@twitter analysis of #edmedia10 – is the #informationstream usable for the #mass

Martin Ebner, Thomas Altmann, Selver Softic
Graz University of Technology, Austria

Summary

In this paper we report the use of an application that enables an automatic analyses of social media content. In this early stage of development our work focuses on data from Twitter[1] as currently to be the most popular and fastest growing microblogging platform. After an introduction about a general concept the conference tweets of a big e-learning conference are examined. It is aimed to show whether there is a possibility to get significant information from a pool of postings or not. The publication concludes that a keyword extraction can be taken as basis for further investigations and treatment of data.

Keywords: Twitter, microblogging, semantic, linked data.

---

[1] http://www.twitter.com (last access September 2010)

Introduction

Most of the content on the Web nowadays is user generated as many of new Web applications focus on user activity (Web 2.0). Especially microblogging gained strong importance in recent years. Current microblogging platforms like Twitter, Jaiku[2], Tumblr[3], Emote.in[4] attract every day masses of users with different social, cultural and educational background. Furthermore microblogs tend to become a solid media for simplified collaborative communication. Templeton (Templeton, 2008) defines microblogging as a small-scale form of blogging made up from short, succinct messages, used by both consumers and businesses to share news, post status updates and carry on conversations.

By communicating people share different kind of information like common knowledge, opinions, emotions, and information resources and their likes or dislikes. They discuss on different topics in more or less open communities (Dejin and Rosson, 2009). Information chunks (small scale messages) produced by masses and structured in adequate way offers new scientific insights which are interesting not only for science but also for commercial use like monitoring of trends, advertising, statistics, reputation management and news broadcasting. Possibility of monitoring and analysis of such messages not only restricted to humans but also for machines would enormously contribute the exploitation of already present on-line social data essence e.g. for educational, commercial or informative purposes. Short form of content posted by microbloggers offers also a solid base for automated content processing and analysis. However current state of research lacks of simplified architectural paradigm approach according to these problems.

Twitter along with Facebook[5] still belongs to the fastest growing social web platforms of the last 12 months[6]. On 22nd of February 2010 Twitter hits 50 million tweets per day[7]. Without any exaggeration it can be said that these two social networks are worth to be researched in more detail (Haewoon et al., 2010).

This publication aims to introduce to a new tool that analyses the twitter stream by keyword extraction. Beforehand the whole environment of the tool is explained as well as the goal of the research. Finally one of the biggest e-learning conferences (ED-MEDIA) is examined as example.

---

[2] http://www.jaiku.com (last access September 2010)

[3] http://www.tumblr.com (last access September 2010)

[4] http://www.emote.in (last access September 2010)

[5] http://www.facebook.com (last access September 2010)

[6] http://ibo.posterous.com/aktuelle-twitter-zahlen-als-info-grafik (last access April 2010)

[7] http://mashable.com/2010/02/22/twitter-50-million-tweets (last access April 2010)

Related Work

Microblogging and Twitter

Although the beginning of first serious microblogs goes back to spring 2006 their leverage on the web grows rapidly. Especially in the area of communication and social networking microblogging is gaining significantly daily (Dejin and Rosson, 2009). Tumblr, Jaiku, Emote.in and identi.ca are only some of current massively used microblogging platforms. Most significant among them is Twitter, which induced a new culture of communication (McFedries, 2007) (Java et al., 2007). The 140 characters restriction for twitter messages can also be compared with a short-message internet-based service platform. Users can send a post (tweet) that is listed on the top of their twitter-stream together with messages of people they are following. Furthermore in principle any user can be followed by anyone who is interested in that user's updates. By nature Twitter or similar services support the fast exchange of different resources (links, pictures, etc.) as well as fast and easy communication amongst more or less open communities. In the same way Java et al. (2007) defined four main user behaviors why people are using Twitter – for daily chats, for conversation, for sharing information and for reporting news.

It has been shown lately that the usage of Twitter at conferences helps to increase reports, statements, and announcements as well as supports fast conversation between participants of the conference at face and online. Nowadays very often so called Twitter-streams on base of hash tag search projected nearby an ongoing presentation (Ebner, 2009) or placed at any other location at the conference support the conference administration, organization, discussions or information exchange. From this point of view microblogging becomes a valuable service reported by different publications (Ebner, 2009; Reinhardt et al., 2009).

Because Twitter and other microblogging platforms are very easy to use masses of information is being produced. Analysis of that content in a systematic manner using standards and adequate automated techniques enables the consumer of information to get valuable data categorized by aspects like locality, time, popularity, categorization etc.

Although Twitter already has an API with advanced search functionality, retrieved data lacks of usability. Bringing these results into a structured form with appropriate domain description using wide accepted vocabularies for a specific knowledge domain would increase the relevance of information retrieved through mining and exploration of such content. Second disadvantage of Twitter API[8] is that results are restricted to the last 3200 tweets.

Microblogging at Conferences

Graz University of Technology (TU Graz), especially the Department of Social Learning (DSL) has been using Twitter since it's existences as well as has started very early to investigate on the utilizability of Twitter on different purposes as for research at

---

[8] http://apiwiki.twitter.com (last access September 2010)

conferences. Already in 2008 (Ebner, 2009) a first study at ED-MEDIA 2008 examined the use of the platform to enhance the keynote presentation by providing a second projector displaying the Twitter stream of the hash tag search «edmedia». It was pointed out that this setting can improve the interaction between the auditorium and the presenter in a new way. The so called «murmuring in the background», the «noise of listeners» becomes visible to a broader audience and support discussions. Following this approach one year later at ED-MEDIA 2009, the whole conference stream before, during and after the conference was examined (Ebner and Reinhardt, 2009). As usual the hash tag was monitored to get all tweets concerning the event. Beside a quantitative analyses (number of daily tweets listed by user) also a first qualitative analyses pointed out key terms used. The publication concluded how or for which purpose people are using Twitter at conferences. It was shown that mainly the exchange of different resources and social activities as well as communicating with the community are crucial factors. Due to the fact, that microblogging is often postulated as an backchannel for non-participants a further research study aimed to underline the importance of Twitter for people following the conference only online on base of that stream (Ebner et al., 2010). Using the EduCamp 2010 in this context a detailed qualitative analyses of tweets was carried out. On the one side it can be summarized that the high number of so called ReTweets (tweets that are being posted a second time by another user to underline the importance of a tweet) handicaps following the conference stream and on the other side only 120 of 2110 (6%) tweets seemed to be useful for a non-locally participant. These results evoke to rethink the way Twitter can be used at conferences or as a «backchannel».

As a first consequence Twitter is to be seen more as a kind of personal digital jotter to store relevant data (hyperlinks, photos, statements, etc.) and as communication tool amongst participants for networking than for a conference service for online participants. So using Twitter to inform a broader community about the event cannot be suggested per se. This outcome leads to the next steps of developing a tool that allows grabbing tweets and storing them offline on personal hard drives. This application allows to search in messages after an event, especially addressing the jotter functionality (Mühlburger et al., 2010). Furthermore the stored data can be transferred to different ontologies for further semantic analyses, which are currently under process.

In this publication we introduce a second tool, called STAT that allows a keyword extraction by presenting a short example of examining postings of the e-Learning conference ED-MEDIA.

System Architecture

Concept

Basically our experimental system paradigm depicted in Fig. 1 describes the basic mining architecture consisting of three fundamental parts: data acquisition, data extraction and analysis as final step.
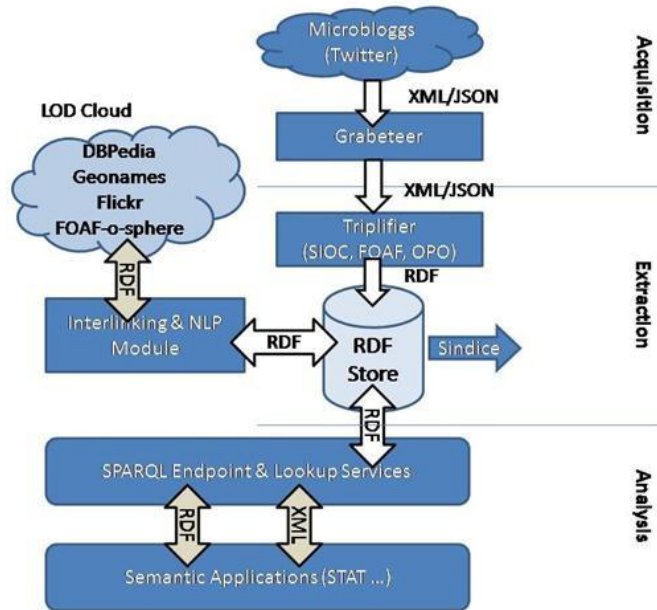
Fig. 1 - Experimental architectural paradigm for semantic microblog mining

In the acquisition step tweets are simply grabbed using Twitter API and subsequently stored into a local database or passed automatically through to extraction phase. Extraction phase «triplifies» the microblogs content into RDF triples and stores the data into a triple store of choice. Also interlinking of tweet bodies is done in this step. Finally in the analysis phase the stored RDF triples are exposed over SPARQL Endpoints or by using the Lookup Services to humans and machines.

The overall concept about the waylinked data and RDF triples are used for semantic analyses is described in Selver et al. (2010. Now we concentrate only on the last part – the tool STAT.

Analysis by STAT

Analysis is done using semantic technologies and simple text based techniques. Especially for Twitter we developed for demonstration purposes a so called STAT (Semantic Tweeter Analysis Tool) Infrastructure. STAT is still in development at this moment, but the first beta-version is already online (http://vlpc01.tu-graz.ac.at/~altmann). However we are aiming to create an analysis system that will be able to answer simple questions about people and their actions and interactions on microblogging platforms. A snapshot of the current status of STAT is depicted in Fig. 2.

Fig. 2 – STAT (Semantic Twitter Analysis Tool) development snapshost

The main functionality can be described by a keyword extraction of number of tweets. In case of conferences a hash tag search is initialized using the so called Twatter Keeper[9] application and via the provided API tweets downloaded and examined.

First study – ED-MEDIA 2010

Design of the study

For this study the ED-MEDIA 2010 conference is used. ED-MEDIA is an international conference on «Educational Multimedia, Hyper- media & Telecommunication» and started in 1993 as follow-up after 6 years of International Conferences on Computers and Learning (ICCAL). The main purpose as stated on their webpage is to serve as a multi-disciplinary forum for the discussion and exchange of information on the research, development, and applications on all topics related to multimedia, hypermedia and telecommunication/distance education. Nowadays it is certainly one of the largest international conferences on these topics. Over 1000 participants every year attend numerous sessions and workshops for 5 days. Two recent publications (Khan et al., 2009) (Mendez and Duval, 2009) pointed out the huge amount of contributions, the relationship of authors, the key players and lots of more trends.

This year conference took place in Toronto, Canada and the used hash tag was #edmedia. A Twitter archive at Twatter Keeper was started and downloaded afterwards.

---

[9] http://twapperkeeper.com/index.php (last access October 2010)

Results

An initial analysis produced the following results:

Which persons (@) were using the hash tag #edmedia

Which @persons write about #edmedia and how often

mebner (235), gsiemens (112), walthern (108), CosmoCat (72), Nona_Muldoon (66), ProfBravus (61), benbull (57), NancyWhite (49), psychemedia (45), LisaMLane (45), Downes (38), schwier (37), klconover (36), cogdog (36), cosmo07 (32), gconole (29), anitsirk (27), mdrapp (24), aoyamassi (22), LizFalconer (20), …

Which other keywords were used with #edmedia

Which keywords were used with #edmedia

rt (452), is (352), i (237), from (190), my (155), about (147), with (122), that (120), learning (120), it (120), this (119), are (117), we (112), be (111), your (104), have (100), not (99), great (97), social (94), presentation (93), media (87), will (86), all (84), what (83), but (82), as (79), by (78), out (77), use (76), how (73), talk (72), good (71), our (70), keynote (67), twitter (67), can (66), me (66), online (65), now (63), so (62), & (60), thanks (59), just (58), get (57), more (56), web (56), if (56), paper (54), ideas (54), do (53), …

Which hash tags (#) were used with #edmedia

Which #hashtags were used with #edmedia

#toronto (20), #ple (19), #hermannmaurer (18), #whoweare (9), #highered (8), #travelbacktoaustria (7), #keynote (7), #poster (6), #grabeeter (6), #frank (6), #roombay (6), #elearning (6), #ple_bcn (6), #ukoer (5), #audioboo (5), #xphone (4), #graz (4), #twitterstream (4), #oer (4), #edtech (4), #uoit5199 (4), #mlearning (4), #film (3), #edreform (3), #equity (3), #moodle (2), #scmedu (2), #workshop (2), #secondlife (2), # (2), #colaab (2), #downes (2), #iste10 (2), #digitalworld (2), #mebner (2), #education (2), #digitalnatives (2), #sakai (2), #electricity (2), #mustsee (2), #stat2 (2), #telbib… (2), #digcult10 (2), #maurer (2), #toronto; (2), #allhailtonyhirst (2), #virtual (1), #aloha09 (1), #m-learning (1), #iphone (1), …

Summary

First of all the analyses range the Twitter user regarding to their tweeting activity concerning conference issues. Therefore it's easy to find out who is very interested in the

topic. In addition this data is used for a deeper analysis by investigating specific persons of interest.

In 4.2.2 keywords used within a tweet containing #edmedia are listed. This list consists of all the regular words in a specific tweet. In this category, the structure of natural language is responsible for a high degree of words that hold no information value out of used context.

Nevertheless, it is possible to draw some conclusions. The top word is «rt». RT is the short form for «ReTweet» and means that anyone is tweeting again a copy of a tweet of a friend. The main purpose for doing a retweet is to open a piece of information to a broader community. Simply it can be stated that a RT is a kind of expression of importance. The second and third most used words are «is» and «I», respectively. These words are very common, and it can be considered to blacklist them. But these words still hold some information. The word «is» shows us that users tweet a lot about the present, rather than about the past or the future («is» instead of «will» or «was»). With 352 mentions «is» is much more common than «will» with only 86 mentions. «I» indicates that users tweet mostly about their personal experiences. The word «my» in fifth place confirms this trend. The first noun in this list is «learning». This corresponds well with the topic of the conference. The first adjective is «great». It indicates that the attendees seem to like this conference, or at least that they tweet mainly about positive experiences. Other words in the list like «social», «media», «presentation», «talk», «keynote», «good», «web» and «twitter» give further clues about topics and reactions concerning this conference. Social media and the web seem to be important topics. There are also a lot of tweets about keynotes and other talks. «Papers» are presented, «ideas» are exchanged, and a lot of users just want to say «thanks».

Finally it is analysed which other hash tags are used together with #edmedia. Obviously, they are easier to interpret than keywords because they are meant to be meaningful on their own. In contrast to the list of keywords, it is not necessary to filter irrelevant words, because hash tags are made of words that the writing user thought to be relevant. The most used tag is «#toronto», the location of ED-MEDIA 2010. The tags «#PLE» and «#ple_bcn» stand for personal learning environment. This was obviously one of the topics at that conference. Other topics seem to be «#grabeeter», «#elearning», «#audioboo» and «#edtech». Furthermore, there are tweets about «#keynotes». The tag «#hermannmaurer» is used many times, which indicates that he might have made an important contribution to the conference. All of these tags are a good starting point for further analysis, or even just a simple search on the web to find out more about the topic at hand.

This first analysis clearly shows an important aspect of STAT: the analysis works best when users tag their tweets sufficiently and appropriately. And then a more detailed investigation can be made to get further interesting information. If for example the hash tag «#hermannmaurer» is taken, following result can be shown by using STAT.

Result of analysis with «#edmedia» and «#hermannmaurer»:

- Which @persons wrote #edmedia together with #hermannmaurer

mebner (11), digitisation (2), okinasevych (2), johnnigelcook (1), yvonhuybrechts (1), ErikDuval (1)

- Which keywords were used with #edmedia and #hermannmaurer
  is (9), rt (6), showing (6), http://bit.ly/az7zd0 (4), movie (4), this (4), as (3), flying (3), 's (3), about (3), future (3), talk (3), credible (2), impression (2), me. (2), video (2), want (2), cars? (2), [must (2), not (2), "nokia (2), source (2), audioboo: (2), disney (2), see: (2), mixed (2), reality" (2), reference? (2), 20 (2), data (2), lost (2), i (2), http://bit.ly/9njwfp (2), http://boo.fm/b146833 (2), media (2), popular (2), xphone!] (2), his (1), learning (1), just (1), "any (1), precondition (1), years (1), topic (1), talking (1), session (1), thanks (1), second (1), beyond (1), visions (1), …

- Which #hash tags are used with #edmedia and #hermannmaurer
  #keynote (5), #xphone (4), #toronto (2)

First of all it can be seen that mainly one single user was using the pair of hash tag #hermannmaurer and #edmedia; this users tweets then seemed to be retweeted by other users considering the number of RTs (6). The other words give us a rough overview about the content of the tweets: «showing», «movie», «talk», «future», «flying», «cars», etc.

The three hash tags used in addition to the two analyzed tags are #keynote, #xphone and #toronto. A possible conclusion is that «#hermannmaurer» gave a «#keynote» presentation at the conference in «#toronto» and that «#xphone» played a crucial role in his talk.

Conclusion

In this publication the possibility of semantic analysis of short postings through microblogging is discussed. A simple tool, called STAT, was used to take a closer look at conference tweets, based on keyword extraction.

First results have shown that it is possible to get meaningful outcomes, for instance filter relationships between words used in order to find important content or concentrate on the users themselves. Of course the interpretation of such analysis is limited and will never replace personal participation, but it gave a short and general overview. This first analyses can be seen as starting point for further semantic analyses for example to interlink other resources automatically. So simple information can be enhanced to make it more reliable.

Bibliography

Boyd D., Golder S., Lotan G. (2010), *Tweet, tweet, retweet: Conversational aspects of retweeting on twitter*. In Proceedings of the HICSS-43 Conference, January 2010.

Dejin Z., Rosson M.B. (2009), *How and why people Twitter: the role that microblogging plays in informal communication at work*, GROUP '09. In Proceedings of the ACM 2009 international conference on Supporting group work, New York, NY, USA, pp. 243-252, URL: http://doi.acm.org/10.1145/1531674.1531710

Ebner M. (2009), Introducing Live Microblogging: How Single Presentations Can Be Enhanced by the Mass, *Journal of Research in Innovative Teaching* (JRIT), vol. 2, n. 1, pp. 91-100.

Ebner M., Reinhardt W. (2009), *Social networking in scientific conferences – Twitter as tool for strengthens a scientific community*. In Proceedings of the 1st International Workshop on Science 2.0 for TEL, Ectel 2009.

Ebner M., Mühlburger H., Schaffert S., Schiefner M., Reinhardt W. and Wheeler S. (2010), Get Granular on Twitter – Tweets from a Conference and their Limited Usefulness 1 for Non-Participants. In *Key competences in the knowledge society*, KCKS, pp. 102-113.

Haewoon K., Changhyun L., Hosung, P. and Moon S. (2010), *What is Twitter a Social Network or News Media?* In Proceedings of the 19th International World Wide Web (WWW) Conference, April 26-30, 2010, Raleigh NC (USA), URL: http://an.kaist.ac.kr/traces/WWW2010.html

Khan M.S., Ebner M., and Maurer H. (2009), *Trends discovery in the field of e-learning with visualization*. In Proceedings of 21st ED-Media Conference 2009, pp. 4408-4413.

Java A., Song X., Finin T. and Tseng B. (2007), *Why we twitter: understanding microblogging usage and communities*. In Proceedings of the 9th WebKDD and 1st SNA- KDD 2007 workshop on Web mining and social network analysis, ACM, pp. 56-65.

McFedries P. (2007), All A-Twitter, *IEEE Spectrum*, October 2007, p. 84.

Mendez X.O.G. and Duval E. (2009), *Who we are: Analysis of 10 years of the ed-media conference.* In Proceedings of 21st ED-Media Conference 2009, pp. 189-200.

Mühlburger H., Ebner M. and Taraghi B. (2010), @twitter Try out #Grabeeter to Export, Archive and Search Your Tweets. In *Research 2.0 approaches to TEL*, pp. 1-9.

Passant A., Hastrup T., Bojärs U. and Breslin J. (2008), *Microblogging: A Semantic Web and Distributed Approach*, 4th Workshop on Scripting for the Semantic Web (SFSW2008), Tenerife, Spain, 2 June 2008.

Reinhardt W., Ebner M., Beham G. and Costa C. (2009), How people are using Twitter during Conferences. In V. Hornung-Praehauser, M. Luckmann, (Eds.), *Creativity and Innovation Competencies on the Web*, Proceedings of the 5th EduMedia 2009, Salzburg, pp. 145-156.

Softic S., Ebner M., Mühlburger H., Altmann T. and Taraghi B. (2010), @*twitter Mining #Microblogs Using #Semantic Technologies*. In 6th Workshop on Semantic Web Applications and Perspectives, SWAP 2010, pp. 1-12.

Templeton M. (2008), *Microblogging defined*, URL: http://microblink.com/2008/11/11/