MonAnt - Monitoring Online Antisocial Behaviour

Milan CÁK, David TRAN DUC, Dung LAM TUAN
Matúš BARABÁS, Peter LIPTÁK, Veronika ŽATKOVÁ
Patrik ŽIDULIAK*

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 2, 842 16 Bratislava, Slovakia
fiit.tp14@gmail.com

Millions of people all over the world using the Internet everyday are, whether being aware of it or not, surrounded by antisocial behaviour. One of the most serious types of such behaviour is especially misinformation and false information spreading [1].

The huge number of online social networks and news portals has drastically changed the way of communications and information sharing. Being a part of this cyber space and knowing how to consume the right information is the most difficult part of all. The advent of social networks makes every user a source of information without checking for any facts. Because of that, the misinformation spreads very fast and everyone can access this information without any difficulty. As the usage of internet increases, the chance to spread misinformation and disinformation also increases many fold. The spread of misinformation has been revealed in many topics such as health, politics, finance and technology. Online environment provides a very good opportunity to misuse it.

Antisocial behaviour in online environment is one of the most recent and serious problems. This topic is researched by a permanently increasing body of research. In many publications, the three main crucial steps of regulation and elimination of antisocial behaviour are identified [2]:

- Characterization: in this step, we try to characterize antisocial behaviour by analyzing its characteristics.
- Detection: the goal of this step is to automatically or semi-automatically detect a antisocial behaviour.
- Elimination: we try to reduce the antisocial behaviour in internet. Many approaches are used, e.g. banning and filtering antisocial content.

As mentioned above about crucial step of regulation, we identified the important step that must be done before all is, how we can collect data for characterization, detection and elimination of antisocial behavior. Our project focuses on collecting antisocial behaviour data with purpose for further research and studies in the monitoring, detection and mitigation of antisocial behaviour area. The main goals of the project are:

- Having as much as possible data in one place, where we can easily apply text analysis and machine learning algorithms to identify the correctness and trustworthiness of data.
- Creating an application used to effectively manage sources of this data. Our goal is to collect as much data as possible, but we do not try to collect all data that is shared in Internet. Having a management application, we can focus in specific fields and have data well organized.
- Managing storage, which is used to store all collected data. We expect external sources, which can push their collected data to our storage. Therefore we develop API, which will be primary interface to the storage. Users, applications with correct authentication and authorization rights can make requests to the storage.

To achieve our goals, we propose a platform named MonAnt. Architecture of our project is shown in figure 1. While designing our architecture, we focus on availability, robustness and scalability of our system. Our system consists of three main modules:

- Web management: Standalone web application for web monitoring management. Users can predefine rules by setting web pages, which information will be retrieved from. The application is

* Master study programme in field: Information Systems
Supervisor: Dr. Ivan Srba, Institute of Informatics, Information Systems and Software Engineering, Faculty of Informatics and Information Technologies STU in Bratislava

IIT.SRC 2019, Bratislava, April 17, 2019, pp. 1–8.

being developed in Python3 with usage of Django framework. The main reason, we chose Django is its portability and simplicity of Python as programming language.

– Web monitoring: Implementations of data providers for specific topics. We use Python3 as programming language for our implementation. To simplify implementations we use Scrapy library for web crawling, BeautifulSoup for scraping HTML files and feedparser for parsing RSS feeds.
– Central storage: API layer for storage of collected data. This plays crucial role in our system. It also mediates communication and data transfer between other platform modules. As database, we choose to use Postgres, one of the most used relational databases currently.

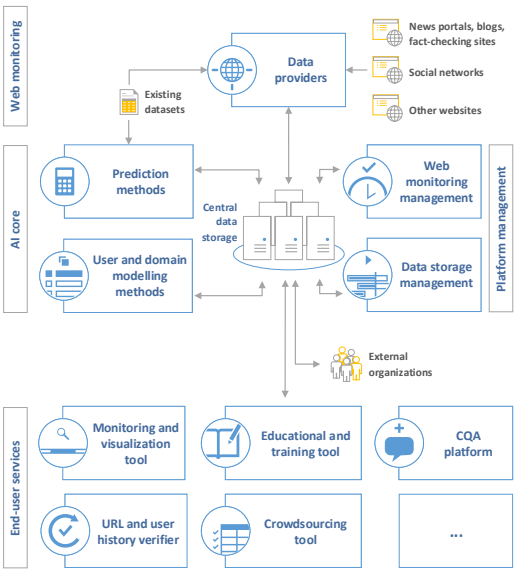

Figure 1. Project architecture [3].

By integration of all three modules, we have scalable system, where we can define rules for monitoring of information spread. Collected datasets are then well organized, easily to use and available for all potential researchers.

The trigger of system is a user, who wants to monitor some sources of information, e.i news portals. The user defines the list of sources, which he wants to monitor the spread of information on. This action activates data providers, which start crawling data from predefined web sources. The collected data are then sent by REST API to storage and be stored by database schemes. With stored data, we provide them to people, who want to analyze them to identify the trustworthiness.

Our platform provides a possibility to be extended by AI core modules. AI core modules are components that ensure handling, scheduling etc. on methods over data from central storage that are collected by our system. It will support whole machine learning flow, search and natural language processing. They can be added in anytime, so many AI algorithms can use our datasets at one moment.

In the final version of system, we want to integrate whole system with Docker[1]. By doing so, from the technological point of view, we can painless manage resources, performance, some aspects of security of the whole system.

This platform is proposed and will be implemented in projects Rebelion and Misdeed. Rebelion is research project that focuses on Automatic Recognition of Antisocial Behaviour in Online Communities. Misdeed is research project that focuses on Misinformation Detection in Healthcare Domain.

Conclusion

Antisocial behaviour in online world is one of the most recent and serious problems. We propose the platform MonAnt for monitoring, detection and mitigation of antisocial behaviour.

References

[1] Fortuna, P., Nunes, S.: A Survey on Automatic Detection of Hate Speech in Text. *J. ACM*, 2018.

[2] Kumar, S., Shah, N.: False Information on Web and Social Media:A Survey, CRC press., 2018.

[3] Srba, I., et al.: Monant: Universal and Extensible Platform for Monitoring, Detection and Mitigation of Antisocial Behavior, 2019.

[1] provides a uniform interface for application isolation to containers