Social Minder: a Tool for Social Media Monitoring and its Use for Detecting COVID-19 Misinformation

Marcos Fernández-Pichel[1], David E. Losada[1] and Juan C. Pichel[1]

[1]*Centro Singular de Investigación en Tecnoloxías Intelixentes (CiTIUS), Universidade de Santiago de Compostela, Rúa Jenaro de la Fuente s/n, 15782, Santiago de Compostela (Spain)*

Abstract

In this work, we introduce Soсıal Mınder, a Big Data platform for Social Media monitoring that allows massive extraction of textual information, and stands on a modular and scalable architecture for efficient real-time and batch processing. This demo is oriented to present a use case that provides users with estimates of credibility for webpages linked in Social Media. Soсıal Mınder can serve multiple research and commercial purposes but we use it here for identifying COVID-19 related misinformation posted on Twitter.

Keywords

Big Data, Real Time, Web Streams, Credibility

1. Introduction

Social Media (SM) has become one of the main sources of information for end-users [1]. However, processing SM data is a challenge and doing it in real time is critical for many added-value applications. For example, according to Twitter [2], the number of daily posted tweets is higher than 500 million (around 5,787 tweets per second). Companies and researchers need tools able to digest this huge amount of information and present it in a convenient and understandable way.

However, SM can be a source of misinformation, which is specially damaging when it comes to health-related content. During the COVID-19 pandemic, for example, dubious and poor quality information about the disease and its treatments was broadcasted on SM [3, 4], sometimes resulting in situations of personal harm [5]. In this work, we present Soсıal Mınder[1], a Big Data platform for batch and real-time social media monitoring, which has been adapted to detect misinformation about COVID-19.

[1]http://tec.citius.usc.es/social-minder/

1.1. Related tools

SM analytical tools are often constrained to work from data provided by official APIs, as Batrinca and Treleaven showed in their thorough survey [6]. One of the advantages of Socıal Mınder is that it allows massive extraction of tweets with its own crawler [7] and works with a modular and scalable architecture that can efficiently ingest large amounts of textual data (see Section 2).

Existing tools for social media monitoring, such as Social Mention, provide a rigid set of functionalities (e.g., general statistics about queries). Socıal Mınder differs from these because it includes a real-time credibility estimation module with self-developed technology. This module was built in the context of previous experimental studies [8], which are freely available for the research community[2].

Although there are some existing initiatives for real-time credibility analysis on Twitter [9], to the best of our knowledge, our platform is the first to integrate this functionality into a complete monitoring system expandable to other web sources, not only Twitter.

Related to our use case, the study by Sharma and colleagues [10] also addressed COVID-19 misinformation on Twitter. However, the main difference here lies in the way that misinformation is detected. These authors proposed a manual annotation technique, based on certain expressions and hashtags, while Socıal Mınder incorporates an automatic algorithm, see Section 2 for more details.

2. Architecture

Socıal Mınder was built on the top of eXtream [11], which is a Big Data framework that permits advanced users to design their own processing topologies. Socıal Mınder is an evolution oriented to the end-user, providing a dashboard for non-expert users. Its system architecture consists of a fixed consumption topology that interconnects several containerised modules (see Figure 1). The functionality of each one is briefly explained below:

- A Twitter crawler [7] that injects text streams into the topology. For a given query, it first tries to recover all historical tweets, and then starts to consume in real-time.
- A sentiment analysis module based on VADER [12], a rule-based classification technology.
- A credibility estimation module that uses a self-developed classification technology, based on our experimental results [8]. It consists of a support vector machine trained on three credibility classification datasets [13, 14, 15]. Since the training data comes from the Web Search domain, only the web pages linked in the tweets are assessed for credibility. Tweets that do not contain any link are skipped.
- A timestamp-aggregator module that groups texts by different temporal granularities (hour, day, month) to perform the analysis.
- Four parallel computation modules that perform different statistical analysis tasks (count texts, extract keywords using TF-IDF techniques, compute aggregated sentiment and credibility) for all temporal granularities available.
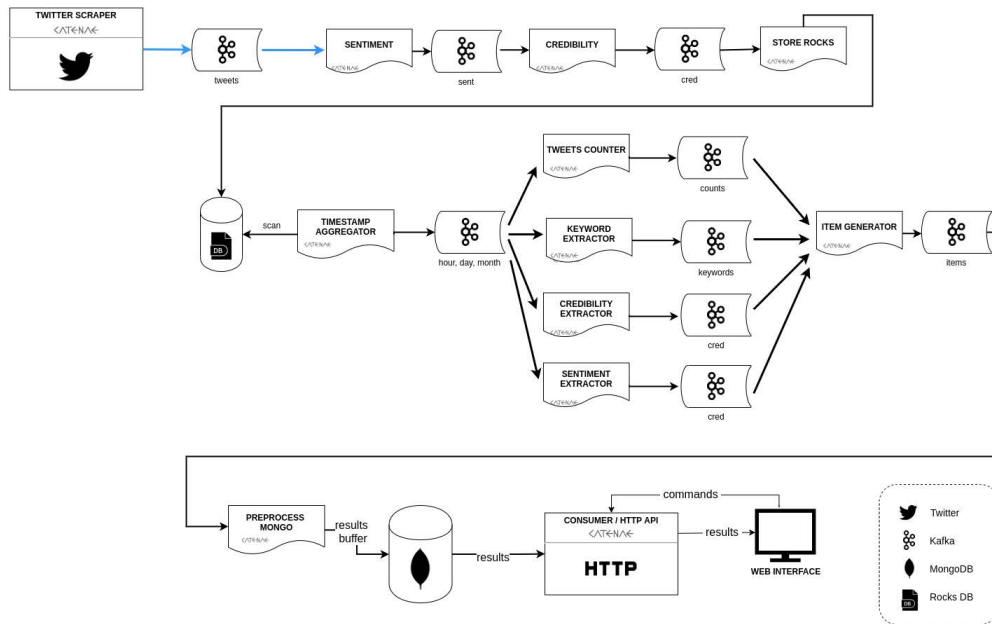
---

[2]https://github.com/MarcosFP97/Health-Rel

Figure 1: Social Minder architecture.

- Two final modules that aggregate results and write them on permanent storage (Mongo DB[3] database).
- A dashboard for non-expert end-users that shows statistics and graphs per query for different granularities.

3. COVID-19 misinformation use case

Socıal Mınder can serve multiple research or commercial purposes. For example, one can develop new SM applications by modifying the profiles of interest. This demo focuses on a use case of Socıal Mınder oriented to monitor misinformation posted on Twitter about COVID-19. We exemplify the tool with a dashboard associated to a sample of covid-related tweets obtained in 2020[1].

Socıal Mınder allows to filter the Twitter stream either by account or by a textual filter. We illustrate this by considering four cases: two reputed accounts (*"@who_europe"*, *"@dhscgovuk"*) and two filters (*"coronavirus treatment"*, *"alternative medicine coronavirus"*). One can expect that the two accounts publish more reliable contents, while the tweets associated to the filters include more dubious information. The sample used to run this demonstration was extracted during the first lockdown period (May 2020) over the full Twitter stream.

The dashboard consists of an upper part with configurable elements, general statistics, keywords extracted from the tweets (computed using TF/IDF) and an initial graph that plots the number of tweets (see Figure 2). The user can choose to analyse tweets submitted by an
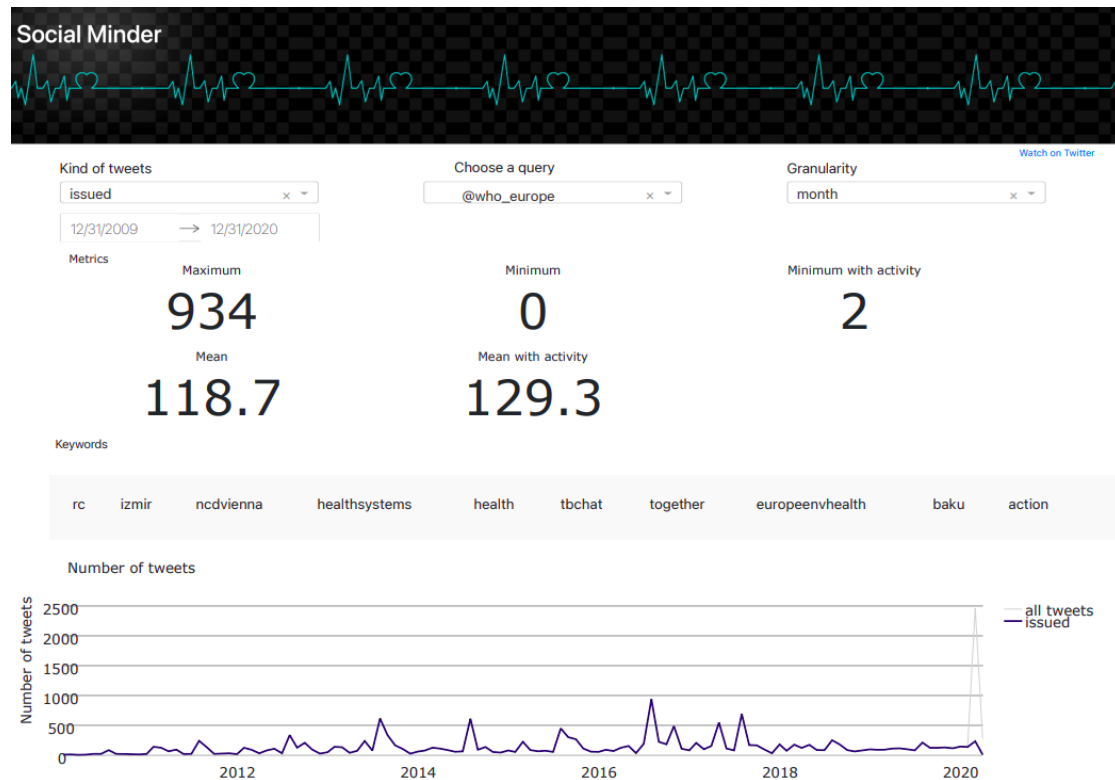
---

[3]https://www.mongodb.com/

Figure 2: Social Minder dashboard (upper part).

account ("issued" in the interface) or "mentions" to an account or to a given keyword query. For this demo, we pre-configured some example queries and the user can click on them and obtain the corresponding results. The granularity of the analysis is also configurable (days, weeks, months). In this upper part, general count statistics and keywords provide the user with a first glimpse of the account/topic in social media.

The bottom part consists of bar graphs that represent the evolution of the sentiment and the credibility of the posted contents (see Figure 3). Using this tool, one can observe, for example, that *@who_europe* tends to publish more credible contents (as estimated by the classifier) that the contents associated to tweets that mention words like *"coronavirus treatment"*.

It might be surprising that some contents from a reputed organisation such as the WHO are classified as "highly uncredible". This may be due to false negatives in our predictive technology, which has still room for improvement. However, as mentioned above, the tool identifies general trends and, in general, is able to distinguish the relative quality of authoritative accounts versus more dubious contents (e.g.,"alternate medicine coronavirus").
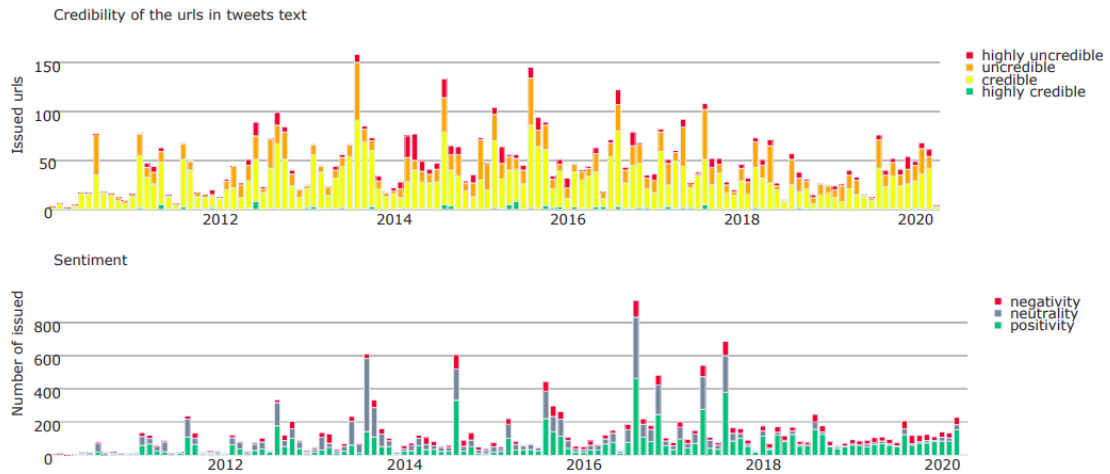
Figure 3: Social Minder dashboard (bottom part).

4. Conclusions

In this work, we presented an end-user oriented tool called Social Minder. It allows monitoring Twitter but it could be expandable to other web sources, and it provides different estimations (like sentiment or credibility) that can be useful for commercial or research purposes, like monitoring a company's account or analysing misinformation trends. Some core modules, such as the experiments[2] that inspirited our credibility estimation technology or the Twitter crawler[4] are freely available to the community.

In this demo, we have shown one possible use case, but this technology could be adapted to monitor new dynamic text streams, new queries, and/or add new modules, just to name a few.

Acknowledgements

References

[1] Reuters Insitute, University of Oxford, Reuters Digital News Report 2021, 2021 (accessed September 07, 2021). URL: https://reutersinstitute.politics.ox.ac.uk/digital-news-report/

---

[4] https://github.com/labteral/bluebird

2021.

[2] Twitter, Twitter for Business, 2021 (accessed September 07, 2021). URL: https://business.twitter.com/en.html.

[3] M. S. Islam, T. Sarkar, et al., Covid-19−related infodemic and its impact on public health: A global social media analysis, The American Journal of Tropical Medicine and Hygiene 103 (2020) 1621−1629.

[4] G. Pennycook, J. McPhetres, Y. Zhang, J. G. Lu, D. G. Rand, Fighting covid-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention, Psychological science 31 (2020) 770−780.

[5] N. Vigdor, Man fatally poisons himself while self-medicating for coronavirus, doctor says, 2020. URL: https://www.nytimes.com/2020/03/24/us/chloroquine-poisoning-coronavirus.html, [Online; posted 24-March-2020].

[6] B. Batrinca, P. C. Treleaven, Social media analytics: a survey of techniques, tools and platforms, Ai & Society 30 (2015) 89−116.

[7] R. Martínez-Castaño, J. C. Pichel, P. Gamallo, Polypus: a big data self-deployable architecture for microblogging text extraction and real-time sentiment analysis, arXiv preprint arXiv:1801.03710 (2018).

[8] M. Fernández-Pichel, D. E. Losada, J. C. Pichel, D. Elsweiler, Reliability Prediction for Health-Related Content: A Replicability Study, in: D. Hiemstra, M.-F. Moens, J. Mothe, R. Perego, M. Potthast, F. Sebastiani (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2021, pp. 47−61.

[9] A. Gupta, P. Kumaraguru, C. Castillo, P. Meier, Tweetcred: Real-time credibility assessment of content on twitter, in: International conference on social informatics, Springer, 2014, pp. 228−243.

[10] K. Sharma, S. Seo, C. Meng, S. Rambhatla, Y. Liu, Covid-19 on social media: Analyzing misinformation in twitter conversations, arXiv preprint arXiv:2003.12309 (2020).

[11] M. Fernández-Pichel, R. Martínez-Castaño, D. E. Losada, J. C. Pichel, eXtream: a System for Real-time Monitoring of Dynamic Web Sources, in: Proceedings of the Joint Conference of the Information Retrieval Communities in Europe (CIRCLE 2020). http://ceurws. org, volume 2621, 2020.

[12] C. Hutto, E. Gilbert, Vader: A parsimonious rule-based model for sentiment analysis of social media text, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 8, 2014.

[13] P. Sondhi, V. Vydiswaran, C. Zhai, Reliability prediction of webpages in the medical domain, in: European conference on information retrieval, Springer, 2012, pp. 219−231.

[14] J. Jimmy, G. Zuccon, J. Palotti, L. Goeuriot, L. Kelly, Overview of the clef 2018 consumer health search task, CLEF 2018 Working Notes 2125 (2018).

[15] J. Schwarz, M. Morris, Augmenting web pages and search results to support credibility assessment, in: Proceedings of the SIGCHI conference on human factors in computing systems, 2011, pp. 1245−1254.