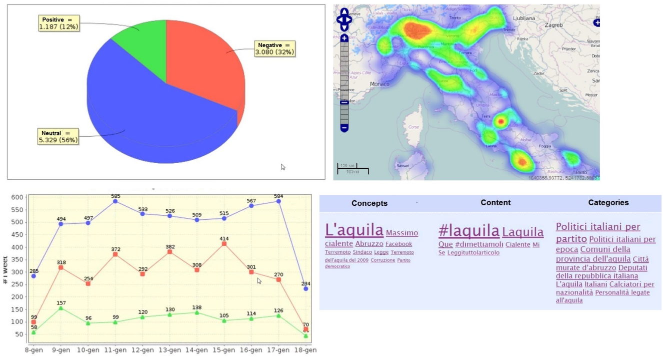
MAPS are used to show the geographical distribution of data gathered through the extraction job. This can be very useful for several scenarios (as an example, the check the citizens' opinion about recent administrative measures over different areas of the town, or to check how popular is a topic in a particular area). Data visualization is performed by adopting the popular *heat map* formalism: the more hot the colour, the more the content extracted from that particular location.

TAG CLOUDS are used to aggregate and organize the output produced by the Semantic Tagger. As shown in Figure 3, we designed three different types of tag cloud, one for each of the output produced by the Semantic Tagger. The *concepts* tag cloud reports the entities returned by the entity linking pipeline, while the *content* tag cloud shows the most popular terms and hashtags. Finally, the *categories* tag cloud is based on the Wikipedia categories attached to the entities identified in the test. All elements in the tag cloud are not static, since the user can click on them. By clicking on a tag, the platform will update the widget by showing the most popular tags which are used in co-occurence with the tag the user clicked on. Finally, CHARTS are used to report some information about the trends emerging from the data. As an example, the distribution of the sentiment over the posts stored for a certain analysis is plotted on a *pie chart*, while a *line chart* is used to show the amount of Tweets posted over time about a certain topic.

The exploitation of maps, charts and tag clouds lets the user analyze data and to have some aggregate views of the latent information hidden in the data, in order to obtain some valuable and reliable insights and analytics from the rough information gathered from social networks. In the next section we will show some of the output produced by the framework for both scenarios.

[5] http://tagme.di.unipi.it/

[6] http://sentic.net/api/en/concept/die/



Concepts	Content	Categories
L'aquila Massimo cialente Abruzzo Facebook Terremoto Sindaco Legge Terremoto dell'aquila del 2009 Corruzione Partito democratico	#laquila Laquila Que #dimettiamoli Cialente Mi Se Leggituttolarticolo	Politici italiani per partito Politici italiani per epoca Comuni della provincia dell'aquila Città murate d'abruzzo Deputati della repubblica italiana L'aquila Italiani Calciatori per nazionalità Personalità legate all'aquila

Figure 3: Analytics Console outputs

3. USE CASES

The combination of semantic processing techniques with sentiment analysis lets the users easily obtain some findings in the data. This paves the way to the development of several smart cities-related services relying on deep (semantic) analytics of social streams. In this section we describe the use cases in which the effectiveness of the platform has already been evaluated.

3.1 L'Aquila Social Urban Network

L'Aquila hit the headlines in April 2009 because of a tremendous earthquake which killed 297 people. Nowadays, the severe trauma to physical and psycho-social structures is still in the phase of recovery. In this scenario, ENEA[7] (Italian National Agency for New Technologies, Energy and Sustainable Economic Development) proposed the Social Urban Network (SUN) project, a smart cities-related project aiming at empowering and revitalizing the urban heritage and the social capital of the city after the dreadful earthquake. SUN relies on the insight that the analysis of the content produced by the citizens on social networks can produce a reliable *snapshot* of the current state of the recovering process. The multidisciplinary facet of the project lies in the fact that typical Artificial Intelligence and NLP techniques have been coupled with psychological research.

Indeed, in the first part of the project a set of social indicators to be monitored has been set. The eight social indicators, shown in Figure 4, have been defined by exploiting standard procedures of psycho-social research [6]. Next, we exploited our framework to extract social content and to implement a methodology which automatically mapped all the content posted on social networks by L'Aquila citizens to these social indicators. In this scenario, the SOCIAL EXTRACTOR has been launched as follows: as regards Facebook, specific pages and groups managed by citizens of L'Aquila (especially those focusing on the discussions about the consequences of the earthquake) have been analyzed. For Twitter, both the GEO heuristic (set on the latitudine and longitude of L'Aquila) and the USER one have been exploited. In the first case, all Tweets localized in a range of 50km from the city of L'Aquila have been extracted, while in the latter all Tweets posted by the main local newspapers, as well as the re-tweets and the mentions to such articles by other users, have been considered.

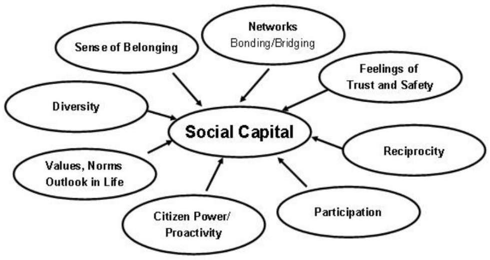
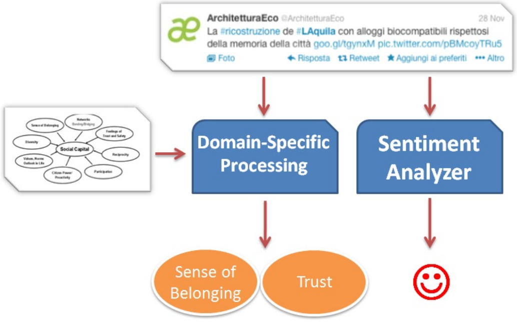


Figure 4: Social Capital indicators



ArchitetturaEco @ArchitetturaEco 28 Nov
La #ricostruzione de #LAquila con alloggi biocompatibili rispettosi della memoria della città goo.gl/tgynxM pic.twitter.com/pBMcoyTRu5

Figure 5: Mapping Social Capital - Example

Given the content extracted from social network, we used the component for DOMAIN-SPECIFIC PROCESSING to tackle the mapping problem as a text classification task. Indeed, this step has been carried out by exploiting a set of labeled examples to learn a multi-class classification model. Next, our model was able to associate each new Tweet or each new post to the social indicators (if any) it referred to. Next, in order to provide each social indicator with a *score*, each extracted content has been processed through the SENTIMENT ANALYZER component, as well. Next, the overall score of each social indicator has been obtained by summing the sentiment score of each content referring to that indicator (according to the classification model). This synthetic score represents the snapshot of the feelings of L'Aquila's citizens about a certain aspect.

The processing carried out by the domain-specific module developed for the SUN project is summarized in Figure 5. Given a Tweet coming from a citizen of L'Aquila[8], the classification algorithms associates to that Tweet (it is about the idea of introducing new sustainable buildings in the town) two social indicators: *Sense of Belonging* and *Trust*. Next, the Sentiment Analyzer algorithm associates to that content a positive sentiment score, which is inherited by both social indicators the content refers to. This process is performed in real-time, in order to continously update the scores associated to each social indicator over time, as new content is published on the social networks.

The final output of the project is presented in Figure 6, where the trend of each social indicator in the timelapse between April and October 2014 is provided. Due to space reasons, we just reported four out of eight social indicators.

[7] http://www.enea.it

[8] for the sake of simplicity, it is reported in its original version in Italian language

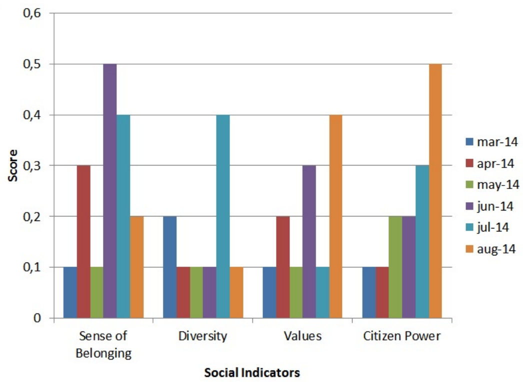


Figure 6: L'Aquila SUN - Social Indicators (Part 1)

The snapshot has been built on the ground of 490,000 social content extracted from Twitter and Facebook in this timelapse. Within the project, these data were exploited by a *community promoter* who monitored in real-time the aggregated score of each social indicator through a visual dashboard, and tackled specific actions aimed at empowering some facets of the social capital when some negative trend emerged.

3.2 The Italian Hate Map

The goal of the project, inspired by the Hate Map built by Humboldt University[9], was to analyze the content produced on social networks in order to *measure* the level of intolerance of the Italian country and to guide the definition of specific interventions (recovery and prevention, for example) on the territory. The analysis was performed by analyzing five different facets, called *intolerance dimensions*: homophobia, racism, violence against women, anti-semitism and disability. Differently from the USA Hate Map, the project aimed at automatically labeling intolerant content and deeply analyzing the Tweets in order to filter out ambiguous or polysemous terms.

In this scenario, our framework was exploited to identify and extract intolerant content. First, the Social Extractor was launched by defining a set of *sensible terms* for each intolerance dimension. The definition of the lexicons associated to each dimension was performed by psychologists with specific experience in this domain. The final list contained 47 terms[10] which were used to set the CONTENT heuristics. In this specific scenario, only Twitter was used as source to extract intolerant content, since due to Facebook policies, no groups or pages with a clear homophobic or racist intent is available on the platform.

Next, semantic processing and sentiment analysis were exploited to remove ambiguous and non-intolerant Tweets. In the case of ambiguous Tweets, all content with no intolerant intent has been filtered out, while in case of Tweets with a neutral or positive sentiment score, they were removed from the map. Finally, the remaining Tweets have been localized in order to produce an heat map as that shown in Figure 4. To this aim, as domain-specific processing some heuristics were introduced to increase the amount of geolocalized Tweets: we exploited social network official APIs to extract the *location* attribute for all users who posted intolerant content. When a specific location was indicated, all content coming from that specific user inherited the information about the location. Similarly, we extracted all the content posted by each user in a 7-days window. If other content (regardless they were intolerant or not) contained information about location, the location itself was used to label all intolerant Tweets from that user.

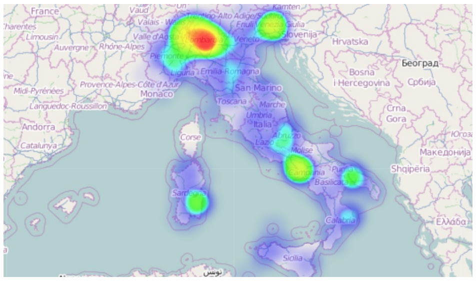


Figure 7: Italian Hate Map - Homophobia

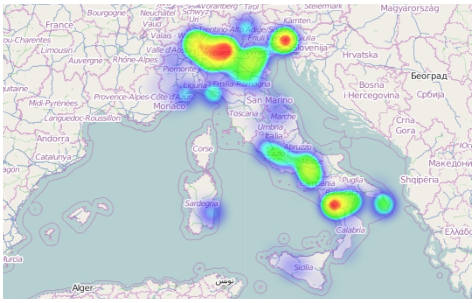


Figure 8: Italian Hate Map - Racism

Some statistics about the amount of Tweets analyzed are reported in Table 1, while Figure 7 and 8 reports the final heat maps. Due to space reasons, we only show two maps. It is also worth to note that the maps have been released as *open data*, in order to help to plan prevention and awareness activities in specific areas of the Italian territory.

Dimension	**#Tweets**	**#Geo**	**%Geo**
Homophobia	110,774	8,501	7,66%
Racism	154,170	1,940	1,24%
Violence	1,102,494	28,886	2,62%
Disability	479,654	3,410	0,75%
Anti-Semitism	6,000	1,150	18,03%

Table 1: Italian Hate Map - Extracted Tweets

4. RELATED WORK

The research line of social (or partecipatory) sensing [2] is based on the insight that the mash-up of crowd-based data can lead to the development of novel services and applications. The first work in the area dates back to 2006 [5], where the concept of *people-based urban sensing* is introduced. Many recent work investigated the effectiveness of this paradigm in urban scenarios. A typical application is represented by disaster management. A popular work in this

[9] http://users.humboldt.edu/mstephens/hate/hate_map.html

[10] The list contains very explicit terms. If requested, it can be put in Appendix of the paper.

research line is due to Abel et al. [1], who developed Twitcident, a platform for a incident or crises detection based on the combination of algorithms for real-time extraction of Twitter data streams with techniques for semantic analysis of content.

Regardless the specific application domain, the research area regarding the application of text analytics algorithms to social media data (as micro-blog ones) falls under the name of Social Media Analytics. According to Zeng et al. [10], social media analytics is supposed to provide tools to collect, monitor, analyze, and visualize social media data in an automated way. A typical application of such methodologies regards the marketing area. More recent attempts focused on the application of such techniques in different domains: the analysis of the sentiment expressed by people on social networks is also the focus of Felicittà, proposed by Allisio et al. [3]. In this paper, the authors propose a lexicon-based algorithm for sentiment analysis of geolocalized Tweets aiming at estimate the level of happines of different areas in Italy. Moreover, Paris et al. [7] shows that through the analysis of social media political institutions and government can easily track the opinion of the people about recent measures. The analysis of Twitter posts about politics is also investigated by Stieglitz et al. [8]. This framework represents the most similar attempt to develop a domain-agnostic framework for the extraction and the analysis of textual streams from social networks. Indeed, in this work the authors implements algorithms for the extraction of posts from Twitter, Facebook and Weblogs and provide users with several extraction heuristics and several visualization widgets.

By the way, differently from our framework, this platform does not implement neither algorithms for semantic content representation or techniques for sentiment analiysis. This is a very important issue, since, as showed by our use cases, semantics and sentiment analysis play a key role for most of the potential scenarios.

5. CONCLUSIONS AND FUTURE WORK

In this work we presented a framework for real-time analysis of human-generated textual streams. We depicted the architecture of the framework and explained the role of each module which compose it. Through our a framework we gave evidence that the combination of techniques for semantic representation, content classification and sentiment analysis can provide valuable findings which are latent in the data.

We also showed two real uses cases of the framework: the Italian Hate Map and the SUN project for the city of L'Aquila. In both cases the framework was able to easily reach the project goals with some simple adaptation to domain-specific requirements. As regards the Italian Hate Map project, we showed the power and the effectiveness of tools for monitoring and mining data for social goods. The outcomes coming from psychological analysis of the maps provide several insights which better explain and localize intolerant behaviors, and this can be useful to prevent them through specific initiatives. In this specific scenario we aggregated single people-based information to build a map describing the current situation of an entire country. As regards the SUN project, we showed that a multi-disciplinary approach combining psycho-social research with computer science can be exploited for mining social data to obtain a valuable and interesting snapshot of people feelings, sentiments and opinions about the current state of the town. Moreover, those information can be exploited to plan some specific intervention aimed at empowering or recovering the situation of the indicator whose score gets worse over time. As regards future work, we will extend processing modules by introducing techniques for social network analysis and to link content-based information with those coming from the Linked Open Data (LOD) cloud.

Acknowledgments. This work fullfils the research objectives of the project "VINCENTE - A Virtual collective INtelligenCe ENvironment to develop sustainable Technology Entrepreneurship ecosystems" funded by the Italian Ministry of University and Research (MIUR)

6. REFERENCES

[1] Fabian Abel, Claudia Hauff, Geert-Jan Houben, Richard Stronkman, and Ke Tao. Twitcident: fighting fire with information from social web streams. In *WWW Companion Volume*, pages 305–308. ACM, 2012.

[2] Charu C Aggarwal and Tarek Abdelzaher. Social sensing. In *Managing and Mining Sensor Data*, pages 237–297. Springer, 2013.

[3] Leonardo Allisio, Valeria Mussa, Cristina Bosco, Viviana Patti, and Giancarlo Ruffo. Felicittà: Visualizing and estimating happiness in italian cities from geotagged tweets. In *ESSEM@ AI* IA*, pages 95–106. Citeseer, 2013.

[4] Erik Cambria, Daniel Olsher, and Dheeraj Rajagopal. SenticNet 3: a common and common-sense knowledge base for cognition-driven sentiment analysis. *AAAI*, pages 1515–1521, 2014.

[5] Andrew T Campbell, Shane B Eisenman, Nicholas D Lane, Emiliano Miluzzo, and Ronald A Peterson. People-centric urban sensing. In *Proceedings of the 2nd annual international workshop on Wireless internet*, page 18. ACM, 2006.

[6] Franco Orsucci, Giulia Paoloni, Mario Fulcheri, Mauro Annunziato, and Claudia Meloni. Smart Communities: social capital and psycho-social factors in Smart Cities. 2012.

[7] Cecile Paris and Stephen Wan. Listening to the community: social media monitoring tasks for improving government services. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 2095–2100. ACM, 2011.

[8] Stefan Stieglitz and Linh Dang-Xuan. Social media and political communication: a social media analytics framework. *Social Network Analysis and Mining*, 3(4):1277–1291, 2013.

[9] Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, and Manfred Stede. Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2):267–307, 2011.

[10] Daniel Zeng, Hsinchun Chen, Robert Lusch, and Shu-Hsing Li. Social media analytics and intelligence. *IEEE Intelligent Systems*, 25(6):13–16, 2010.