iParticipate: Automatic Tweet Generation from Local Government Data

Christoph Lofi[1], Ralf Krestel[2]

[1] IFIS, Carolo-Wilhelmina Universität Braunschweig, Germany
lofi@ifis.cs.tu-bs.de
[2] L3S Research Center, Leibniz Universität Hannover, Germany
krestel@L3S.de

Abstract. With the recent rise of Open Government Data, innovative technologies are required to leverage this new wealth of information. Therefore, we present a system combining several information processing techniques with micro-blogging services to demonstrate how this data can be put to use in order to increase transparency in political processes, and encourage internet users to participate in local politics. Our system uses publicly available documents from city councils which are processed automtically to generate highly informative tweets.

1 Introduction

More and more public institutions and governments provide access to official documents via the Internet. This fosters *transparency* [2] and the possibility for common citizens to participate more actively in political and governmental processes. But the huge amount of information available and the diversity of documents (meeting minutes, requests, petitions, proposals, ...) makes it difficult for non-experts to detect interesting information snippets. Especially the younger generations want to stay informed but are not willing to spent much time digging for related information.

With e-Democracy and e-Government on the rise [6], the use of information technologies for governmental processes becomes a regular means of communication between the government and its citizens. Leveraging all the public information available today could not only yield *social value* but also *commercial value*. Important proposals and eventually decisions are made public by officials but are not easy to find by a broader audience. In particular local politics are not covered by mass media and thus remain obscure and non-transparent despite the goal of Open Government Data.

Besides the informational aspect, open public data can also encourage active *participation* in political discussions and decisions if combined with modern information systems techniques. Politicians can get quick feedback and capture the overall mood towards a question or argument. Communication technology nowadays offers various ways of interaction and participation [3]. Web 2.0 applications enable everyone to share their thoughts, e.g. via blogs, in social networks such as Facebook, or using microblogging (e.g. Twitter).

In this paper we present an innovative approach to combine Web 2.0 applications and Open Government Data, directly delivering relevant information to support citizens. We connect and link data offering a space for discussion and sharing of ideas. In detail, we present a system which automatically generates tweets in Twitter based on open local government data.

2 Automatic Tweet Generator

In our demonstrator, we use publicly available data from the city of Hannover, Germany[3]. This data contains various types of documents, such as meeting minutes, proposals, and petitions. The documents are part of official administration processes. The city council, for example, governs the citywide affairs, while for each city district an additional district council manages local issues, such as traffic planning, school and kindergarten organization, maintenance and reorganization of parks and recreational areas, and local town planning. Furthermore, districts may suggest measures to the city council. Each of the district councils meets monthly. For each meeting, members of the council, but also every citizen, may submit petitions to be added to the agenda of the upcoming meeting. These petitions are published online as individual documents as soon as they are received. Briefly before the meeting, the full agenda is published. After the meeting, all decisions are documented in meeting minutes, while the initial petitions are updated to include the passed resolutions.

Currently, although all documents are openly published, the whole process is quite opaque for the average citizen: documents are hard to find and usually only identified by a numeric id. The missing link structure makes it difficult to follow a certain process or stay informed about a particular issue. Here, our prototype systems steps in to increase transparency and convenience of Open Governmental Data. In particular, the following issues are addressed:

- Obtaining relevant documents from local open-data enabled governmental websites and archives
- Automatic topic classification of documents for better accessibility
- Extraction of hash-tags for intuitive navigation between interrelated tweets
- Automated Twitter tweet generation for always up-to-date push-style notifications of notable local political developments

2.1 System Architecture

Figure 1 gives an overview of our system. The first step in generating tweets is to poll the city archives for new documents and store them in an internal database. Each document is then analyzed, classified, and linked to previously obtained documents. Then, the expected "interestingness" of each document is established based on its type, content, and relation to other documents. For each city district,

[3] https://e-government.hannover-stadt.de/lhhsimwebre.nsf (german). For demonstration purposes, we will use documents machine-translated to English.

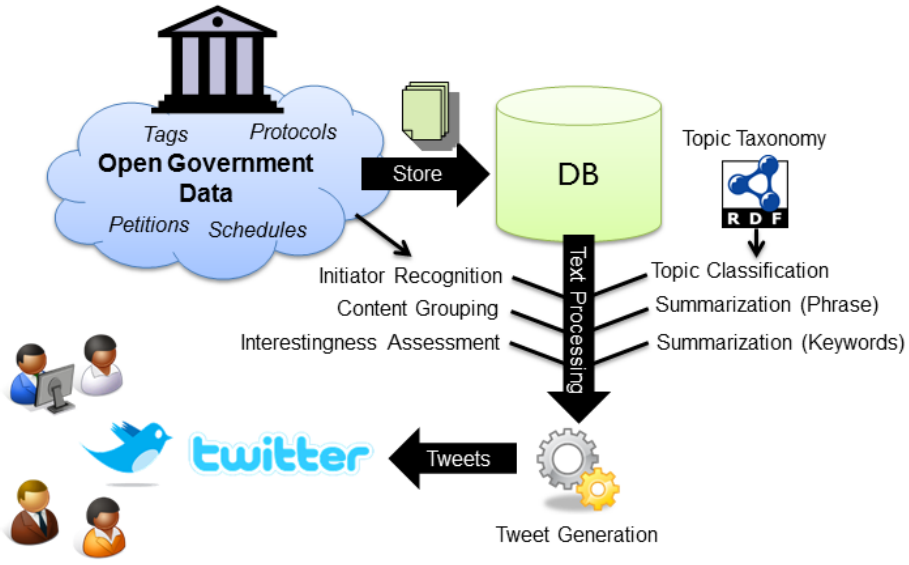

Fig. 1. System Architecture Overview

a twitter account is available allowing users to follow the local politics in this district. Documents which are considered interesting by the system are then transformed into a Twitter tweet, and published via the respective district's account. Each tweet also contains a link to the full document. This task raises two major challenges: a) selection of suitable hash-tags for finding, grouping, and organizing tweets and b) summarization of the document respecting Twitter's highly limited text length (140 characters).

Hash-tags are generated by several different means: Topic hash-tags (like 'schools', 'traffic', 'parks', etc.) briefly summarize the document's content. This classification with respect to a pre-defined topic taxonomy can be achieved by using a machine learning algorithm like support vector machines (SVM) [4]. By training the SVM with a small sample set, all following documents can easily be classified with high accuracy. Hash-tags can then be used to group documents together belonging to a single process or issue. For example, all petitions, agendas, and protocols of one particular district meeting use the same, unique hash-tag, thus allowing for easy navigation between related tweets. Source tags encode the documents political source, i.e. usually containing the shortcut for the responsible political party or spokesperson. These tags help to better explore the political activity within a city. City tags group all tweets of a city's districts together for broader overviews. Finally, different techniques for automatic text summarization [5] and key phrase extraction [1] can be used to summarize and tag the content of a document.

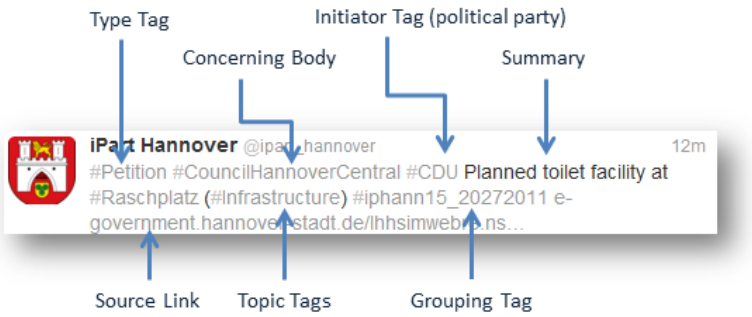

Fig. 2. Example Tweet

An example tweet illustrating the functionality of our automatic tweet generator is shown in Figure 2. Hannover uses its Twitter account (iPart Hannover) to publish news, in this particular case about a petition to build new toilets in the city center.

3 Summary & Outlook

In this paper we presented a system to automatically generate tweets from Open Government Data. Our system can help citizens to stay informed about local affairs. Further, it allows internet users to discuss, follow, and get involved in local politics adding transparency and encouraging participation. For future work, we plan a long-term user study to evaluate the system and estimate the possible impact of new communication technologies like Twitter on local politics.

References

1. Frank, E., Paynter, G.W., Witten, I.H., Gutwin, C., Nevill-Manning, C.G.: Domain-specific keyphrase extraction. In: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence. pp. 668–673. IJCAI '99, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1999)
2. George, J.M.: Can it help government to restore public trust?: Declining public trust and potential prospects of it in the public sector. In: Sprague, R. (ed.) Proceedings of the 36th Hawaii International Conference on Systems Sciences (HICSS'03). IEEE Computer Society (2003)
3. Gibson, A.: Local by Social: How Local Authorities Can Use Social Media to Achieve More for Less. NESTA (2010)
4. Joachims, T.: Learning to Classify Text Using Support Vector Machines – Methods, Theory, and Algorithms. Kluwer/Springer (2002)
5. Mani, I.: Advances in Automatic Text Summarization. MIT Press, Cambridge, MA, USA (1999)
6. Parycek, P., Kripp, M.J., Edelmann, N. (eds.): CeDEM11 Proceedings of the International Conference for E-Democracy and Open Government. Edition Donau-Universität Krems (May 2011)