Trend-based and Reputation-versed Personalized News Network

Olga Streibel
Corporate Semantic Web
Free University Berlin
Berlin, Germany
streibel@inf.fu-berlin.de

Rehab Alnemr
Hasso Plattner Institute
Potsdam University
Potsdam, Germany
rehab.alnemr@hpi.uni-potsdam.de

ABSTRACT

Web users while collaborating over social networks and microblogging services also contribute to news coverage worldwide. News feeds come from mainstream media as well as from social networks. Often feeds from social networks are more up-to-date and, for user's view, more credible than those that come from mainstream media. But the overwhelming amount of information requires to personally filter through it until one gets what is really needed. In this paper, we describe our idea of a personalized news network built on current Web technologies and our research projects by filtering Twitter and Facebook messages using both trend mining and reputation approaches. Based on the example of Egyptian revolution, we explain the main idea of personalized news.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Information filtering

General Terms

Algorithms, Design, Experimentation

Keywords

trend mining, reputation-based systems, semantic web technologies

1. INTRODUCTION

Social networks like Delicious[1], Diaspora[2], Facebook[3], Flickr[4], LinkedIn[5], Twitter[6], Xing[7], YouTube[8] etc. have become very popular among users on the Web. In recent years, Facebook attracted hundred millions of users worldwide, increasing its membership from over 100 million in 2009 to over 500 million in 2011[9].

Around 175 million[10] of Web users in 2010 had a Twitter account. Everyday there are 95 million[11] of tweets worldwide and "more than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) each month" shared on Facebook[12]. Owing to these novel forms of communication, everybody could follow the developments during the flood in Rockhampton in Australia 2010/2011 since residents of this town created a public Facebook group reporting in real-time about the flood[13].

In mainstream media, the political events in Iran in 2009 have been described as Twitter-Revolution[14] since many people communicated about these events using the microblogging service Twitter. Furthermore, the political developments, and revolutions, in North Africa beginning from January 2011 could be followed on Facebook, Twitter, Flickr, Bambuser, etc. During this period, social networks became nearly the only trusted source of information. In Egypt, the media blackout and manipulation of facts led millions of users to extract information from several social network sources to be informed of what was really happening. Two big groups were formed in Twitter (with tag #jan25) and Facebook (R.N.N group)[15] to receive direct feed from the demonstrators in the streets. Public Facebook status updates, tweets, bookmarks, and pictures represent immediate knowledge about our world, generated by Web users. Among this content, many reports on breaking news emerge in real time.

[1] `http://delicious.com/` visited June 2011
[2] `http://joindiaspora.com` visited June 2011
[3] `http://www.facebook.com` visited June 2011
[4] `http://www.flickr.com` visited June 2011
[5] `http://www.linkedin.com` visited June 2011
[6] `http://www.twitter.com` visited June 2011
[7] `http://www.xing.com` visited June 2011
[8] `http://www.youtube.com/` visited June 2011
[9] `http://www.facebook.com/press/info.php?factsheet` visited June 2011
[10] `http://twitter.com/about` visited June 2011
[11] as for September 2010
[12] `http://www.facebook.com/press/info.php?statistics` visited June 2011
[13] `http://www.facebook.com/pages/Rockhampton-And-CQ-Floods-2010/184869738205940` visited June 2011
[14] `http://www.washingtontimes.com/news/2009/jun/16/irans-twitter-revolution/` visited June 2011
[15] `https://www.facebook.com/RNN.World?ref=ts` visited on June 2011

SMUC'11, October 28, 2011, Glasgow, Scotland, UK.

Considering this development from the computer science point of view leads to the conclusion that one can harvest user-generated content in an automatic way in order to provide breaking news reports from user to user. Contrary to the mainstream media, which offers breaking news reports either already processed or created only by a strict number of people (journalists), breaking news reports extracted from users posts allow for: multiple sources of information, reports from people directly involved in the event that they are reporting on, multiple views on individual events, etc. However, we are aware of disadvantages in this approach: the more sources of information, the more confusing information emerges and the question of trust in information is becoming more significant than in the case of a single-sourced breaking news reports.

We analyzed the usefulness of breaking news reports based on users posts and found out that sophisticated filtering is the core of the process of informing oneself. This filtering can be expressed in two stages that are the most important ones in the process of gathering information: discovering the trend in information and estimating the reputation of the information.

In this paper we present our approach of merging the two processes to attain such personalized news network that provides breaking news reports from users to users. By using this network, one will be able to have a personalized version of the news based on the current trends and one's trusted network. In section 2 we briefly discuss related research followed by a motivating example. Based on our analysis of the process which we called "informing oneself", in section 4 we illustrate the idea of the personalized news network. In sections 5 and 6, we explain the trend mining processes (used for trend estimation) and using reputation objects (used for applying trust) as a semantic artifact that help in realizing the idea.

2. RELATED RESEARCH

Previous research (1998-2004) on trend analysis proved that automatic trend detection from texts is possible, e.g. based on the content of news stories it is possible to predict trends in stock prices [11]. Several research works from the topic detection and tracking research field (TDT) [1] concentrated on statistical models for trend analysis. TDT identified core research tasks such as story segmentation, first story detection, cluster detection, topic tracking, and story link detection that still inspire research direction of current work, i.e. [14][21]. On the other hand, works summarized in [10] under the emerging trend detection research field (ETD) show diversity of general trend analysis systems. Some works focus on constructing complete systems for trend detection from news, e.g. [8][19], some on applying known algorithms, e.g. self-organizing neural networks [15], for the trend mining task. Current researches (2009-2011) concentrate often on probability based algorithms, i.e. [6] [9] and propose trend detection approaches adapted to the novel forms of online texts streams: i.e. Twitter[12][7]. To our best knowledge, the current trend analysis approaches seem to consider only the fact that there are new forms of news (i.e. Twitter). We think that considering two general aspects of current Web are of an emerging importance for the trend mining approaches: a) users share in real-time information on the Web and b) using Semantic Web technologies (i.e. RDF[16] standard and LinkedData initiative[17]) that enable machine-readable information interlinking allows for linking the explicit knowledge on the Web. Considering the Web as the most important source of texts and regarding these developments lead us to a novel viewpoint on trend mining research. We propose to take a knowledge-based perspective on trend analysis[18]. Regarding the problem of information filtering with a purpose of extracting real-time, meaningful and trustful information we chose to focus on the techniques for trend mining combined with a reputation approach.

Using reputation as a base for trust is becoming a critical factor especially that it is becoming easier to publish information about oneself -or anyone- online through platforms such as social-networking sites and photo-sharing services. Online reputation systems are the biggest and most obvious examples of reputation systems. Specialized news sites like Kuroshin.org[18], Slashdot[19], & Zdnet[20] are applying the concept of rating for their network participants. Nevertheless, news coming from social platforms such as Facebook or Twitter are neither rated nor weighed by a measure of trust.

3. MOTIVATING EXAMPLE

There are many possibilities for informing oneself about what is happening in the world. One method is to read Twitter messages and Facebook status updates as illustrated in Figure 1.

Figure 1: Informing oneself from timelines

Based on an informal survey we conducted, we analyzed the process of informing ourselves while reading *daily* Twitter timelines and Facebook "top news" during the unrest in Egypt from January 26th to February 11th, and discovered interesting issues about the process itself. We noticed that the process mainly involves *filtering out* the relevant information, and can be summarized in 5 steps:

[16]`http://www.w3.org/RDF/` visited June 2011
[17]`http://linkeddata.org/` visited June 2011
[18]`http://www.kuro5hin.org/` visited June 2011
[19]`http://www.slashdot.org/` visited June 2011
[20]`http://www.zdnet.com/` visited June 2011

1 Estimating trending messages:

1.1 Which topics are emerging in the timeline?

1.2 What are trending[21] hashtags[22] in general?

1.3 Which tweets include the trending hashtags and what are they about? What updates appear as the top news on Facebook?

2 Choosing trending and interesting tweets: Which timeline messages fit into my field of interest and piques my curiosity today?

3 Estimating information's reputation: Who are the authors of trending and interesting messages? Which of them are interesting according to my own field of interests and according to my own subjective criterions:

3.1 Is the author a real person or an Internet robot?

3.2 Do I trust in her/his information?

3.3 What does this person write about in general?

3.4 Does this person write more interesting messages to me?

4 Extending the list of trending and interesting messages by messages written by the authors of high reputation i.e. are in my immediate web of trust

5 Reading the information and external links in the tweets that are trending, interesting and reliable: linking to external news (blogs, mainstream news portals)

Figure 2: An example of situation-irrelevant tweet from February 3^{nd} 2011 that uses trending term #Cairo

By monitoring the process within the time period of January 26th - February 11th, an example of trending messages on Twitter were messages marked with the hashtag #jan25 and #Cairo. Most users *interested in the political developments worldwide* could notice an increasing number of messages in their timelines containing information about the situation in Egypt. Terms such as "Egypt", "revolution", "president Mubarak", "protest", "Tahrir square", etc. seemed to appear more frequently than usual. And more Twitter users started to retweet posts containing these words. However, misleading and irrelevant information was also posted using these trending terms for other purposes than reporting what was happening in Egypt those days (e.g. figure 2).

In order to distinguish useless information from valuable one, one had to search for more information about the authors of the tweets. i.e. reading their Twitter profiles or their past posts. In cases where it was not possible to find more information about the authors, the statistics were significant in estimating if the given author is a person who could post reliable information, i.e.: how many messages on similar topic did the author post, how many followers does the author have, how many other users retweeted the posts of this author and how did they comment on this post. After estimating the reputation value of the authors and hence the trust value of the trending and interesting posts, one could continue with reading the chosen tweets and the information contained within them (often the tweets contained links to other social networks posts: blogs, pictures, videos). Based on this experiment (staying properly informed about the events in Egypt), we noticed that trend estimation and reputation calculation steps were crucial.

[21]based on Twitter's own trend estimation for trending tags in tweets

[22]`http://hashtags.org/` visited June 2011

4. PERSONALIZED NEWS NETWORK

Based on our analysis as described above, we propose a novel news system which we call *personalized news network (PNN)*. The main idea is to offer the user reports on breaking news tailored to the user's interests, filtered out from other users posts. We concentrate particularly on the filtering process.

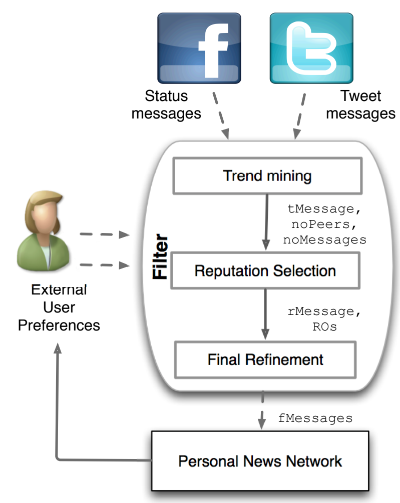

Figure 3: Personal News Network Architecture

4.1 The Filtering Process

The filtering process is carried out by implementing three phases of filtering. They are:

Trend Mining.
where

input:	stream of status messages from Facebook friends and specified groups: `stMessages` and Twitter followers: `twMessage`
output:	a stream of triples `tMessage` (trended messages), `noPeers` (number of people posted messages with the same labeled trend), and `noMessages` (number of messages posted with the same labeled trend)

Reputation Filtering.
where

input:	a stream of trended messages `tMessage`
output:	a sorted stream of trended messages and their reputation objects `rMessage` , `RO`

Knowledge Retrieval.
where

input:	ordered stream of trended messages where each message has its reputation attached to it
output:	the final stream of messages to be displayed in the PNN for the user (personalized messages)

4.2 Advantages

The advantages of our filtering technique for the news messages lie in using the $q = two$ approaches of trend mining and on estimating a measure of trust for the news based on reputation filtering. The advantages are:

1. *Early news:* As mentioned in the introduction, one of the main advantages of getting news from a social network is to be informed earlier than being informed by watching/reading regular news media.

2. *Multiple Sources:* The filter has multiple information sources such as Twitter and Facebook.

3. *Trend estimation based on our trend mining approach:* Focus on field of user's interest, on statistic and semantic trend values for messages

4. *Reputation calculation based on our Reputation Object (RO) approach:* Focus on the "importance value" of the calculated trending messages

5. *Trustable:* The reputation of the information plus the trust in its source (followers on Twitter or friends on Facebook).

6. *User control over feeds:* The filter takes as input user preferences (time window parameters, threshold values, etc.) that are used to prioritize the flow of the shown news messages for the user.

5. ESTIMATING TREND

Continuing with the example of informing oneself by reading Twitter posts from users' reporting on unrest in Egypt from January to February 2011, we introduce very briefly our trend estimation approach. According to the general definition as provided by related research on trend mining, trends in texts are defined as emerging topic areas and an emerging topic in texts is a topic that "increases in interest and utility over time"[10]. In order to estimate a trend we need to define a time window, outliers, interestingness, utility and the trend indication. The definitions are as following:

Def. Time window and time slice
t_{window} is a time interval in which trends can occur. A day is an example of time window
t_{slice} is a subinterval of time window. If its starting point lies at t_0 the end point has to lie at $t_k < t_n$. An hour is an example of a time slice.

$$t_{window} = [t_0 ... t_n] \wedge t_{slice\,k} = [t_0 ... t_k]$$

$$t_{window} :=< t_{slice\,k}, ..., t_{slice\,n} >$$

$$|t_{slice\,k}| = |t_{slice\,n}| \wedge k, n \in \mathbb{N} \wedge k < n \tag{1}$$

Def. Outliers
Outlier is a term that appears significantly often in a given time slice comparing with the whole window. Outlier as value can be determined for every term by calculating:

$$outlier(w)_{t_{slice}} := TF_{(w,|P|_{t_{slice}})} * IPF_{(w,|P|_{t_{window}})} \tag{2}$$

$$IPF_{(w,|P|_{t_{window}})} := \log \frac{|P|_{t_{window}}}{PF(w)_{t_{window}}}$$

whereas: $TF_{(w,|P|_{t_{slice}})}$ says how frequent does term w appear in the posts of particular time slice[23]. $|P|$ expresses the total number of given posts. $IPF_{(w,|P|_{t_{window}})}$ determines the appearance of the particular term w in the whole window.

If we consider the beginning of the Egyptian unrest reports on 25th of January and the $t_{window} = day$,$t_{slice} = hour$, the terms as "#jan25", "Egypt", "revolution" were outliers on this day and would have the most significant outlier values among other terms of chosen time window.

Def. Interestingness

$$interest(w)_{t_{slice}} = f(w)_{t_{slice}} := \log \frac{TF_{(w,|P|_{t_{slice}})}}{|W|_{t_{slice}}} \tag{3}$$

whereas: $|W|$ is the number of all terms considered.

$$interest(w)_{t_{window}} :=$$

$$< f(w)_{t_{slice}\,k}, f(w)_{t_{slice}\,k+1}, \ldots, f(w)_{t_{slice}\,n} > \tag{4}$$

expresses increasing interest if:

$$f(w)_{t_{slice}\,k} < f(w)_{t_{slice}\,k+1} < \ldots < f(w)_{t_{slice}\,n}$$

[23] the calculation is based on the principle of the weightning method TFIDF[16] by including time as an calculation dimension

The interest values of our example terms "#jan25", "Egypt", "revolution" were constantly increasing over the time slices of time windows beginning from January 25th.

Def. Utility
The utility can be described by the number of resources that, in given time window, have been tagged with the term w divided by the number of all resources tagged in this time window:

$$util(w)_{t_{window}} := \log \frac{|R|_{(tag=w)t_{window}}}{|R|_{(tag)t_{window}}} \quad (5)$$

whereas: $|R|$ is the number of resources (posts, status messages, tweets) in the given system.

Similar to the interestingness, we can identify increasing and decreasing utility for every term while looking slice for slice in whole time window, of the utility value increase or decrease. Regarding our example, the utility values of "#jan25", "Egypt" were significantly high since more and more messages were tagged with these terms.

Def. Trend indication
Only terms with significant high outlier values are to be considered as trend indicating. Terms with their outlier values below a certain threshold can be omitted.

$$trendind(w)t_{window} = \frac{interest(w)t_{slice} * util(w)t_{window}}{ratio(t_{window})} \quad (6)$$

whereas:

$$ratio(t_{window}) = |t_{window}|$$

is the size of time window given by the number of its time slices. A feature in text (i.e.: term, term pair, concept) is *trend indicating* if: a) it has significant outlier value and b) its interest and utility values, in relation to the frequency of the time window, are increasing. The trend indication of this feature is the value as defined in Def. 6. and the more trend indicating features are contained in a post, the higher is its trend indicating value. We also define:

Trend features:
The stepwise weighting method based on functions as presented above allows us to select the trend features out of a given text corpus in a chosen time window. Trend features are in particular, the trend indicating terms in texts.

Trend feature enhancement:
The particular terms of high trend indicating values are, if treated separately, without any meaning. Only if we know that "Cairo" is the capital of "Egypt", "Tahrir Square" is a place in Cairo, we can conclude that these terms are also semantically connected and probably describe one particular place on earth. In order to verify any connections between the trend indicating terms, we use knowledge from an ontology and finally express the terms as RDFS[24] concepts. A trend ontology[18] or any ontology that describes the domain in which we are looking for trends is applicable for enhancing the calculated terms (in our example: political domain). In our example, we propose first to look up in Dbpedia[5][25] since it is one of the most popular sources of structured knowledge on the Web[26].

The *Trend estimation algorithm* that we used to realize the first phase of the filtering process is:

```
//PRE: stMessages, twMessages
PREPROCESS:
  findBestTimeSlice(PARAM:time window);
  parsMessage();
  calculateStopWordList(PARAM:time window);
 for each stMessage,twMessage{
   removeStopwords();
   stemm();
   tokenize();}
    tsVectors = createTimeStampedMessageVectors();
     END PREPROCESS;
TREND FEATURE SELECTION:
  for each term in tsVectors{
  calculateOutlierV();
  calculateInterestingnessV();
  calculateUtilityV();}
  for each term in tsVectors{
  if (term.OutlierV > threshold &&
      term.InterestingnessV increases in tsVectors &&
      term.UtilityV > threshold )
      calculateTrendindication(term);
     tTermList = addTermToTrendingList();}
     END TREND FEATURE SELECTION;
TREND FEATURE ENHANCEMENT
 for each term from tTermList{
  if(lookupOntology(term)){
   createRDFdescription(term);}}
   for each message from tsVectors{
    if (message contains term from tTermList){
    createRDFdescription();}}
//POST: tMessages, noMessages, noPeers
```

6. CONTEXT-AWARE REPUTATION

Most of the existing work on reputation systems focuses on improving the calculation of reputation values where reputation is mostly represented in a singular value form. This work focuses on how to represent reputation to reflect its real-world concept (i.e. non-general, context specific and dynamic) and to facilitate reputation information exchange or reputation interoperability in any domain. The argument is that in most reputation systems, the context of a reputation value is not embedded within the given reputation information because it has the single value format. Since reputation changes with time and is used within a context and every domain has its own information sources as well as its own requirements, the representation -not the calculation- of reputation should be unified between communities in order to facilitate knowledge exchange. In this model reputation is represented as a new form of reputation value: *Reputation Object (RO)*. This object holds information on the reputation of an entity in multiple contexts. The ontology's components are: a `ReputationObject` `hasCriteria` one or multiple instances of class `Criterion` or `QualityAttribute` (for a service, the criterion describing service reputation is referred to as a quality attribute). The criterion is collected using a `CollectingAlgorithm` and

[24] `http://www.w3.org/TR/rdf-schema/` visited in August 2011
[25] `http://dbpedia.org` visited in August 2011
[26] "3.64 million "things" with over half a billion "facts" " `http://wiki.dbpedia.org/Datasets` visited in August 2011

`hasValue ReputationValue`. Each criterion instance has a `ReputationValue` (which includes the `currentValue\`, its time stamp, and a simple list of its previous values called `historyList`) that in turn has the range of values defined in `PossibleValues`. It describes the data type that the criterion can have or a specific set of values (literals or resources URI) evaluating this criterion. Each time there is a new entry value for this criterion, a `currentValue` is calculated using the `ComputationAlgorithm` which is the reputation computation function used with this criterion.

Since it is not always easy to identify intuitively what the highest reputation value is among the defined possible value set, the `PossibleValues` class has an `orderedList` that is ordered from the relatively highest reputation value to the lowest. It has also the possibility to define a comparison and ordering function; `OrderFunction` to compare between values within each criterion. The ontology is implemented using Protégé-OWL and also a Java library is implemented to facilitate the integration within any system [4] [2]. It has been used with several use cases.[3][17][13]

For this work, the purpose of using this ontology of reputation is that the messages coming as an input after the trend mining process are prioritized and rearranged based on their reputation. Also, it provides structured information for the knowledge retrieval phase to filter out the messages that will be displayed to the user. We fixed three contexts to be represented in the reputation objects: *who, tweetsstatus, sources*. The *value* of these contexts is always `RDF` triples [20] for both expressiveness and interoperability. The idea is that a message's reputation depends on who posted it, how many tweets or status messages about each trend or topic is involved, and how many friend/foaf/followers are involved in this trend. For simplicity, we show the structure of the RO as a matrix (though in the implementation it is not a flat matrix).

Table 1: Message's Reputation Object

Contexts	Values as RDF triples
who	`<FOAFAgent,has,weight>`
tweetstatus	`<noPeers,posted,noMessages>`
sources	`<FOAFAgent,netPeers,relativeContexts>`

where `relativeContexts` is a URI referring to similar messages, `netPeers` is the network size of the source who posted the message, `weight` is 1 if the message originated from a direct friend and 0.8 if the message origin is a friend-of-a-friend. In our experiments we noticed that one of the effective point to the variable `netPeers` is that sources with a low number of friends tend to be a scam. We have experienced such problems while following the "R.N.N" news group on facebook, where thousands of fake accounts were created as a form of anti-revolution act by the government to spread false news.

In the case of our motivating example, based on observations of 200 posts in the news group, we discovered that 82% of the people posting scam/fake messages have less than 78 friends. Therefore, we propose to discard posts by people who have less than x friends (hence the threshold of `netPeers` $<x$ in the algorithm, whereas we set $x = 80$). Based on the external input from the user that has her preferences, the process is:

```
PROCESS:MAIN
read tMessage,source,noPeers,noMessages
get netPeers(source)
if netPeers < x
   discard
else
   For all tMessage
     computeRO()
get userPreferenceList
sort(userPreferenceList)

PROCESS:computeRO()
if source=friend
   weight=1
else
   weight=0.8
set ROParameters()

PROCESS:sort(userPreferenceList)
get priority P2 from userPreferenceList
For all tMessages
   select the tMessage with the highest P1 values
   save in FilterList1
get priority P2 from the userPreferenceList
For all tMessages in FilterList1
   select the tMessage with the highest P2 values
   save in FilterList2

POST:Ordered List of tmessage,RO
```

The process constructs the reputation object (RO) for each message (based on the conditions described in last paragraphs) and gets from the user her preferences; i.e. if she cares more about messages posted only by her friends then the message list is sorted using the `who` triples, and if she cares more about messages posted by many people and many times then the message list is sorted based on the highest `noPeers` and `noMessages` and so on. At the end the messages are rearranged and with their ROs attached are passed to the final processing stage.

7. CONCLUSION AND FUTURE WORK

This paper reflects a real situation that we experienced during the Egyptian revolution from 25^{th} of January to 11^{th} of February 2011. In order to stay informed, we were frequently flipping TV channels and reviewing updates on the Web from different mainstream media, as well as navigating social networks updates from Twitter and Facebook. The experience made us realize two critical facts. The first is that mainstream media seemed to be delayed in breaking news reports, and that the real-time information comes from the combination of social networks updates. This was our motivation to start the conceptual design and the implemention of the PNN, as a way of staying informed as well as a tool to test our combined research topics. The combination of our research approaches that we are focusing on - knowledge-based trend mining and reputation calculation - turned out to be a powerful method for filtering interesting real-time news. The second realization is the importance of social networks over mainstream media. Watching the Egyptian news (that at the time was controlled by the government) and comparing it to the other information sources, we realized that almost none of the broadcasted news was true. Reviewing the mainstream news from several countries, in several languages, made us realize that news reported there are not showing the broadness of the situation as it could be derived from social networks and Web blogs. Gathering news from different tweets and Facebook status updates and groups were not only much more informative but also seemed more trustworthy.

For the aforementioned arguments, we developed the concept of a personal news network to provide the user with personalized news, based on a calculated trend and on reliable sources. The application is currently under ongoing development but first experiments based on our research approaches show promising results of our filtering process.

8. ACKNOWLEDGMENTS

This work has been partially supported by the InnoProfile-Corporate Semantic Web project funded by the German Federal Ministry of Education and Research (BMBF) and the BMBF Innovation Initiative for the New German Laender - Entrepreneurial Regions. The authors would like to thank Radoslaw Oldakowski for his critical and useful feedback on this paper, Kelly Ann Callahan for her correction on English writing.

9. REFERENCES

[1] J. Allan, editor. *Topic Detection and Tracking. Event-based Information Organization.* Kluwer academic publishers, 2002.

[2] R. Alnemr, J. Bross, and C. Meinel. Constructing a context-aware service-oriented reputation model using attention allocation points. *Proceedings of the IEEE International Conference on Service Computing*, 2009.

[3] R. Alnemr, S. König, T. Eymann, and C. Meinel. Enabling usage control through reputation objects: A discussion on e-commerce and the internet of services environments. In *Special issue of Trust and Trust Management, Journal of Theoretical and Applied Electronic Commerce Research*, 2010.

[4] R. Alnemr, A. Paschke, and C. Meinel. Enabling reputation interoperability through semantic technologies. In *ACM International Conference on Semantic Systems*. ACM, 2010.

[5] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web.*

[6] L. Bolelli, c. Ertekin, and C. L. Giles. Topic and trend detection in text collections using latent dirichlet allocation. In *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*, ECIR '09, pages 776–780, 2009.

[7] S. Goorha and L. Ungar. Discovery of significant emerging trends. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD2010, 2010.

[8] S. Havre, E. Hetzler, P. Whitney, and L. Nowell. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8:9–20, 2002.

[9] N. Kawamae and R. Higashinaka. Trend detection model. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 1129–1130, 2010.

[10] A. Kontostathis, L. Galitsky, W. M. Pottenger, S. Roy, and D. J. Phelps. *A Survey of Emerging Trend Detection in Textual Data Mining.* Springer-Verlag, 2003.

[11] V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan. Mining of concurrent text and time series. In *In proceedings of the 6 th ACM SIGKDD Int'l Conference on Knowledge Discovery and Data Mining Workshop on Text Mining*, pages 37–44, 2000.

[12] M. Naaman, H. Becker, and L. Gravano. Hip and trendy: Characterizing emerging trends on twitter. *J. Am. Soc. Inf. Sci. Technol.*, 62:902–918, May 2011.

[13] A. Paschke, R. Alnemr, and C. Meinel. The rule responder distributed reputation management system for the semantic web. In *RuleML-2010 Challenge, Washington DC, USA.* ACM, 2010.

[14] S. Petrović, M. Osborne, and V. Lavrenko. Streaming first story detection with application to twitter. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, HLT2010, pages 181–189, 2010.

[15] K. Rajaraman and A.-H. Tan. Topic detection, tracking, and trend analysis using self-organizing neural networks. In *Proceedings of the 5th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, PAKDD '01, pages 102–107, 2001.

[16] G. Salton. *Automatic Text Processing.* Addison-Wesley, 1989.

[17] M. Schnjakin, R. Alnemr, and C. Meinel. A security and high-availability layer for cloud storage. In *The 2nd Int. Workshop on Cloud Information System Engineering (Springer CISE 2010)*, 2010.

[18] O. Streibel and M. Mochol. Trend ontology for knowledge-based trend mining in textual information. In *7th International Conference on Internet Technology: New Generations*, pages 1285–1288, 2010.

[19] R. Swan and D. Jensen. Timemines: Constructing timelines with statistical models of word usage. In *KDD-2000 Workshop on Text Mining.*

[20] W3C. Resource description framework (rdf). `http://www.w3.org/TR/rdf-concepts/`.

[21] J. Zeng and S. Zhang. Incorporating topic transition in topic detection and tracking algorithms. *Expert Systems with Applications*, 36(1):227 – 232, 2009.