HAL
open science

Online Forums vs. Social Networks: Two Case Studies to Support eGovernment with Topic Opinion Analysis

Timo Wandhöfer, Beccy Allen, Steve Taylor, Paul Walland, Sergej Sizov

Online Forums vs. Social Networks: Two Case Studies to support eGovernment with Topic Opinion Analysis

Timo Wandhöfer[1], Beccy Allen[2], Steve Taylor[3], Paul Walland[3] and Sergej Sizov[4]

[1] GESIS – Leibniz Institute for the Social Sciences, Cologne, Germany
{timo.wandhoefer}@gesis.org
[2] Hansard Society, London, UK
{beccy}@hansardsociety.org.uk
[3] University Southampton, IT-Innovation Centre, Southampton, UK
{sjt,pww}@it-innovation.soton.ac.uk
[4] WeST – Institute for Web Science and Technologies, University of Koblenz, Germany
{sizov}@uni-koblenz.de

Abstract. This paper suggests how eGovernment and public services can apply "topic-opinion" analysis (developed in the EC IST FP7 WeGov project) on citizens' opinions on the Internet. In many cases, discussion tracks on the Internet become quite long and complex. Stakeholders are often interested in gaining a quick overview of such a discussion, including understanding its thematic aspects, identifying key arguments and key users. The topic opinion analysis that is part of the WeGov toolbox aims to provide appropriate summarization techniques by identifying latent themes of discussion (topics), most relevant contributions and arguments for each topic, as well as identifying the most active users that influenced a certain aspect of discussion. In this paper we focus on online forums and social networks as digital places where users discuss potential political issues. Therefore we setup two different case studies to validate the accuracy and usefulness of analysis results of the topic opinion analysis.

Keywords: eGovernment, Twitter, Facebook, Online Forums, Topic Opinion Analysis, Sentiment, Controversy

1   Introduction

Governments and public institutions are increasingly working with citizens to give them more of a stake in the policy-shaping process, for example through public consultations on new legislation [5]. E-participation platforms foster communication and interaction between politicians and government bodies on the one side, and citizens on the other [5]. Notwithstanding the benefits brought about by existing eParticipation platforms, there remains the unsolved challenge of how to involve a larger number of affected individuals, groups and communities in discussions than is currently achieved through dedicated web sites. This problem has, for example, been analyzed with the 10 Downing Street Debate Mapper, being a case in point [6]. They found that

very few people (7% of specifically addressed 309 invitees) took part using the dedicated Debate Mapper website, but many of them did comment about the same subject on other Web platforms.

The use of social networking platforms has a significant part to play in political engagement. Social Media [3] and blogs [2] have high potential for the eGovernment to interact with citizens. From a sociological point of view, platforms like Twitter are interesting for analyzing the dissemination of topics and as well for analyzing the opinions and sentiments of the society regarding particular topics [7]. Beyond that, online platforms have the power to influence the process of opinion making [2]. That's why [1] uses the label: "*new politics of listening*". For instance in the UK the Internet and social networks are everyday life functionality for Parliamentarians [6].

There is thus a huge potential of online discussion places, but there is a problem of making sense of the huge amounts of text in them. The aim of this paper is to suggest topic opinion analysis while validating to cases.to support eGovernment by exploiting the potential of online discussions and addressing the problem of "too much information". In the next section, we introduce the WeGov toolbox resulting from the EC IST FP7 WeGov project, and especially its topic opinion analysis component that provides summarization for political decision-makers. Subsequently we explain the process model behind this paper and describe two case studies that were selected to evaluate the topic opinion analysis with real data. Finally, we draw general conclusions.

2 Background

Both case studies that are described within this paper are based on the WeGov Toolbox as the technical framework. The WeGov toolbox supports diverse components for analyzing huge amount of online available text for stakeholders[1].

2.1 WeGov Toolbox

The WeGov Toolbox (hereafter "the toolbox") is a web-based system that enables the user to collect and analyze postings[2] and users from social networks and the HeadsUp forum. The toolbox is deployed and hosted at a server, and the user connects to this using their web browser. The key features of the toolbox are as follows.

- The user can specify and run searches on the social networks Facebook and Twitter, and raw data reflecting users' comments is collected.

---

[1] Here we are referring to the definition from [17]: "*Any group or individual who can affect or is affected by the achievements by the organization's objectives*". Beyond that stakeholders within the context of this paper are potential end users of the WeGov toolbox within the field of politics and public administration.

[2] A posting (abbreviated post) is a digital user contribution within online forums, blogs or social networks. Generally a post is a text message. Here a post subsumes seed posts, status updates and comments.

- — On Facebook, the user can monitor public groups and pages - the user can instruct the toolbox to collect posts and comments on those posts from a Facebook group or page by specifying the URL of the page.
  - — On Twitter, the user can search for keywords or hashtags.
- Searches can be scheduled, so that they repeat automatically. This is useful for collecting data over an extended period, which is particularly suitable for monitoring a news story. The system is designed so that when a search is executed multiple times by a schedule, it will not collect any duplicate posts, as duplicates can skew analysis results.
- Search results can be fed into the toolbox's two analysis components to provide summaries and automated insights into the (sometimes very large) data set returned from the social networks.
  - — Behavior analysis has been developed by the Open University, Knowledge Media Institute (KMi), and monitors the discussion activity, categorizes users into behavior types and highlights key posts and users [11,12,13].
  - — Topic-opinion analysis has been developed by the University of Koblenz [8], and determines themes of the documents (posts, comments, etc.) in the discussion by identifying sets of terms that frequently occur together in multiple posts and grouping them together into topic groups. In addition, opinions are determined by sentiment analysis, and the topic groups can be measured in terms of whether they express positive or negative opinion.
- We have adopted a methodical approach for the development process of the toolbox with frequent and iterative end user engagement, such as the German Parliament, the German State Parliament of North Rhine-Westphalia, the EC Parliament, city administrations, parties and NGOs [9] so as to get requirements and feedback on the toolbox's functions and usability [4]. As part of user engagement, a number of use cases were designed [16] showing how the toolbox analysis tools could provide a two-way dialogue with citizens, and the work reported here develops one of these use cases.
- An important aspect of the work in the WeGov project is to protect the rights and privacy of citizens and policy makers. To address this, a legal and ethical analysis was conducted to provide us with an understanding of data protection issues and give an insight into transparency. This work has influenced the design and use of all parts of the toolbox, and has been reported elsewhere [10]. The impact it has on the work here is that we only collect postings from publicly accessible sources.

2.2 Topic Opinion Analysis

In many cases, discussion tracks in social media become long and complex. Stakeholders of the toolbox technology (such as politicians, political researchers, active users) are often interested in gaining a quick overview of such a discussion, including understanding its thematic aspects, identifying key pro- and contra- arguments and finding the most influential users. However, completely reading hundreds (or even thousands) of posts is too time-consuming to be practical. There is thus a huge need to summarize the discussion tracks, and the Topic-Opinion analysis component of the

toolbox provides this by identifying latent themes of discussion (topics), most relevant contributions and arguments for each topic, as well as identifying the most active users that influenced a certain aspect of discussion. [8]. The topic-opinion tool employs state of the art methods of Bayesian learning and opinion mining [14,15] for finding the most relevant pieces of information that should be presented to the user, and the methods are briefly described next.

Modeling topics: Probabilistic Bayesian models are used for mining the latent semantic structure of the online discussion. The toolbox approach can be seen as an extension to the state-of-the-art method named "Latent Dirichlet Allocation" (LDA) [14]. The collection of postings is represented by means of probabilistic distributions over terms (words) that appear in particular discussion postings with different frequencies. The analysis runs across many posts, and looks for words that occur together in the same post. Topics are formed from groups of words that frequently occur together in a post, and the more posts that contain the same group of words, the stronger the topic is. Each topic is therefore characterized by its most relevant terms. A post can be in more than one topic (for example if it contains words commonly occurring in two topic groups), and consequently, postings are represented by means of probalistic membership of topics (e.g. a post can be 50% in topic 1, 25% in topic 2 and 25% in topic 3). Postings that belong to a certain topic with high probability are considered as most characteristic examples for the certain aspect of online discussion. [8]

Modeling opinions: The toolbox employs state of the art techniques for mining user opinions and affect states. Conceptually, they are based on structured vocabularies that indicate the emotional state of a post's author (e.g. skepticism, positive or negative emotions, anger, etc.). Consequently, postings with strong opinions or emotions are selected for presentation to the user. [8]

Topic-opinion summarization: Results of topic and opinion analysis are combined for presentation to the user. First, candidate postings are chosen with respect to their high relevance regarding particular discussion aspects (i.e. topics). Second, for each pre-selected posting, opinion/emotion analysis is performed. The output is constructed in such a way that a) all topics identified in the dataset are appropriately reflected, and b) postings chosen for each topic reflect different opinions and emotions. As a result, the output contains a limited number of "must-see-first" contributions from the online discussions, covering a broad spectrum of its contextual and emotional facets. [8]

Topic-opinion analysis is intended to provide quick summaries of the themes in a debate and the opinions expressed by the citizens on digital places. As an example of this, Figure 1 shows the topic analysis results when the input was multiple sets of responses on Twitter to the query 'cyprus'.
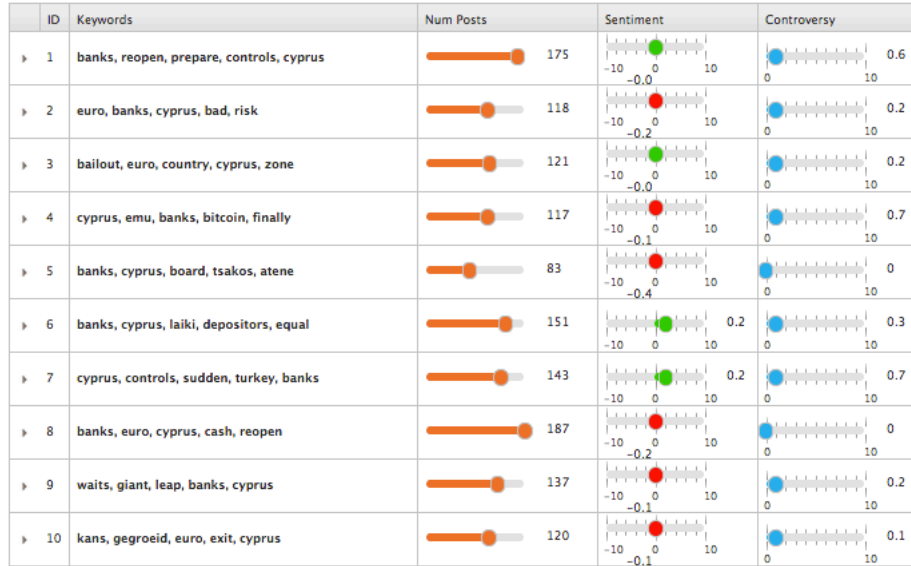
Fig. 1. Topic opinion analysis

Each line includes a list of five keywords that build the topic (e.g. "banks", "reopen", "prepare", "controls", "cyprus"). The next column shows the number of tweets that are sorted to each topic (e.g. 375 tweets for the first topic). The last two columns show the sentiment and controversy of tweets that are measured for each of the twelve topics. The indication of sentiment shows if the tweets that are related to one topic are rather positive, neutral or negative. The indication of controversy shows the ratio of positive and negative posts.

3   Applied Process Model

Figure 2 shows the applied process model how stakeholders were engaged, both to determine requirements and to evaluate the toolbox. The idea behind this approach is to identify potential use cases that are in the end users' daily working lives. These cases are therefore of value to the end user, and can be used for validating the toolbox and its analysis results. Figure 1 shows two examples of such use cases ("HeadsUp" and "social networks"), and these are discussed throughout the paper to illustrate how the topic opinion analysis can be applied in everyday politics.

The top row in Figure 2 shows users on the Internet (the digital society) – for instance users of online forums or social networks. The second row shows stakeholders, and how they interact with the users on the Internet. In the use cases, the stakeholders already perform (often manual) analyses on the data they get from citizens on social networks. The results of their existing analyses are shown in the bottom row – here we call these data the "control group". The control group is compared with the

toolbox's analyses of the same data. For instance the operator of the HeadsUp[3] discussion forum (cp. left) analyses the forum discussions manually to get an insight on the debate. Another example is the policy maker (cp. right) who extracts topics from social networks to get insight into the discussion.
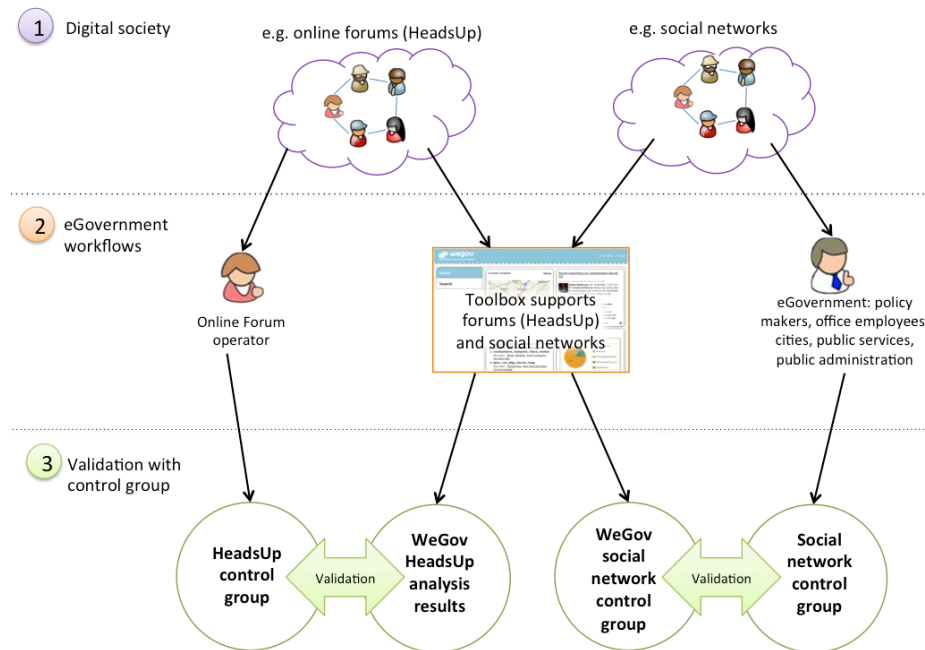

Fig. 2. Applied Evaluation Model

4   HeadsUp Case Study – Online Discussion Forums

Civil society groups run forums and blogs to connect with their members and supporters, but often analyzing the themes of these discussions is often beyond the organizations' resources. The toolbox could play an important role in helping small not-for-profit organizations, larger media organizations, as well as politicians and policy-makers to understand feedback across a range of communication channels.

HeadsUp is a UK initiative, launched in June 2003 to promote political awareness and participation amongst young people. It is an online debating space for 11-18 year olds that gives them the opportunity to debate political issues with their peers, elected representatives and other decision-makers.

Five debates happen each year, each lasting three weeks and are fitted around both the school and parliamentary calendar. The forum discussions are based around political topics of interest to young people, as well as those related to key political events,

---

3   URL: http://www.headsup.org.uk/content/ (Retrieved 13/3/13).

issues of debate in Parliament and the media, and current government policy. Each forum is supported by background materials and teaching resources to ensure that the discussions are of a high quality.

The discussions are analyzed by the Hansard Society and are summarized in a report, which is disseminated widely. The report contains the key themes of the debate with direct quotes from participants, other information about the forum and the political context at the time the debate happened.

The core reason for analyzing the forums and distributing the report is to allow young people to have their voices heard by those that make decisions on their behalf, and to highlight that their perspectives are often different to those of adults. This is a vital aspect of HeadsUp: the report provides a channel to feed back information from the forums to policy-makers, politicians and journalists; thereby allowing young people's perspectives to inform a wide audience of those with the power to effect change.

4.1 Methodology

HeadsUp was used to evaluate the usefulness of the toolbox toolkit with regard to forum data and it provided a case study using a real world data set. It was a useful test case because the reports of each forum were generated before the WeGov project was started, so they were a good, independent control group for comparison with the results emerging from the toolkit. Three forums of different sizes were used:

- Sex Education – Do you get enough? (36 posts)
- Youth Citizenship Commission: are young people allergic to politics? (317 posts)
- How equal is Britain? (1186 posts)

The output of the toolbox was compared to the forum reports to assess how well it determined the **themes of the debate** as analyzed by a human. The toolbox's assessment of **sentiment** was compared with the reports and a selection of posts was double-checked for accuracy by a human. The **user interface** and the options available to view the data were assessed for their usefulness when populated with forum data.

4.2 Findings

The toolbox has applications outside social networks. Comments from blogs and forums or other data sets could be analyzed using the toolbox. This could help both, small non-profits, or large media organizations to analyze large-scale interactions.

- The toolbox is best at dealing with large quantities of data, amounts that could not be analyzed effectively by a human without significant resources to do so.
- The toolkit performs well on relatively in-depth data - this lends itself to blogs and forums that encourage more considered and less immediate responses.
- The toolbox also performed well in showing the nuances between different elements of a wider debate e.g. the women's sport debate (sexism in sport & the types of sports played by men and women vs. mixed sports).

- The toolkit works best when analyzing medium length comments that focus on one issue and when spelling and vocabulary are good.

4.3   Improvements

- A plain English explanation of how the algorithm understands and processes data is very important to ensure users trust the results. An explanation of irregularities such as:
  - Why the same data sometimes yields different results?
  - Why keywords appear in the order and frequency that they do?
- Showing the hidden workings of the toolkit such as:
  - The relative influence of a greater number of key words e.g. via a tag cloud.
  - Highlighting positive/negative words that contribute to sentiment scores.
  - A separate group for excluded comments so they are still visible to the user.
- Implementing more options for the users to refine the data and customize it to their situation and needs. For example:
  - The ability to exclude certain posts or words from analysis.
  - Splitting up long posts into sections that can be analyzed separately to avoid the conflicting analysis of longer posts.

Although elements of the toolkit interface and the algorithm could be improved to help users understand the results better, the toolbox worked very well with the longer more in depth posts that are more common to blogs and forums than to social media.

5   Facebook and Twitter Case Study - Social Networks

The intention of this case study was the validation of usefulness of topic opinion analysis of social media for politics. Therefore we designed three use cases how this technology may support politicians' everyday life:

1. **Local Facebook topics**: Within this use case, we monitored a sample of at least ten Facebook pages represent a geographical areas like an MP's constituency. Here topic opinion analysis was applied to extract the topics that people discussed on the pages. Each topic is a combination of words that represents a theme of the discussion, and comes with key users, and key comments.
2. **Monitoring topics on Twitter**: The intention of the second use case is to identify subtopics on Twitter. For instance, the general debate on climate change covers subtopics like green energy, new kinds of technologies. Here we collected data three times a day from Twitter by searching for e.g. "climate change", and used topic analysis to detect the topics of the results. Because the results are already filtered by the search, the analysis produced subtopics.

This case study was conducted with a number of governmental representatives as end-user stakeholders: two members of the German Bundestag, four employees that work directly for a member of the German Bundestag, two members of the State Par-

liament North Rhine-Westphalia[4], one small German city (Kempten), one big German city (Cologne), and with a German state chancellery (Saarland). In total this evaluation consisted of 11 questionnaires and 12 semi-structured interviews following the questionnaires. The questionnaires and interviews were based on an individual analysis report that was created from four weeks of data collected from Facebook and Twitter using the toolbox search tools and scheduler.

5.1 Methodology

To address the aims above, we configured the toolbox to collect data relevant to our proposed interviewees – we created user accounts for them, and set up automatic scheduled searches that were relevant to them. This enabled us to demonstrate and evaluate the analysis components with the external users that would contain subject matter they were interested in. Our reasoning behind this was that if they were interested, they would be better engaged, and therefore the quality of feedback would be better than if we had used arbitrary searches. Feedback from previous meetings with end user stakeholders showed us that local or constituency-based searches were of high importance to them, so these were strongly featured in the searches we set up.

- **Strategy**: Our strategy was the preparation of an individual analysis report. Each report was created using the same structure, but with data targeted to the end-user stakeholder it was intended for, and included approx. ten Facebook pages, related to the local area. For Twitter we used approx. five keyword searches using phrases around the end-user stakeholder's areas of interest. The unique data profile was initially created by the WeGov project team and was updated over several iterations by the feedback end users provided concerning their profile. For the collection from Facebook pages, we used the Facebook search tool, where we queried the constituency and the names of cities and towns within the constituency. The pages represent a selection of the available pages related to or managed by cities, public institutions, associations, local associations, arts and culture, politics, tourism and the local press. Pages with more likes, posts and comments were selected before those that displayed less public engagement. If the MP had "liked" one of the selected pages this information was noted.
- **Analysis Report**: After four weeks of data collection, the data was analyzed by the toolbox following the "Social Network" use case pattern described above, and the results were collated into reports. The reports included a description of the evaluation strategy and the results at a glance, on one page where possible, and were sent to end users approx. two weeks before the interviews to allow time for them to prepare their comments and feedback.
- **Questionnaire**: In addition to the analysis report, the participants got a questionnaire that covered concrete examples from the report. All questionnaires contained the following information and included the same questions. The only difference

---

[4] URL: http://en.wikipedia.org/wiki/Landtag_of_North_Rhine-Westphalia (Retrieved 9/3/13).

was the sample of analysis results that were tailored to the target end user, and contained questions similar to the following examples: Is the topic clear? What is the label for this topic? Do you know the topic from presswork? Is this an interesting topic?

- **Follow-up Interviews**: Follow-up interviews were conducted to receive more indepth assessments about the analysis results, which were provided within the analysis report and the questionnaire. Here the interview focused on the reasons that stakeholders answered the questionnaires in the way that they did.

5.2   Findings

- **Sensible and expected topics:** The toolbox is providing topics that were sensible and expected given the source data. All topics from local Facebook pages that were assessed as understandable were known beforehand. The reasons being: stakeholders are well informed about topics that arise or are discussed within their constituency. Stakeholders follow local social network channels and are part of social network discussions – so they are 'aware of the public area'. Regarding the further twitter analysis results the assessment was often the same: stakeholders monitor topics – therefore they are 'aware of the public area' and which subtopics being discussed. Within the samples that were shown to the interviewees the subtopics were identified and the topic of discussion was clear to them. Therefore the analysis is able to provide the topics that are relevant for the queried search on Twitter; if there are enough tweets.

- **Quality of topics:** When comparing the use cases 'Facebook topics' and 'Twitter topics', the Twitter results were more useful to the stakeholders. For topics like 'federal armed forces' all of the relevant subtopics were identified. Concerning Facebook the topics were better understandable and helpful for the interviewees when they were extracted from Facebook pages with high discussion activity (e.g. Angela Merkel[5] or the press).

- **Different meanings for topics:** All interviewees mentioned that the combination of five words for one topic could have multiple meanings. It is often the case that two or three words fit together and another word has a completely different meaning for the group of words as a whole. Another problem is that single words can also have different meanings. For instance 'dear': one interviewee mentioned that it was not clear to him if this word means the form of address, a verb, an adjective or if it is part of a substantive. Depending on the single meaning of the word the combination with other words can have different meanings.

- **Less clear topics with local Facebook pages:** All interviewees observed that the topics are often unclear for local pages. The reason why 42% of the 110 topics were assessed as understandable topics is due to the fact that policy makers know what's happening in the area of their electorate. The interviewees confirmed that the number of 42% in the questionnaire is a very optimistic number, because the interviewees often made a guess what the meaning of the topic could be. Most of

---

[5]   URL: https://www.facebook.com/AngelaMerkel (Retrieved 13/03/2013).

the topics were clear to them, because they know the 'real world' case and can therefore suggest the topic. All interviewees confirmed further that this background information is necessary for most of the provided topics. The interviewees argued that the analysis is only as good as the input data. On the local area there are not so many political debates that are public. But the results with Twitter have shown that the topic analysis is able to provide useful results.

The validation of sentiment analysis was not part of the questionnaire. But the analysis report covered at least one example similar to figure 1 that has been discussed during the expert interviews. Most of the interviewees can guess at the meaning of 'sentiment' and 'controversy' within the toolbox, and can use these indicators to choose a topic, and to read the posts contained within the group. But:

- It's not 'clear' to them why a discussion is either positive or negative as the visualization provides only one scale. For instance it may help to show the total number of both - the number of positive and also the number of negative comments.
- When combined with 'controversy', the 'sentiment' is less clear. End users have difficulties understanding what the discussion looks when only seeing both scales.
- The 'controversy' scale it is easier to understand when viewed separately. In general the UI needs improvement to provide the end users with a better understanding of its parameters.

6   Conclusions

This paper shows two case studies how to apply and validate topic opinion analysis for user comments on the Internet. While the first case study focused on the HeadsUp online discussion forum the second case focused on Facebook local pages and Twitter as social networks. Even if the approaches are different, both case studies follow the same process model and show added value as well as possible boarder-lines.

Both evaluation approaches were very effective with respect to the quality of validation and the end users' feedback. While the HeadsUp case focused on the accuracy and reliability of analysis data the social network case focused on the usability of analysis results to be used within the decision-makers' everyday life. However both cases were very time consuming. For instance the analysis reports and the extracted sample for the questionnaire needed current and personalized data - Facebook pages and topics for Twitter of interest to the stakeholder. Therefore this approach needed research time on the social web and continuous coordination with the end users to design an individual data report. Including all steps that were necessary to run this study, about one week was needed for each end user.

The toolkit returns on average a topic group for every 30 posts (HeadsUp) when the number of topic groups is not set manually by a user. However, the comments will not be distributed equally across the topic groups. Without being able to manually set the number of topic groups returned, the results were very hard to understand. With a medium sized forum of around 300 posts the outcome may be understandable but with smaller or larger forums the topic groups are either not refined enough or there

are so many topic groups that patterns are hard to see or too many similar topic groups are returned. However, it is important to note that the data being tested on the toolkit had already been analyzed manually so there was already an understanding of what the debates were about; discussions that were previously unseen may be more challenging for a user to understand.

In the case of HeadsUp, the toolbox could be helpful in analyzing forum data, particularly the larger forums with hundreds or thousands of comments. The toolkit takes seconds to analyze hundreds of comments, whereas human analysis takes days to see similar results. In the social network case the interviewees argued that the toolbox is a tool that is between them and the large amount of social network data. Therefore the toolbox needs to consider that the behavior of the social network may change frequently – for instance through new privacy settings on Facebook or the way that political parties in Germany, the pirates, have revolutionized discussions on social network using open and transparent methods. However the results were very useful on Twitter to inform on particular topics to see the width of a debate.

The toolbox is best at dealing with large quantities of data, amounts that could not be analyzed effectively by a human without significant resources to do so. The toolkit performs well on relatively in-depth data - this lends itself to blogs and forums that encourage more considered and less immediate responses. The toolbox also performed well in showing the nuances between different elements of a wider debate within the HeadsUp case and the Twitter case. With local Facebook pages the quality of results worsen due to the fact of the quality of input data and less political conversations. Instead of monitoring a bunch of local Facebook pages the interviewees proposed to select less pages, but with more qualitative and more active political debates.

Although the toolbox was primarily conceived of as a project focusing on the analysis of political conversations on social media, it also has applications for forums and blogs. Most websites now support comments and sites such as the BBC or Daily Mail regularly have hundreds of comments on each article.

Civil society groups also run forums and blogs to connect with their members and supporters. Analyzing the themes of these discussions is often beyond the resources these organizations have. The toolbox could play an important role in helping small not-for-profit organizations, larger media organizations, as well as politicians and policy-makers to understand feedback across a range of communication channels.

References

1. Coleman, S.: Connecting parliament to the public via the Internet: two case studies of online consultations. In: Information, Communication and Society, 7, 22. (2004)
2. Coleman, S.: Blogs and the New Politics of Listening. In: The Political Quarterly, 76, 8. (2005)
3. Hrdinova, J., Helbig, N. & Stollar Peters, C.: Designing social media policy for government: Eight essential elements. Center for Technology in Government. New York. (2010)

4. Joshi, S.; Wandhöfer, T.; Koulolias, V.; Van Eeckhaute, C.; Allen, B. & Taylor, S.: Paradox of Proximity – Trust & Provenance within the context of Social Networks & Policy. In: ACM Proceedings of The 4th International Conference on Social Informatics. (2012)

5. Koop, A., Feil, B., Habbel, F.-R., Klötzer, J., Kubicek, H., Lührs, R., Märker, O., Mitterhuber, R. & Trénel, M.: Leitfaden Online-Konsultation: Praxisempfehlungen für die Einbeziehung der Bürgerinnen und Bürger über das Internet. (2010)

6. Miller, L. & Williamson, A.: Third Phase Report: August 2007 – August 2008. Digital Dialogues. London, UK: Hansard Society. (2008)

7. Savage: Twitter as Medium and Message. Researcher are mining Twitter's vast flow of data to measure public sentiment, follow political activity, and detect earthquakes and flu outbrakes. In: Communications of the ACM, 54, 18-20. (2011)

8. Sizov, S.: GeoFolk: latent spatial semantics in web 2.0 social media. In: B. Davison & T. Suel & N. Craswell & B. Liu (Eds.), Proceedings of the third ACM international conference on Web search and data mining, pp. 281-290. (2010)

9. Wandhöfer, T; Taylor, S.; Alani, H.; Joshi, S.; Sizov, S.; Walland, P.; Thamm, M.; Bleier, A. & Mutschke, P.: Engaging politicians with citizens on social networking sites: the WeGov Toolbox. In: International Journal of Electronic Government Research (IJEGR), Vol. 8/ No. 3, p. 22-43. (2012)

10. Wilson, C. & Fletcher, R.: Appendix A, Legal Analysis of Issues Surrounding Social Networking Sites. In: WeGov Deliverable 5.1. (2010)

11. Rowe, M. & Angeletou, S.: Predicting Discussions on the Social Semantic Web. In: Semantic Web Conference. Heraclion, Crete. (2011)

12. Rowe, M., Angeletou, S. & Alani, H.: Anticipating Discussion Activity on Community Forums. In: IEEE International Conference on Social Computing. Boston, USA. (2011)

13. Rowe, M., Fernandez, M., Alani, H., Ronen, I., Hayes, C. & Karnstedt, M. Behaviour analysis across different types of Enterprise Online Communities. In: Web Science Conference (WSC), Evanston, USA. (2012)

14. Blei, D. M.; Ng, A. Y. & Jordan, M. I.: Latent Dirichlet Allocation. In: Journal of Machine Learning Research, 3, 993-1022. (2003)

15. Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K. & Harshman, R.: Indexing by Latent Semantic Analysis. In: Journal of the American Society for Information Science and Technology, 41, 391-407. (1990)

16. Addis, M., Taylor, S., Fletcher, R., Wilson, A., Fallon, F., Alani, H. & Wandhöfer, T.: New ways for policy makers to interact with citizens through open social network sites - a report on initial results. In: Internet, Politics, Policy, Oxford, UK. (2010)

17. Mitchell, R, Agle, B. & Wood, D. 1997. Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts. Academy of Management Review, 22, 853-866.