

Political Viewpoint Analysis from Social and Web Sources

Xavier Tannier (LIMSI/ U. Paris Sud, CNRS) Ioana Manolescu (INRIA, U. Paris Sud)

Duration: 5-6 months, the starting date is flexible (ideally March 1st, 2016)

Location: Orsay, France

Keywords: **Natural Language Processing, Text Mining, Information Extraction, Information Retrieval, Social data management, Databases** (the tasks will be adapted according to interest and level of qualification of the candidate).

The work is to be carried in close collaboration between the INRIA OAK team¹, LIMSI-CNRS² and the major French newspaper Le Monde³. The work is supported by a Computational Research Journalism Award from Google⁴.

1 Context

The French political arena comprises many political parties, spanning from the extreme left to the extreme right. In 2012, no less than 10 candidates ran for the presidential election (12 in 2007), with sensitivities such as communism, socialism, ecology, centre left, centre right, right, far right, extreme right, or even one literally called “hunting, fishing, nature and tradition”. Further, as in any democratic countries, ideas and values intertwine with group tactics, personal ambitions, communication strategies, that make politicians take stands which do not always directly copy those of their political party.

For all these reasons, deciphering politician candidates’ positions and deciding which candidate is the closest to our own opinions is a complex task for citizens. Similarly, analyzing statements and facts into perspective is complex work for journalists.

The goal of the internship is to **automatically build (from a variety of sources such as online news articles, Twitter feeds, structured databases etc.) topical threads** that will organize and visualize **claims made by politicians**, in order to help journalists and citizens decode them and distinguish between personal opinion, communication tools established by the parties, and voluntary distortions of the reality.

Data sources. We will use different textual and/or structured data sources as input to our extraction process:

- Newspaper articles from Le Monde web site and printed version.
- Social network data, that comes already endowed with metadata specifying e.g., the author and date of every information item published in a social context, and possibly previously published items to which the new item refers e.g., re-tweets);
- The link structure (e.g., news articles citing each other, links appearing in tweets), from which we will extract information on a thread continuation;
- Background knowledge concerning political affiliations of people, as well as their position in the political chessboard (importance and side), can be exploited to contextualize the social item content;

¹<https://team.inria.fr/oak/>

²<https://www.limsi.fr/index.en.html>

³<http://www.monde.fr>

⁴<http://googleresearch.blogspot.fr/2015/06/google-computational-journalism.html>

- Possibly, data such as voting intention polls, in order to track events to have an influence on them.

Scientific areas. This project involves scientific fields such as **Natural Language Processing, Information Extraction, Information Retrieval, Social data management, Database, Data visualization.**

2 Description

Depending on the level of qualification and duration of the candidate internship, (s)he will work on one or several of the following steps:

- *Using existing metrics (such as mutual information, tf-idf, lexical specificity) for extracting differences and similarities* between claims from different political sides, or different political personalities.
- Implementing and adapting *unsupervised algorithms* such as topic models (*e.g.* LDA) on different types of claims.
- Building a *unified framework for collecting, organizing and querying claims* from various sources.

Required competencies are: good software development skills, strong qualification in one or several of the scientific areas involved in the project (demonstrated *e.g.*, by academic results or past successful projects), good communication skills and willingness to work in a team.

On a daily basis, work will take place in a collaborative team comprising the internship supervisors, an INRIA engineer whose task it is to oversee and coordinate the development of our unified platform for data-based fact checking, and probably other interns. The work is related to a longer scientific effort within the four-years ANR project ContentCheck (Content Management Techniques for Fact-Checking: Models, Algorithms, and Tools); the project starts in January 2016. The internship may lead to a PhD within the project.

The internship will take the form of a full-time INRIA employment contract. The intern will be paid 1100 €/month.

3 Contacts

- Ioana Manolescu (ioana.manolescu@inria.fr), <http://pages.saclay.inria.fr/ioana.manolescu/>
- Xavier Tannier (xavier.tannier@limsi.fr), <https://perso.limsi.fr/xtannier/en/>