@twitter Try out #Grabeeter to Export, Archive and Search Your Tweets

Herbert Mühlburger[1], Martin Ebner[1], Behnam Taraghi[1],

[1] Graz University of Technology, Social Learning, Steyrergasse 30/I,
8010 Graz, Austria
{muehlburger, martin.ebner, b.taraghi}@tugraz.at

Abstract: The microblogging platform Twitter is beside Facebook the fastest growing social networking application of the last years. It is used in different ways, e.g. to enhance events (conferences) by sending updates, hyperlinks or other data as a news-stream to a broader public. Until now the stream ends with the end of the event. In this publication a new application is introduced that allows information retrieval and knowledge discovery by searching through local stored tweets related to a corresponding event. The architecture of the prototype is described as well as how the data is being accessed by a web application and a local client. It can be stated that making tweets available after the end of an event, enhances the way we deal with information in future.

Keywords: Knowledge discovery, information retrieval, Twitter, search, microblogging

1 Introduction

Twitter[1] and Facebook[2] are the fastest growing platforms of the last 12 months[3] [12]. On 22nd of February 2010 Twitter hits 50 million tweets per day[4]. Without any exaggeration it can be said that these two social networks are worth to be researched in detail [10] and are of interest for scientists and educators. After a period of testing first results emerge on this form of communication and interaction in science [7] as well as in the area of e-learning [3] [5] [9]. Although Twitter is widely known to be the most popular microblogging platform, a short introduction is given. Templeton [14] defined microblogging as a small-scale form of blogging, generally made up of short, succinct messages, used by both consumers and businesses to share news, post

[1] http://twitter.com (last access: 2010-04)

[2] http://facebook.com (last access: 2010-04)

[3] http://ibo.posterous.com/aktuelle-twitter-zahlen-als-info-grafik (last access: 2010-04)

[4] http://mashable.com/2010/02/22/twitter-50-million-tweets/ (last access: 2010-04)

status updates and carry on conversations. Due to the restriction to 140 characters it can also be compared with a short-message service that is based on an internet service platform. Maybe the factor of success of this application relies on its simplicity - users can send a post (tweet) that is listed on the top of their wall together with messages of their friends. Furthermore 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, thoughts) as well as fast and easy communication amongst more or less open communities [2]. In the same way Java [11] defined four main user behaviours why people are using Twitter - for daily chats, for conversation, for sharing information and for reporting news.

Taking a look at the usage of Twitter at conferences we notice the increase of reports, statements, announcements as well as fast conversation between participants. So called Twitter-walls nearby the projection of an ongoing presentation [4] or placed at any other location at the conference support the conference administration, organization, discussions or knowledge exchange. From this point of view microblogging becomes a valuable service reported by different publications [13].

One of the most recent studies on using Twitter at Web 2.0 conferences [1] examined tweets on a semantic basis [6]. The analysis showed that the idea of microblogging usage for distributing or explaining conference topics, discussions or results to a broader public seems to be limited. The authors pointed out that the use of Twitter during conferences should follow logics, like

- Usage as backchannel for conference participants
- Usage of document and illustrate connections
- Usage as a public notepad to collect relevant ideas, quotes or links
- Usage as evaluation tools

Basically there are two core issues - Twitter should be used first for communication between participants instantly and second for documentation on their own. Especially in case of documentation this will be only useful if users are able to create a kind of archive where they can store their tweets.

This publication deals with the research question, what can be the advantages of a web-based application that can also be used offline (without Internet connection) for information retrieval and knowledge discovery based on a micro-content system like Twitter.

"Grabeeter – Grab and Search Your Tweets" is the name of the application that has been developed in order to fulfil these requirements. The next chapter describes Grabeeter in more detail by giving an overview of the system's architecture and its particular features.

2 Architecture of Grabeeter

The architecture of Grabeeter (see Fig. 1) consists of two main parts. The first part is a web application that retrieves tweets and user information from Twitter through the Twitter API[5]. The second part of Grabeeter consists of a client application developed in “JavaFX[6] technology for accessing the stored information on a client side.

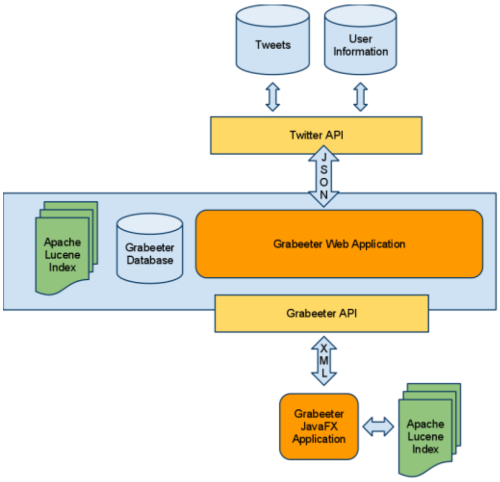


Fig. 1. **Architecture of Grabeeter**

As illustrated in Fig. 1 the Grabeeter web application implements the Twitter API in order to retrieve tweets of predefined users. The tweets are then stored in the Grabeeter database and on the file system as Apache Lucene[7] index. In order to ensure an efficient search the tweets must be indexed. The Grabeeter web application provides access to the Grabeeter database through its own REST style [8] API. This enables client applications to retrieve tweets and user information in an easy way by implementing this API. In difference to the Twitter API Grabeeter API provides all stored tweets and makes no restriction over time.

The Grabeeter client application is developed using JavaFX in order to be independent from different operating systems as well as to provide an easy process to

[5] http://apiwiki.twitter.com/ (last access: 2010-04-21)

[6] http://www.sun.com/software/javafx/ (last access: 2010-04-16)

[7] http://lucene.apache.org/java/docs/ (last access: 2010-04-21)

upgrade the client application using Java Web Start[8]. Furthermore it provides an easy way to store the retrieved tweets on the user's local file system for later offline processing. The following sections describe the different parts of Grabeeter in detail.

2.1 Grabeeter Web Application

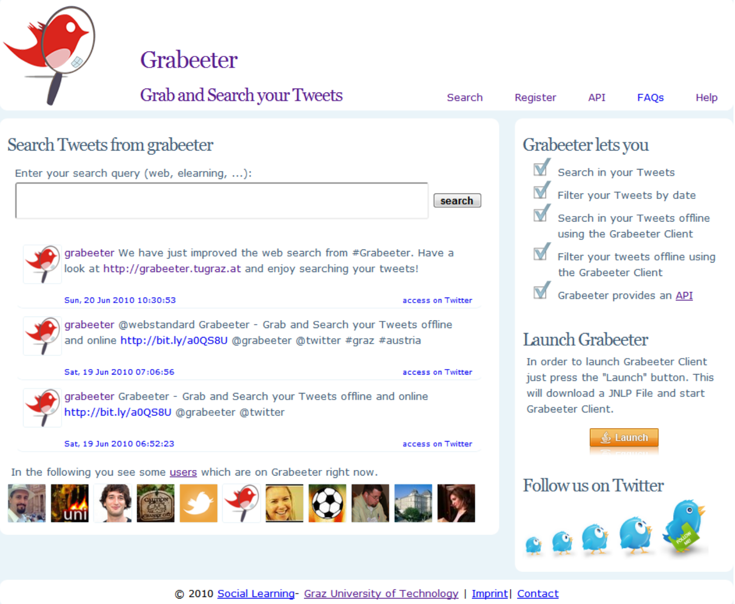
The Grabeeter web application enables users to archive their tweets in the Grabeeter database and to perform a search on the stored tweets through a web interface. The tweets are not only stored in the database but also indexed by Apache Lucene in order to support an efficient search on the tweets. These tweets can be accessed then by client applications through the Grabeeter REST style API[9].

As illustrated in Fig. 2 users are able to carry out a search on the stored tweets online or launch the Grabeeter JavaFX Client application by pushing the "Launch" button and search their tweets using the client application.

The workflow of the Grabeeter web application is as follows: At first users register their Twitter usernames at the Grabeeter web application. These usernames are stored in a text file which is parsed later by a cron job. The cron job runs a PHP script that retrieves all accessible tweets for the given usernames. Later another cron job updates the tweets for all monitored users on a scheduled timetable.

[8] http://java.sun.com/javase/6/docs/technotes/guides/javaws/index.html (last access: 2010-04-21)

[9] http://grabeeter.tugraz.at/developers



Grabeeter

Grab and Search your Tweets

Search Tweets from grabeeter

Enter your search query (web, elearning, ...):

grabeeter We have just improved the web search from #Grabeeter. Have a look at http://grabeeter.tugraz.at and enjoy searching your tweets!

Sun, 20 Jun 2010 10:30:53 access on Twitter

grabeeter @webstandard Grabeeter - Grab and Search your Tweets offline and online http://bit.ly/a0QS8U @grabeeter @twitter #graz #austria

Sat, 19 Jun 2010 07:06:56 access on Twitter

grabeeter Grabeeter - Grab and Search your Tweets offline and online http://bit.ly/a0QS8U @grabeeter @twitter

Sat, 19 Jun 2010 06:52:23 access on Twitter

In the following you see some users which are on Grabeeter right now.

Grabeeter lets you

- ✓ Search in your Tweets
- ✓ Filter your Tweets by date
- ✓ Search in your Tweets offline using the Grabeeter Client
- ✓ Filter your tweets offline using the Grabeeter Client
- ✓ Grabeeter provides an API

Launch Grabeeter

In order to launch Grabeeter Client just press the "Launch" button. This will download a JNLP File and start Grabeeter Client.

Follow us on Twitter

© 2010 Social Learning- Graz University of Technology | Imprint| Contact

Fig. 2. **Grabeeter Web Application**

Due to Twitter's REST API Limit[10] it is only possible to access the latest 3200 tweets (statuses) via the API of a given user on Twitter. So in case a user has less than 3200 tweets on Twitter at the time of registration on Grabeeter all of the user's tweets are archived. From that time on all future tweets are stored and the entire first (3200 or less) tweets remain accessible and searchable too. In that way all tweets of a user ever become saved and searchable. If a user has more than 3200 tweets on Twitter at the time of registration on Grabeeter it is only possible to retrieve the latest 3200 tweets of this user from Twitter due to the Twitter limit. But from that time on all of the future tweets are archived and searchable through Grabeeter.

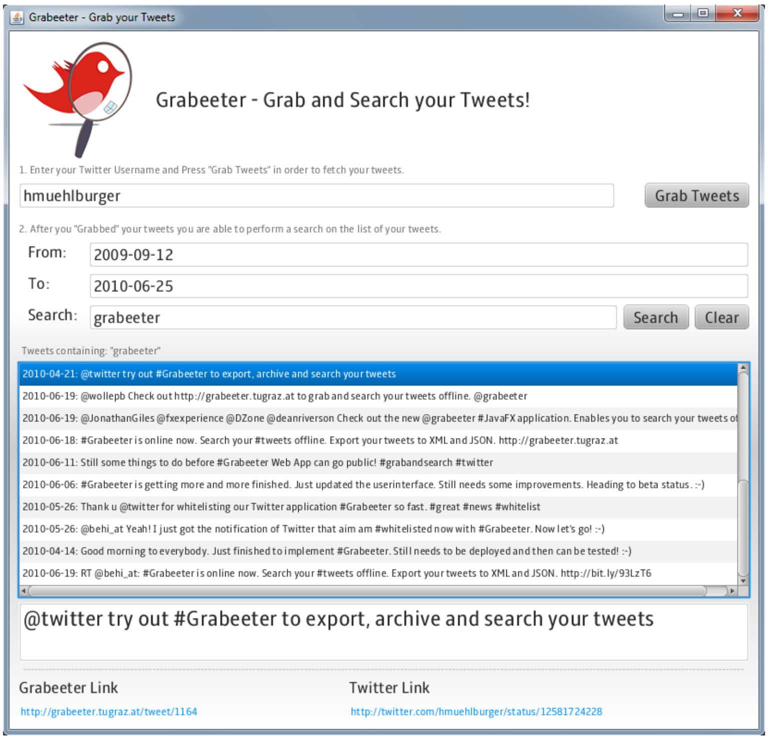
Later processing of the stored tweets enables us to achieve more enriched data sets by adding different kind of metadata to the stored information. However this step is not yet implemented and is described in more detail in section 4 regarding future work.

[10] http://apiwiki.twitter.com/Things-Every-Developer-Should-Know#6Therearepaginationlimits (last access: 2010-04-21)

2.2 Grabeeter Client Application

The Grabeeter client application was developed using JavaFX technology. It was tested on different operating systems such as Windows XP, Ubuntu Linux 10.04 and MacOS X running the latest Java SE Runtime Environment.

In order to start the Grabeeter Client application the user clicks the “Launch” button provided on the Grabeeter website (see Fig. 2). While the application starts a shortcut is created on the user’s local desktop. Through this shortcut the user is able to restart the application later on instead of using a browser.



Grabeeter - Grab your Tweets
Grabeeter - Grab and Search your Tweets!
Grab Tweets
Search
Clear

1. Enter your Twitter Username and Press "Grab Tweets" in order to fetch your tweets.

hmuehlburger

2. After you "Grabbed" your tweets you are able to perform a search on the list of your tweets.

From: 2009-09-12

To: 2010-06-25

Search: grabeeter

Tweets containing: "grabeeter"

2010-04-21: @twitter try out #Grabeeter to export, archive and search your tweets
2010-06-19: @wollepb Check out http://grabeeter.tugraz.at to grab and search your tweets offline. @grabeeter
2010-06-19: @JonathanGiles @fxexperience @DZone @deanriverson Check out the new @grabeeter #JavaFX application. Enables you to search your tweets of
2010-06-18: #Grabeeter is online now. Search your #tweets offline. Export your tweets to XML and JSON. http://grabeeter.tugraz.at
2010-06-11: Still some things to do before #Grabeeter Web App can go public! #grabandsearch #twitter
2010-06-06: #Grabeeter is getting more and more finished. Just updated the userinterface. Still needs some improvements. Heading to beta status. :-)
2010-05-26: Thank u @twitter for whitelisting our Twitter application #Grabeeter so fast. #great #news #whitelist
2010-05-26: @behi_at Yeah! I just got the notification of Twitter that aim am #whitelisted now with #Grabeeter. Now let's go! :-)
2010-04-14: Good morning to everybody. Just finished to implement #Grabeeter. Still needs to be deployed and then can be tested! :-)
2010-06-19: RT @behi_at: #Grabeeter is online now. Search your #tweets offline. Export your tweets to XML and JSON. http://bit.ly/93LzT6

@twitter try out #Grabeeter to export, archive and search your tweets

Grabeeter Link
http://grabeeter.tugraz.at/tweet/1164

Twitter Link
http://twitter.com/hmuehlburger/status/12581724228

Fig. 3**. Grabeeter Client Application**

The user provides a Twitter username to the client (see Fig. 3) and starts the grabbing of tweets by clicking the button “Grab Tweets”. In order to initially grab its tweets the user has to have an internet connection. The Grabeeter Client application then connects to the Grabeeter Database through the Grabeeter API in order to retrieve the tweets. The retrieved tweets are then stored on the local file system in a structured XML format. This enables other applications to access the locally stored tweets for their own purposes.

The Grabeeter application then loads the locally stored tweets and creates an in memory Apache Lucene index. Users are then able to perform a full text search and filter their tweets by specifying a time period.

Initially the Grabeeter Client application works in online mode in order to retrieve and store the entire recent tweets using the Grabeeter API. After restarting the application the locally stored tweets are loaded and indexed again. Therefore users are able to perform searches on tweets without having internet connection and so being independent from web services.

3 Discussion

The following lists interesting aspects that occurred during the development of Grabeeter using JavaFX and the Twitter API.

„Drag-To-Install“: One very utile feature of JavaFX is the „Drag-To-Install“ possibility. It is the ability for an application to be dragged-out of the browser window and being “installed” on the operating system by dropping it onto the operating system’s desktop. The term “installed” means here that a shortcut is created on the desktop and that the JavaFX application is added to the Java application cache on the corresponding operating system. This feature seems not to function properly on MacOS systems so far. From this point of view a new version of the client can be updated in the background without knowledge of the user.

Twitter API restrictions: As already mentioned Twitter REST API requests are restricted to the latest 3200 tweets of a user. There is no chance for any application to access the first tweets, in case the user has already more than 3200 tweets.

Twitter capacity problem: Sometimes the Twitter API is over capacity. In this case no data can be retrieved from the API. This might delay the archive process in Grabeeter web application.

Beside these restrictions Grabeeter may have an interesting effect on the change of writing style: Due to the fact that the suggested tool is able to retrieve data the user is able to document his/her experiences from an event over a time period. This leads to reassess about how we have to use microblogs in general and how we have to write our tweets in order to regain relevant data. Overall this means tweets are written primarily for users themselves and not for a broader public which is a very new aspect to the basic intention of Twitter. With the help of the tool it is now possible to retrieve all tweets concerning a specific hashtag (e.g. event) within a clear defined time frame. Any collected hyperlink can be reused by searching for the specified event and clicking on the appropriate tweet.

If users register on Grabeeter before they reach 3200 tweets on Twitter it is possible to archive and retrieve all tweets from these users. For Grabeeter performs incremental updates and stores all tweets in its archive all tweets of a user are stored continuously from the beginning up to future tweets.

According to our research question in the beginning we like to point out the advantages of the tool Grabeeter:

- Micro-content (tweets) is achievable due to the fact that any tweet can be retrieved at anytime from a local hard-drive
- Micro-content is storable in a way that the user can distinguish between different events
- Micro-content is searchable along keywords, hashtags, time frames as well as different entities (URLs, @, …)

From a technical point of view update process is easily and independence of devices and operating systems is guaranteed.

4 Conclusion and Future Work

Grabeeter was launched in May 2010. The web application as well as the JavaFX client can be accessed at http://grabeeter.tugraz.at.

The rapid improvements in the mobile technology have led to an ascending trend of using mobile applications in recent years. Consequently more users use mobile devices to access online applications. It is planned to build the Grabeeter client as a mobile application for different platforms (Android, iPhone, JavaFX devices …). The adaptations that must be performed are mainly the view adjustment and an appropriate look and feel for the mobile environments.

The next main extension of Grabeeter will be the capability not only to retrieve the search results of a simple search query to the user, but also to combine multiple search queries over multiple users for the analysis of the archived data sets, for data exploration and a better knowledge discovery. Use of semantic technologies and interlinking techniques for this purpose would definitely enrich the data sets and enhance the usefulness of stored tweets. The first step will be to describe the archived data sets semantically, to “triplify” the data sets and convert them to RDF triples by applying the existing vocabularies used for microblogging.

The tweets of each user can be extracted and analysed towards relevant keywords to get a feeling about the main topics for e.g. a specific event. The text fragments in tweets can be extracted and interlinked with resources in the Linked Open Data[11] (LOD) cloud such as DBpedia, Flickr, Geo-names, etc. The Twitter users can be interlinked with FOAF profiles in the LOD cloud too. Having the data sets triplified and interlinked with LOD it will be more efficient to analyse the collected data from the Twitter API. It will become possible to perform a more accurate knowledge

[11] http://richard.cyganiak.de/2007/10/lod/ (last access: 2010-04-16)

discovery and retrieve search results not only within tweets gained from the Twitter API, but also in interlinked resources of the World Wide Web.

Furthermore a SPARQL[12] endpoint can be provided in Grabeeter web application to let different monitoring and analysing client applications to perform SPARQL queries over semantic data sets. As an example searching for tweets containing a geographic term such as “Vienna” would return also the tweets that contain the term “Wien”, which is the German word for Vienna. Search queries can be made even much more complex:

- Get tweets that contain links to photos related to the place where conference xy takes place.
- Get tweets that are related to informatics and semantic technologies.

It can be summarized that the described application allows retrieving status updates from the most famous microblogging platform Twitter for information retrieval on a local hard drive. Furthermore through the combination of tweets from different Twitter users with predefined keywords or hashtags the knowledge discovery seems to be opened up in a new dimension. For the first time the documentation of events by just simply tweeting of statements, hyperlinks or media files becomes possible. Grabeeter is built to enhance the usefulness of microblogging on conferences and allows retrieving data that was produced just on the fly.

References

1. Bernhardt, T., Kirchner, M.: Web 2.0 meets conference – the EduCamp as a new format of participation and exchange in the world of education. In: Ebner, Martin / Schiefner, Mandy (Eds.): Looking Toward the Future of Technology-Enhanced Education: Ubiquitous Learning and the Digital Native. IGI Global, Hershey, 2009
2. Boyd, d., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In Proceedings of the HICSS-43 Conference, January 2010.
3. Costa, C., Beham, G., Reinhardt, W., Sillaots, M.: Microblogging in Technology Enhanced Learning: A Use-Case Inspection of PPE Summer School 2008. In: Proceedings of the 2nd SIRTEL workshop on Social Information Retrieval for Technology Enhanced Learning, 2008
4. Ebner, M.: Introducing Live Microblogging: How Single Presentations Can Be Enhanced by the Mass. Journal of Research in Innovative Teaching (JRIT), 2 (1), p. 91- 100, 2009
5. Ebner, M., Lienhardt, C., Rohs, M., Meyer, I.: Microblogs in Higher Education – a chance to facilitate informal and process oriented learning?. Computers & Education, ISSN 0360-1315, DOI: 10.1016/j.compedu.2009.12.006., 2010

[12] http://www.w3.org/TR/rdf-sparql-query/ (last access: 2010-04-21)

6. Ebner, M., Mühlburger, H., Schaffert, S., Schiefner, M., Reinhardt, W.: Get Granular on Twitter - Tweets from a Conference and their limited Usefulness for Non-Participants, accepted paper at Workshop (MicroECoP). WCC 2010 conference, Brisbane, Australia, 2010

7. Ebner, M., Reinhardt, W.: 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, September 2009

8. Fielding R.: Architectural Styles and the Design of Network-based Software Architectures. 2000, http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm (last access: 2009-01)

9. Grosseck, G., Holotescu, C.: Can we use twitter for educational activities?. In Proceedings of the 4th International Scientific Conference eLSE ”eLearning and Software for Education”, April 2008

10. Haewoon K., Changhyun, L., Hosung, P. Moon, S.: What is Twitter a Social Network or a News Media?. Proceedings of the 19th International World Wide Web (WWW) Conference, April 26-30, 2010, Raleigh NC (USA), April 2010, http://an.kaist.ac.kr/traces/WWW2010.html (last access: 2010-04)

11. Java, A., Song, X., Finin, T., Tseng, B.: 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, pages 56– 65. ACM, 2007.

12. McGiboney, M.: Keep on tweet’n. March 2009, http://www.nielsen-online.com/blog/2009/03/20/keep-on-tweetn/ (last access: 2010-04)

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

14. Templeton, M.: Microblogging defined, http://microblink.com/2008/11/11/microblogging-defined/, 2008, (last access: 2010-04)