

Proposal for Master's Degree Programme

Title: Improvement and Evaluation of a System for Dynamic Management of Users and Retrieval of Posts from Twitter

Advisor and Contact

Pedro R. M. Inácio (inacio@di.ubi.pt), Ph.D.

João Paulo Cordeiro (jpaulo@di.ubi.pt), Ph.D. (co-Advisor)

Summary / Scope

Nowadays and most of the times, scrutiny of public sources is made manually on a regular basis (e.g., a security expert searches and selects news everyday). After a while, this task is repetitive and cumbersome, and sometimes prone to error. More recently, social networks like Twitter have been fueling and changing the way people follow the news and trends, since several experts also share their own selection, effectively taking the role of curators for an increasingly number of sources of information. The fact that Twitter forces messages to have less than 140 characters has a secondary effect of also summarizing some of those needs in an efficient manner. Nonetheless, the number of contributors increases everyday, and the community is also affected by those that share about anything, without properly checking the sources, curating the content or summarizing the news correctly. Keeping up with such a dynamic community and dealing with the noise is a difficult and challenging task.

A previous project on this topic [1] successfully delivered a prototype of a system to automatically add, follow, and remove sources of information (twitter users) and to automatically classify posts. The system is supported by two main algorithms, which use metrics such as the number of followers, posting frequency, and the number of likes or reshares of posts, to manage users and select potentially relevant information. The prototype worked as a proof of concept and needs refining and in-depth evaluation.

The overall objective of this master's programme is to evolve and evaluate the aforementioned system. This work will focus on fine-tuning the existing algorithms, metrics and formulas, assessing if additional metrics could be used to make the system better and improving the way information is displayed in the user frontend of the prototype.

During this master's programme, the student will have the chance to engage in discussions with other people involved in the area of natural language processing and information security, as well as the opportunity to improve his or her knowledge in several computer science fields, namely information security, social networks, programming, web technologies, natural language processing, automatic text summarization, and information retrieval and extraction.

Objectives

This master's program has four main objectives:

1. Study the impact of tweaking weights associated with the metrics used in the formulas;
2. Assess if new metrics are useful for the classification of twitters and posts, integrate them accordingly, and study the impact of their inclusion against the previous version of the system;
3. Improve the overall look and feel of the part of the system facing the user;
4. Evaluate the prototyped system both from the point of view of effectiveness and usefulness of the classification, and from the point of view of user-friendliness, namely in terms of how the information is displayed.

Tasks

In order to achieve its objectives, the following tasks are proposed as an initial work plan for this master's project:

Task 1 Contextualize with the problem at hands and with the objectives of this project, as well as with the technologies involved and with the previously developed prototype. Revision of the specialized literature and related works. Definition of the method for evaluation of the system (2 months);

Task 2 Detailed analysis of the formulas devised for the previous iteration of the system and study of the impact of tweaking weights associated with the metrics used in the formulas (1 month);

Task 3 Proposal of of new metrics for the classification of posts and twitters, followed by their integration in a fork of the system (1 month);

Task 4 Assessment of the usefulness of the new metrics in the classification effectiveness and comparison with the previous version of the system (1 month);

Task 5 Improvement of the way information is presented to the user of the system (1 month);

Task 6 Evaluation of the system (1 month);

Task 7 Writing of the master's dissertation, technical documentation and a conference paper (3 months, eventually distributed and interleaved with the time periods of other tasks).

Timetable

The following table presents an approximate scheduling for the execution of the previously identified tasks. The execution of a given task in a given month is marked with a cross (x).

Task \ Month	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul
1	x	x									
2			x								
3				x							
4					x						
5						x					
6							x				
6								x	x	x	

Expected Outcomes and Dissertation

The main expected outcomes of this master's project are the improved formulas and algorithms for classifying posts and twitters, duly integrated in the prototyped system (another output of the project), as well as an evaluation of its effectiveness and user-friendliness. Potential interesting findings and the description of the functioning of the system may be the subject of a scientific conference paper. The final dissertation, entitled "*Improvement and Evaluation of a System for Dynamic Management of Users and Retrieval of Posts from Twitter*", should be written in English, aiming for the international dissemination of the work.

References

- [1] Ivan Sousa Fernandes, Dynamic OSINT System Sourcing from Social Networks, M.Sc. in Computer Science and Engineering, Universidade da Beira Interior, 2016;