Utilising Linked Social Media Data for Tracking Public Policy and Services

Deirdre Lee, Adegboyega Ojo, Mohammad Waqar,

Digital Enterprise Research Institute (DERI), NUI Galway, Ireland
{firstname.lastname}@deri.org

Abstract. What are the concerns of citizens in relation to the new proposal for a ring-road? What are their views on means-testing for medical cards? Are self-employed people finding the online tax system helpful? Traditionally, such question would be answered through commissioned surveys, targeted consultations or journalistic research. Today, citizens are expressing views on topical issues such as public policy and services voluntarily on social media sites. Facebook, Twitter, LinkedIn, etc. contain views and arguments on a wide range of issues, which can be useful for informing policy-makers on public opinion. However, public policy isn't the only topic discussed by citizens online; how can policy-makers distinguish relevant data from opinions on who should win X-Factor or what is the best sandwich filling? In this paper, we propose a solution for systematically tracking particular topics of interest across a range of social media sites. We also discuss how this solution may be employed for tracking topics related to specific public policy or service of interest.

Keywords: Social media, Linked Data, public policy

1 Introduction

The proliferation of social media sites on the Web brings with it exciting opportunities for everyone to share ideas and broadcast opinions to the world. Facebook is a social networking service with over 1 billion monthly users[1]. Twitter is a microblogging service with over 500 million users[2]. Foursquare is a location-based social networking service with over 25 million users[3]. With continuous posting, tweeting and checking-in on these, and many other, social media sites, monumental

[1] http://news.cnet.com/8301-1023_3-57566550-93/facebook-by-the-numbers-1.06-billion-monthly-active-users/

[2] http://techcrunch.com/2012/07/31/twitter-may-have-500m-users-but-only-170m-are-active-75-on-twitters-own-clients/

[3] http://gigaom.com/2012/10/01/how-foursquare-is-building-a-revenue-strategy-around-local-search/?utm_source=feedburner&utm_medium=feed&utm_campaign=Feed%3A+OmMalik+%28GigaOM%3A+Tech%29

amounts of information are being produced by users from around the world, in many different languages.

Governments are aware of the potential of social media, but up to now have focused on using social media as a means to communicate with citizens. Many governments from United States thru United Arab Emirates to Australia; have developed policies enabling public agencies to use social media to promote more flexible and near real-time government-citizen interaction. Governments set up official profiles mainly to disseminate information about news and events. Many politicians also now have personal, official accounts, which they use for keeping their constituents up to date on some of their own movements; what events they are attending, whom they are meeting and latest news updates. Politicians also sometimes engage with citizens directly through their social media accounts.

However, while these official accounts may help in the real-time dissemination of news and the building of a more transparent government, they don't harness the full potential of social-media for public authorities. Citizens are continuously producing content, some of which may be related to public policy and services. But if there is no direct link to a government account, or even the use of a common hashtag, much of this information will probably never be discovered by public authorities. In addition, there are also many other technical challenges that impedes the discovery of relevant data from social media content, such as:

- A wide variety of data sources
- Each social media site has a different API to access the data (if there is an API at all)
- Each API has unique constraints in relation to how much data can be accessed, how often, etc.
- The APIs return data in a variety of formats, e.g. JSON, XML, etc.
- There are, rightfully, many privacy limitations on social media data
- Social media data is noisy, informal, slang-filled, and thus, difficult to process

Corporations are aware of the value of listening to their customers and invest heavily in online brand-monitoring and reputation analysis. They are then able to respond appropriately with modified products or special promotions. In an era, where a huge emphasis is being placed on Open Government, public servants that listen and respond to citizens is equally, if not more, important, to businesses listening to their customers. However, with limited resources and specialised expertise, governments have limited capacity to effectively harness the collective intelligence of citizens embodied in social media contents for civic purposes, for example, to guide the decision-making process, to sense and track public sentiments on public policies, to monitor the quality of public services, or to explore different views regarding public policies, procedures and services.

In this paper, we will present a Linked Data solution that will enable governments to continually monitor specific policies or services of interest, gauge citizen-opinion and react accordingly.

2 Social Media Data as Linked Data

Linked Data refers to data published on the Web in such a way that (i) it is machine-readable, (ii) its meaning is explicitly defined, (iii) it is linked to other external data sets, and (iv) can in turn be linked to from external data sets (Bizer 2009). Linked Data requires the identification of entities with URI references that can be dereferenced over the HTTP protocol into a common semantic data model, such as RDF. The use of Linked Data is advantageous as it inherently facilitates interoperability and integration of data from disparate data sources. It also opens up possibilities for carrying out inference and analytics over the data.

In order to extract meaningful information from the abundance of social media data, Linked Data can be used, modelling the data in a homogenous format and identifying relationships within the content. Once represented as Linked Data, the content can also be linked to the Linked Open Data (LOD) cloud, which represents all public datasets that have been published in Linked Data format, by contributors to the Linking Open Data community project and other individuals and organizations, based on metadata collected and curated by contributors to the Data Hub[4]. The LOD Cloud currently contains 295 datasets, with over 31 billion RDF triples, from a variety of domains, including media, geographic, government, publications, and life sciences. Using Linked Data to facilitate integration of social media data is currently being adopted in various use-cases, for example, to combine social media data with sensor streams[5], or to combine social media data with Open Data (Kalampokis et al. 2012).

We propose to employ Semantic Web and Linked Data technologies to achieve information integration within the government enterprise, as well as analysis over social media content.

3 Proposed Solution: Social Media Linked Data Space (SMLDS)

The Social Media Linked Data Space (SMLDS) provides a customised topic monitoring service over a number of social media data sources. It accepts queries, based on a number of predefined parameters, such as keywords, location, and time-range, and processes these queries in real-time over a number of disparate social media sites. The resultant data is collated, enhanced using Natural Language Processing (NLP) techniques, such as Named Entity Recognition (NER) and interlinked with the Web of Data. The resultant data can be accessed in a homogenous manner by the user, downloaded, viewed and analysed.

The architecture of the SMLDS is shown in Figure 1. A user creates a new query to monitor activity around a certain topic. The SMLDS crawls data from a number of different social media and review sites, curretnly Facebook, Twitter, Qype, Bestbuy and Foursquare. (This list is extensible to include any data sources that expose their data via an API.) Relevant data is harvested by the SMLDS and captured as Linked Data in a triple store in accordance with a predefined Semantic Model. This data can

[4] http://lod-cloud.net/
[5] http://superstreamcollider.org/

then be enriched using NER, opinion extraction, sentiment analysis, data analysis, etc. The design of the SMLDS was guided by the following requirements:

- Simple and easy to scale.
- Components can be distributed.
- More components can easily be added to the system without disrupting the flow of the system.
- Use standard protocols HTTP and SPARQL for communication.
- Platform and language independent

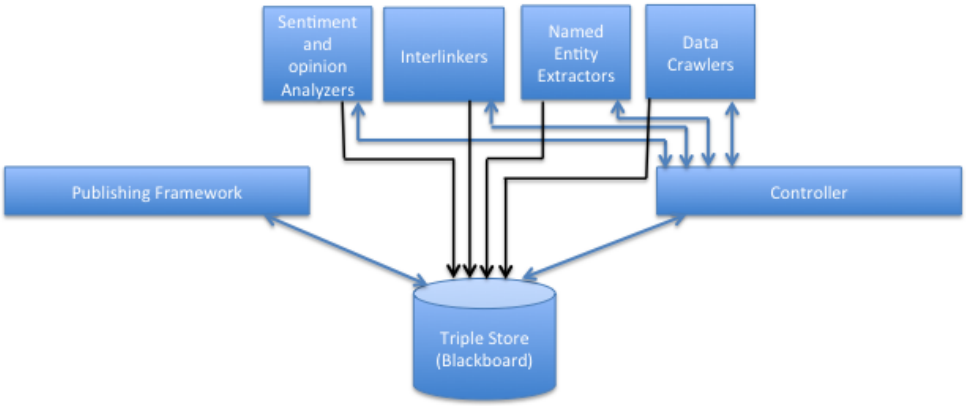

Figure 1: Social Media Linked Data Space (SMLDS) Architecture

4 Application of SMLDS

Rich services can be built around SMLDS to improve decision making and service delivery, including the tracking of public polices and specific services differentiated by geographies and demography, identifying ideological biases in public policy discourse, sensing public sentiments on on-going government programs, detecting odd events for instance resulting from government-citizen communication gap, and monitoring quality of services based on opinions.

5 Conclusion

The challenge of how to extract topics of interest amongst the non-relevant contents on social media sites have prevented governments from fully harnessing the opportunities provided social media. The Social Media Linked Data Space proposed in this paper offers a concrete solution to this problem. With the first release of SMLDS under testing, our next goal is to develop specialised Named Entity Extractors for a pilot public policy or service domain using DBpedia Spotlight[6].

6 Acknowledgements

The work presented in this paper has been funded in part by the European Union under Grant No. 286714 (Linked2Media – Research for SME Associations, SME-2011-2) and by Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2).

References

Bizer, C., 2009. The Emerging Web of Linked Data. *Intelligent Systems*, 24(5), pp.87–92.

Kalampokis, E. et al., 2012. Augmenting Open Government Data with Social Media Data 2 Open Government and Social Media Data. In *W3C Using Open Data (PMOD 2012)*.

[6] http://www.wiwiss.fu-berlin.de/en/institute/pwo/bizer/research/publications/Mendes-Jakob-GarciaSilva-Bizer-DBpediaSpotlight-ISEM2011.pdf