OBTAINING SOCIETAL FEEDBACK ON LEGISLATIVE ISSUES THROUGH CONTENT EXTRACTION FROM THE SOCIAL WEB

Dimitris Spiliotopoulos[1], Guenther Schefbeck[2], Dimitris Koryzis[3]

[1]Innovation Lab, Athens Technology Centre, Athens, Greece, email: d.spiliotopoulos@atc.gr
[2]Parliamentary Archives, Austrian Parliament, Vienna, Austria, email: guenther.schefbeck@parlament.gv.at
[3]Digital Archives, Hellenic Parliament, Athens, Greece, email: dkoryzis@parliament.gr

Keywords: *social web, legislation, digital archiving, language analysis*

Abstract: *Social web is the main channel for people to express their opinion on life issues, including politics. It is, therefore, the most interesting and promising medium to exploit in order to harvest the feedback of the society. Legislation directly affects the people, thus their involvement in such discussions on the web is extremely high. This work reports on the set of specific tools required for obtaining rich, clean, structured input on the peoples' response on legislative issues through the social web.*

1. Introduction

The social web presents a well of knowledge that any student or researcher of the society heartbeat should not ignore. The integration of several language technologies, such as entity, event and topic recognition, opinion mining and aggregation describes the core requirements for successful social web analysis. However, it is not an easy task. There are, certainly, main solutions that mine the web for opinions and certain keywords like people (e.g. politicians, celebrities, etc.) or company names and brands (e.g. Coca Cola, Apple, etc.). The former target the social conversations and respective users. The latter are brand monitoring. People use social networks to respond to the news, share information, address problems, and entertain others.

In the social networks, the opinions of the people can be found in many forms, direct comments, reviews, polls, links to discussions, and so on. The social media also come in many forms, user-centered social pages (like Facebook), microblogs (like Twitter), blogs, wikis, news pages with comments, multimedia (like Flickr and YouTube). The comments of the users can span more than one of these networks, for example a user may post a comment on Twitter with an in-text link to a YouTube video that bears a direct connection to that comment.

Apart from the above, there is also a lot of meta-information that can be harvested and, if interpreted correctly, can present a very valuable and understandable view of how people respond to the stimuli. On the political domain, events like political announcements about economy, elections, draft laws and so on are of high importance and generate huge feedback in the social media.

This paper presents how the ARCOMEM system[1], a dedicated approach on social information analysis and preservation may collect and analyze user feedback from social media in order to provide the researcher the necessary analytics to evaluate the opinion of the people on legislative issues.

---

[1] ARCOMEM: Archive Communities Memories, www.arcomem.eu, FP7-ICT-270239.

2. Motivation

Regarding politics, there is a definite increase on the people involvement in discussion about new laws that are drafted. The society, via the several available social media and blogs, refer to notions, ideas and details on certain aspects of legislative nature. But they mostly discuss on the impact a new law has or would have to the society. Moreover, the detailed information on how the actual society views their members is of high importance. There are references and distinctions between middle, low and high class, educated and non-educated groups, to public servants and private sector, students and teachers, politicians and the voters, and so on. There are opinions on specific actions of a government, on specific political parties, politicians and political analysts, for media organizations that report the news, and so on. There are also references and opinions expressed on actual events that are of consequence to the people, a treaty that was signed, new economy measures that were announced, legislation drafts that are to be discussed in the Parliament, a visit from a foreign president and so on.

Everything that happens in politics may they be events, actions of people, or otherwise could trigger long feedback from the societies that are affected. And the social web is the most fertile place for such discussions to take place. And it has become such an important medium that governments realize how much it means to be able to listen to the will of the people from the social web (Schefbeck et al., 2012). A recent example is the Finnish government's amendment to the national constitution allowing the citizens to submit petitions to the Open Ministry[2]. This is a crowd sourcing platform for legislation drafts that the government is obliged to examine before putting them to vote. The opinion of the citizens matters to the citizens themselves but also to their government and the social web the place where the opinions are expressed.

Harvesting the social web looking for the people's feedback on legislative issues is a very promising, citizen-friendly, democratic way to obtain feedback. In order to achieve quality results, certain innovative technologies, such as intelligent crawling, semantic analysis, opinion mining, can be deployed.

3. The ARCOMEM system

The ARCOMEM system is an integrated environment for socially-aware digital preservation (Risse et al., 2012). It is using social web contextualization as well as extracted information on events, topics, and entities for creating richer and socially contextualized digital archives. This requires thorough analysis of the crawled web pages and their components. The components of a web page are called *web objects* and can be the title, a text paragraph, an image or a video, but also other text snippets like user comments that are usually found at the end of an article, or social media posts as they are displayed in a dedicated text box on the web page.

The ARCOMEM system may utilize semantically enhanced crawl specifications that extend traditional URL based seed lists with semantic information about entities, topics or events. These crawl specifications are complemented by small reference crawls to learn more about the crawl topic and intentions of the user. The combination of the original crawl specification with the extracted information from the reference crawl is called the intelligent crawl specification. This specification, together with relatively simple semantic and social signals, is used to guide a broad crawl that is followed by a thorough analysis of the crawled content.

---

[2] Open Ministry, Finland, http://avoinministerio.fi

Using the ARCOMEM system, the political researcher may set up a customized job for online collection of resources and subsequent offline analysis of the collected data. An extensive set of parameters, generic and data specific may be used:

Generic input:

- Date/time, start and end of the crawl, frequency (schedule)
- Crawler selection
- Crawler specification, initial seed lists

Semantic level data input:

- Targeted entities (people, events, locations, geo-location, other)
- Targeted social networks
- Specific URLs and/or users from social media, languages

The specification above yields a large and diverse set of collected web pages, blogs, wikis, social media posts and pages. Consequently, all the data that are collected are fed to the analysis modules that perform *entity recognition*, *event extraction*, *topic recognition*, *sentiment analysis*, *social graph mining*, *social text diversification* and other analyses.

4. Social feedback and content collection

The ARCOMEM system analysis, in the end, provides a semantic database that contains both the raw data that have been collected and the analyzed results in the form of RDF triples[3]. Using semantic queries and a specialized search and retrieval interface (Spiliotopoulos et al., 2012), the user may search for specific keywords (Figure 1).
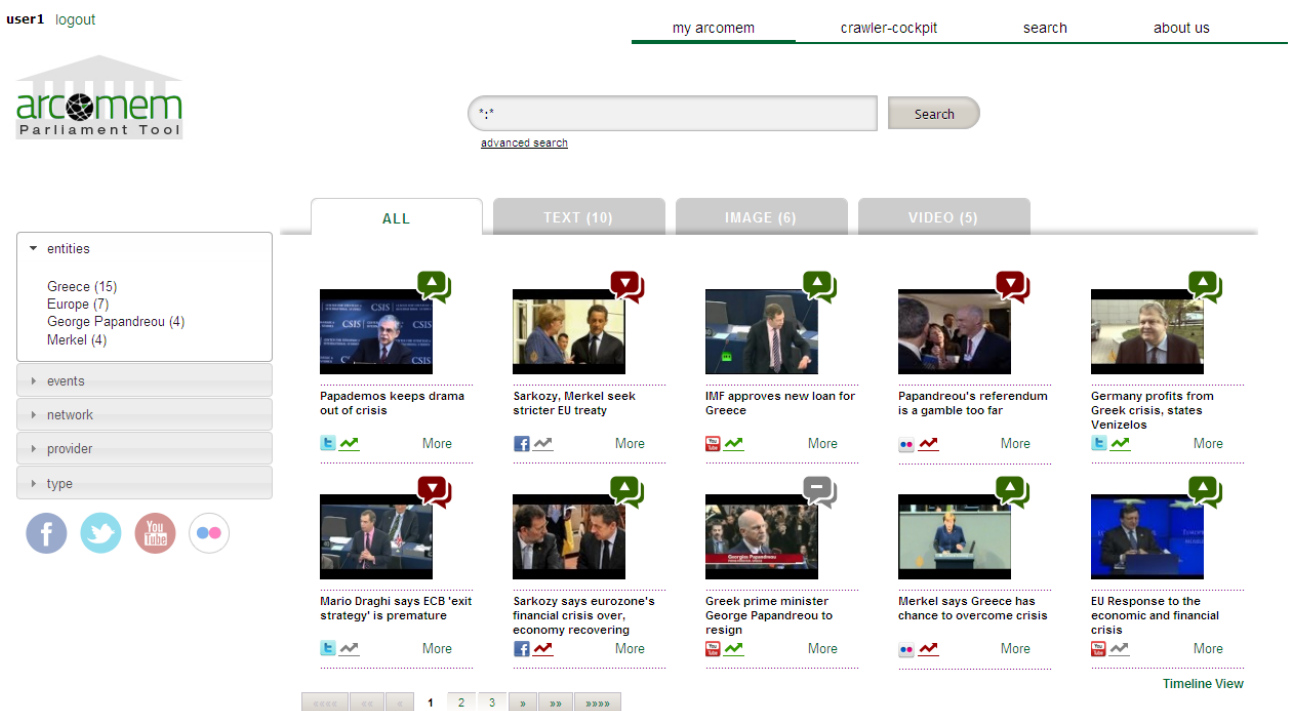

Figure 1: Searching the web archive

---

[3] http://en.wikipedia.org/wiki/Web_Ontology_Language

The keywords may be either specific entities (such as politicians or ministers that are involved in the draft of a legislation) or free text that describes the domain (such as copyright law, digital media, movies, music, Europe, *country_name*). The results contain all web resources where the search parameters were retrieved for as well as the important semantic information. This information is:

- Entities and lists of related entities to the search

- Events and locations that were identified as relevant to the aforementioned entities

- Topics of discussions

- Collective opinions from social network users relevant to the search parameters

- Statistically significant data from the social web analysis, such as opinions of Twitter users on a person over a period of time

- Diverse texts from social networks

- Major users in terms of involvement as well as acceptance from the other users in the social networks and across social networks

All the above constitute a large amount of already analyzed and aggregated data that the user may access, for instance, the type of groups the people refer to (age groups, financial groups, etc.), what they discuss about, opinions on certain people or events, options to view the most diverse comments from the social media in terms of opinion or other parameters, user involvement statistics relevant to the user profiles (age, location, etc.). An expert political analyst, an educated citizen, a news broadcaster, a government consultant may use this information to understand the potential implications that certain aspects of a proposed law would have for the people, the positive and negative views, the important or significant issues that the people are concerned with.

5. Conclusion

Social web is a pool of data that can be used to identify the societies' pulse over the political domain. Legislation issues are accessible on the web and the response of the affected citizens is on the social web. Feedback from the people may be interpreted by using special tools to collect and analyze the citizen input itself as well as statistical information that can be calculated based on social network metrics, user profiling and, most importantly, opinion mining. Such feedback, would be or paramount value both to law makers and to societies and a rich well of common knowledge for the democratic dialogue that should accompany legislation drafting.

6. Acknowledgements

7. References

Schefbeck, G., Spiliotopoulos, D., Risse, T. (2012): The Recent Challenge in Web Archiving: Archiving the Social Web, International Council on Archives Congress, Brisbane, Australia, 20-24 August 2012.

Risse, T., Peters, W. (2012): ARCOMEM - From Collect-All ARchives to COmmunity MEMories.. In: World Wide Web Conference 2012 (WWW2012) - European Project Track, Lyon, France, April 2012

Spiliotopoulos, D., Tzoannos, E., Stavropoulou, P., Kouroupetroglou, D., Pino, A. (2012): Designing user interfaces for social media driven digital preservation and information retrieval, Lecture Notes in Computer Science, Volume 7382/2012, Computers Helping People with Special Needs, Springer-Verlag Berlin Heidelberg, pp. 581-584, 2012.