



## Turning user generated health-related content into actionable knowledge through text analytics services



Paloma Martínez<sup>a,\*</sup>, José L. Martínez<sup>b</sup>, Isabel Segura-Bedmar<sup>a</sup>, Julián Moreno-Schneider<sup>a</sup>, Adrián Luna<sup>b</sup>, Ricardo Revert<sup>a</sup>

<sup>a</sup> Computer Science Department, Universidad Carlos III de Madrid, Spain

<sup>b</sup> MeaningCloud LLC, USA

### ARTICLE INFO

#### Article history:

Received 1 December 2014

Received in revised form 2 October 2015

Accepted 13 October 2015

Available online 10 November 2015

### ABSTRACT

In the last years, the habit of discussing healthcare issues with family and friends, even with unknown people, in the context of social networks has increased and processing user generated content has become a new challenge. This can help in on-line crowd surveillance for different applications (pharmacovigilance and filtering health contents in blogs among others) as well as extracting knowledge from unstructured text sources. In this article, a system that monitors health social media streams is described. It is based on several text analytics processes supported, among others, by MeaningCloud, a commercial platform which provides meaning extraction from texts in a Software as a Service mode. In this architecture, several domain resources are integrated to detect drugs and drug effects such as CIMA (official information about authorized drugs in Spain maintained by the Spanish Agency of Medicines and Health Products), MedDRA (Medical Dictionary for Regulatory Activities) and the SpanishDrugEffectDB database that contains relations between drugs and effects. Different ways of visualizing data considering time lines and aggregated data have been implemented. In order to show performance, an evaluation has been carried out over Named Entities Recognition (NER) and Relation Extraction (RE) tasks.

© 2015 Elsevier B.V. All rights reserved.

### 1. Introduction

Current definitions of Social Media [17] include several sources of user generated data, from Twitter to specialized blogs, through Facebook. Users of these Web 2.0 applications share information about any subject, including issues related to their health condition. The number of people with Internet access seeking for health information through the net ranges from 70% to 75% in the U.S. Besides, 42% of them used social media to get information about health issues. Moreover, mobile technology creates an ecosystem where people are continuously accessing to the Internet and this changes the way people interact with healthcare professionals.

In this context, there is an increasing volume of digital interaction that produces a big stream of data with meaningful information that companies and organizations need to access. In networks and forums such as PatientsLikeMe<sup>1</sup>, DailyStrength<sup>2</sup> or Saluspot<sup>3</sup> patients talk to each other about their feelings regarding

a health problem, the way their bodies react to a given drug, how they mix different drugs to fight against a specific disease they have and many other issues related to their health condition. They can access health-related content as well as connect and collaborate with other patients looking for health issues.

As an example of the importance of Social Media interactions in the health sector, according to a study developed by Price Waterhouse Coopers, 45% of consumers said information from Social Media would affect their decisions to seek a second opinion<sup>4</sup>. Age distribution of social media users must also be considered; there are surveys indicating that 89% of 18–29 year olds use social media in contrast with the 43% of people aged 65+. The difference of utilization by age groups will diminish over the next years and decades as digital natives increase their involvement and influence professionally and privately within their networks.

In order to analyze this market, the heavily regulated environment around health companies and prevention of direct-to-patient interactions must be taken into account, especially in

\* Corresponding author. Tel.: +34916249454.

E-mail address: [pmf@inf.uc3m.es](mailto:pmf@inf.uc3m.es) (P. Martínez).

<sup>1</sup> <http://www.patientslikeme.com/>.

<sup>2</sup> <http://www.dailystrength.org/>.

<sup>3</sup> <https://www.saluspot.com/>.

<sup>4</sup> PricewaterhouseCoopers, Social media “likes” healthcare, <http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml>.

Europe. This prevents pharmaceutical companies to get involved in social networks campaigns and only half of the top 50 pharmaceutical companies in the world interact with patients through social networks. It is also worth mentioning that, outside the U.S., there are a lot of regulatory restrictions forcing pharmaceutical companies to behave in a conservative way. Nevertheless, the interest in listening patients' opinions through social networks as a first step through bidirectional communication with patients is increasing.

Among all health issues, ADRs (Adverse Drug Reactions) are an important health problem due to the fact that they are the 4th cause of death in hospitalized patients [30]. Thereby, the area of pharmacovigilance has captured special attention because of the elevated and increasing frequency of drug safety events [4] along with their high associated costs [28]. Medicine regulatory agencies such as the US Food and Drug Administration (FDA) require clinicians to report every suspected ADR due to the fact that many of them are not spotted in the course of clinical trials. Nevertheless, studies such as [20] claim that ADRs are under-estimated considering that they are communicated by voluntary reporting systems.

Patients can report ADRs using Web-based spontaneous reporting systems (SRS) implemented by medicines agencies such as EMA (European Medicines Agency) and FDA. These SRS have different structures and contents and almost all of them are based on voluntary reporting, except for pharmaceutical companies, which are required to report suspected adverse events once they come to their attention. These companies report adverse drug events to the FDA when there is an identifiable patient, reporter and suspect drug. However, these requirements are not applied to social media.

Patient reports often provide more detailed and explicit information on ADRs than the ones from healthcare specialists [16]. They usually offer a wider or complementary view of the ADR and its possible impact on the patient. Another benefit of patient reporting is that adverse effects caused by OTC (over-the-counter, medicines that are sold without prescription) medications could be analyzed. An important contribution of SRS is getting patients to have a more central role in their treatments. However, despite the fact that these systems are well-established, the rate of spontaneous patient reporting is very low probably because many patients do not know them and may feel either confused when describing their symptoms or even unable to describe them.

On the other side, every medicine is carefully monitored after it is placed on the market, but there are some special drugs, labeled with a *black triangle*<sup>5</sup>, that are intensively controlled. This is due to the lack of information available about these medicines compared to others, for example, because they are new in the market or there are few data about its long-term use. In this context, it is therefore essential that the safety of all medicines continues to be monitored while they are in commercial use and that suspected ADRs are reported in order to keep up to date drug packages inserts corresponding to these drugs. Currently, this pharmacovigilance work is carried out by domain experts on a manual basis, by analyzing scientific literature as well as clinical trials documents and spontaneous reports.

Harpaz et al. [14] remarked that new methods that integrate data extracted from SRS narratives and knowledge extracted from experimental preclinical discovery drugs sources are required. Furthermore, patient-generated content concerns also discussions about treatments and opinions about drugs that could lead to valuable knowledge. Patients use Social Media to self-report adverse drug events three times more than reporting to FDA [10] and 90% is the estimated rate of ADRs that patients do not report

[23]. Thus, it is reasonable to think that health-related social media can be used as a complementary data source to collect ADRs as well as data about the incorrect use of drugs. In other words, monitoring the abuse and misuse of medicinal products, for instance by people who have problems understanding medical language.

Transforming health-related social media streams into useful knowledge by extracting information from messages is not a trivial issue and requires sophisticated tools to tackle challenges with the overall objective of protecting public health. The two main challenges are (1) to analyze patients sharing experiences and (2) to manage highly informal patient-oriented language, something difficult to deal with as there are barely any resources regarding it.

To cope with these information challenges, Natural Language Processing (NLP) technology is a key aspect and should integrate usable tools to deliver real-time insights to decision makers about surveillance and pharmacovigilance tasks. In this context, this contribution describes the application of text analytics processes to extract information from these real-time Social Media streams relevant for the healthcare sector and the challenges that must be faced. The information to be extracted goes beyond drugs and diseases mentions to show also relationships among medications and ADRs and indications. It also covers trend evolution of those named entities mentions and ADRs detected in patients' conversations.

Text analytics processes to be applied for this purpose cannot be generic, but adapted to the health domain. This requires specific dictionaries and ontologies covering drugs, diseases, body part names and other topics to be integrated in named entity recognition processes and to cope with colloquial expressions and laypeople terms for drugs, diseases and other entities with the aim to discover what is being talked about.

The work presented in this article integrates different semantic resources and processes in a complete framework to face real-time text analytics on social media, in particular for monitoring drug-related medical events (ADRs, indications, symptoms, diseases, ...) in online Spanish social networks about healthcare, being able to process a large volume of data in real-time<sup>6</sup> and to address the abovementioned challenges. This system is a result of TrendMiner European project.

Remainder of the article is organized as follows. Section 2 is devoted to review the state of the art in drugs and ADRs entity recognition in medical literature and social media. Section 3 describes the functionality of the proposed system, its architecture and integrated semantic resources. Section 4 describes the evaluation carried out and finally, conclusions and future work are given in Section 5.

## 2. Related work

Due to the fact that users are active consumers and producers of health-related contents in Internet, in the last years extracting knowledge from unstructured contents (mainly texts) in health domain has received a great attention. The main reason is these sources could reveal important public health issues. Many efforts have been devoted to the application of NLP techniques to gather information about health issues, such as diseases, symptoms, drugs, adverse events and others from texts. A comprehensive overview of the application of text mining techniques to biomedical knowledge extraction from scientific literature, clinical narratives and on-line health web sites is given in [35].

Focusing on analysis of social media to mine data about patient health, there are many works that use Twitter both to detect pre-established health conditions and unknown trends. Parker et al. [22] described a method to identify emerging public

<sup>5</sup> [http://www.ema.europa.eu/ema/index.jsp?curl=pages/special\\_topics/document\\_listing/document\\_listing\\_000365.jsp](http://www.ema.europa.eu/ema/index.jsp?curl=pages/special_topics/document_listing/document_listing_000365.jsp).

<sup>6</sup> The system can be accessed at <http://trendminer.daedalus.es>.

Download English Version:

<https://daneshyari.com/en/article/508852>

Download Persian Version:

<https://daneshyari.com/article/508852>

[Daneshyari.com](https://daneshyari.com)