Real Time Discussion Retrieval from Twitter

Dmitrijs Milajevs
University of Groningen
P.O. box 716, 9700 AS
Groningen, The Netherlands
dimazest@gmail.com

Gosse Bouma
University of Groningen
P.O. box 716, 9700 AS
Groningen, The Netherlands
g.bouma@rug.nl

ABSTRACT

While social media receive a lot of attention from the scientific community in general, there is little work on high recall retrieval of messages relevant to a discussion. Hash tag based search is widely used for data retrieval from social media. This work shows limitations of this approach, because the majority of the relevant messages do not even contain any hash tag, and unpredictable hash tags are used as the conversation evolves in time. To overcome these limitations, we propose an alternative retrieval method. Given an input stream of messages as an example of the discussion, our method extracts the most relevant words from it and queries the social network for more messages with these words. Our method filters messages that do not belong to the discussion using an LDA topic model. We demonstrate this concept on manually built collections of tweets about major sport and music events.

Categories and Subject Descriptors

H.3 [**Information Storage And Retrieval**]: Information Search and Retrieval

Keywords

social media; discussion retrieval; streaming event data

1. INTRODUCTION

Microblogs capture discussions that are interactions between people about diverse, yet related topics. Microblog services provide a way to retrieve a topic (for example by hash tag), but lack the ability to provide all the tweets of the discussion.

This work proposes a method for discussion retrieval based on keyword search. Since microblog services make it difficult and time consuming to retrieve past data, we concentrate on streaming data collection, which has less limitations.

The paper is organized as follows. Section 2 makes an overview of the related work. Section 3 shows how a hash tag changes its meaning in time and how a discussion explodes beyond the expected boundaries by example of the hash tag `#pp11` and tweets about two music festivals in 2011. Section 4 gives an overview of our discussion retrieval method focusing on the unpredictability of a discussion and keyword ambiguity. Given a stream of tweets that are highly associated with the discussion, our method identifies the trending words and queries for tweets that contain them. Irrelevant tweets are filtered using an LDA topic model.

Section 5 evaluates the proposed method on data sets about major music and sport events. Finally, section 6 concludes the work.

The two main contributions of this work are the observation that hash tags are not effective for advanced tweet retrieval and a two stage retrieval method for streaming message data that explicitly decouples high-recall retrieval and precision oriented filtering.

2. RELATED WORK

For our proposed method for discussion retrieval, two steps are crucial. First, given a dynamic, core stream of tweets, we need to detect keywords that characterize the stream. For tweets, this problem is similar to that of detecting trending topics. Second, given a set of keywords, we need to select relevant tweets from a stream. This problem is similar to ad hoc document retrieval, although it should be noted that we assume a more dynamic perspective than is usually assumed, as both the keywords and the stream are dynamic and evolving over time.

Trends are the core notion for many Twitter based applications. [2] define a trend as *a word or phrase that is experiencing an increase in usage, both in relation to its long–term usage and in relation to the usage of other words.* [12] use Hodrick-Prescott Trend Filtering to discover trends. The authors estimate trend components of topics and then figure out whether a topic is emerging by introducing a margin based loss function which penalizes static or decaying topics. [9] are looking for trending words that co–occur with a product of interest. According to them, an interesting phrase *should both be mentioned frequently and should be relatively unique to the product with which it is associated.* The significance score of phrase is the ratio of phrase co-occurrence with product to phrase occurrence raised to the power 0.96. This approach requires the global phrase occurrence frequency, which is not available in our setup.

The work in [10] is a study of a compact representation of a stream of tweets as a timeline. The authors assume that events are tightly associated with peaks in user activity, so it makes sense to detect the peaks, and then to label them with few descriptive words. Once a peak is detected, the terms of the tweets in the peak are ranked using TF-IDF, the top ranked terms label the peak. Evaluation on a collection of

WWW 2013 Companion, May 13–17, 2013, Rio de Janeiro, Brazil.
ACM 978-1-4503-2038-2/13/05.

tweets shows that the system captures the major events, however, it did not recognize minor events.

In [11], the authors describe a method of trend discovery based on burst detection. The method is based on three values. f is the observed frequency of a term in a current window of length n. μ is the mean of the frequencies of the term over the windows of length n. σ is the standard deviation of the term frequencies over windows of length n. Given these three values the trending score of the terms in the current window is:

$$\frac{f-\mu}{\sigma} \tag{1}$$

In the streaming data processing scenario, Formula (1) is very effective to implement because the mean and the standard deviation can be computed without iteration over the frequencies of a term in every window, but updated as a new window is finished.

We adopt the formula in (1) for predicting a dynamically evolving set of keywords for a given core stream.

The TREC microblog track[1] studies how recent tweets can be retrieved that are relevant to a given topic.

A number of systems assign special meaning to hashtags, for instance by searching for relevant hashtags given a topic, and then using the (segmented) hashtags as query [8], for for query-expansion, or for results re-ranking [3, 1]. In the following section, we argue that while hashtags are useful for retrieving a core stream of tweets related to an event, they are not sufficient for retrieving all relevant tweets, as not all tweets contain hashtags and, also, the distribution and the meaning of the hashtags may change over time. The changing meaning of hashtags resembles the dynamic nature of web search queries observed in [7].

We model the event of interest using an LDA model built over a core stream of tweets. The microblog track takes as starting point a query, and not a (manually developed) core stream, as we do. Nevertheless, some systems in the microblog track apply techniques that acquire similar topic profiles. A common approach is to exploit links in relevant tweets to obtain additional text for topic modelling.[14] expand queries with links to external resources such as Wikipedia to obtain topic profiles. [6] achieve promising results using, among others, a (labeled) LDA model for semi-supervised twitter topic categorization.

Note that while the microblog track includes a notion of recency, it does not take into account the fact that a stream of data needs to be searched. In our setting, topics evolve over time, and also, the twitter stream changes over time. As a consequence, the frequency of terms, IDF scores, etc., cannot be computed off-line. We present a method that takes into account the dynamic nature of the discussion related to an event (characterized as a core stream of tweets) as well as the document stream from which relevant tweets have to be selected.

3. HASHTAG PROPERTIES

This case study explores hash tag properties by example of the hash tag `#pp11` and tweets about two music festivals — Dutch Pinkpop and Belgian Pukkelpop.

We base our study on a tweet collection built by the University of Groningen [13]. The collection consists of tweets with typical Dutch words, so the tweets are mainly in the Dutch language and cover a broad variety of topics.

From the whole collection we filtered out the tweets that contain *pinkpop*, *pukkelpop* or `#pp11`.

Figure 1 shows tweet frequency per hour for the words of interest.

The Pinkpop festival takes place in Landgraaf every year. Pinkpop 2011 took place from the 11th to the 13th of June. Figure 1 reflects it by the frequency peak of the word *Pinkpop*. 26% of tweets with the word *Pinkpop* do not contain any hash tag.

The analysis of the time line throughout 2011 shows that the Pinkpop Twitter community reacts to the news stories. Every time the festival lineup was announced or extended the frequency of tweets rose.

Pukkelpop 2011 was planned to take place from the 18th to the 20th of August. As Figure 1 shows, during the festival, the word *pukkelpop* experienced increase of usage. 37% of selected tweets do not contain any hash tag.

Unfortunately, the festival was stopped due to a severe storm. Twitter was heavily used to organize help. The Twitter user `@daHawkeyeCaller` gives an overview[2] of how Twitter was used during the storm. According to him, the main challenge in using Twitter to help others was to unclutter the social stream from emotional and already known posts and access valuable information.

The conversation went beyond the official and expected hash tag `#pkp11`. For example, `#hasselthelpt`, a special hash tag consisting of Hasselt (the name of the closest city to the festival) and *helps*, was used for asking or providing help. `@WimLuyckx` suggested to use `#hasselthelptmooi`[3] for expressing sentiments and do not make it difficult to retrieve the tweets that ask for help. `#ppok` was used to communicate that the author of a tweet is fine.

Further examination of the retrieved tweets in 2011 shows the following.

Twitter activity correlates with media news. In spring, the peaks of festival names frequency are triggered by the news stories about the upcoming events. Line up announcements boost user activity on Twitter.

During an event Twitter users produce the most content in comparison to the time before and after the event.

Hash tags are ambiguous. Depending on a tweet's content and time, a hash tag may refer to various entities. `#pp11` in June is most probably about Pinkpop, while in August it is about Pukkelpop.

Conversation goes beyond expected hash tags. At Pukkelpop, people were generating the content by asking or offering help, instead of discussing the news. Consequently, new hash tags appeared in the conversation. `#hasselthelpt`, `#ppok` and `#ppshelter` are among them.

A hash tag based retrieval method is not sufficient for discussion retrieval. More than a quarter of captured tweets did not contain any hash tag making the recall low. Thus more elaborated linguistic analysis is required for a robust retrieval.

4. DISCUSSION RETRIEVAL

This section explains how a discussion is retrieved from social media. In general, a discussion retrieval system should

[1] http://trec.nist.gov/

[2] http://dahawkeyecaller.tumblr.com/post/9111077769

[3] The tag is translated to English as *Hasselt helps beautifully*.

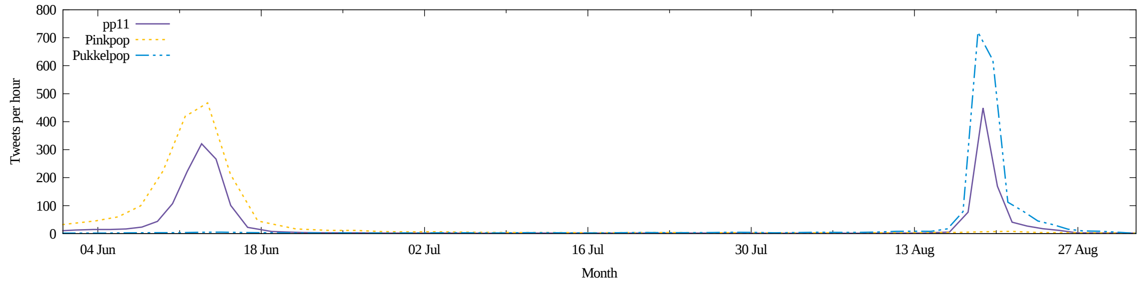

Figure 1: Frequencies of the words *Pinkpop*, *Pukkelpop* and the hashtags `#pp11` **in the Summer 2011.**

be able to identify relevant messages in a global stream of messages.

Discussion retrieval is a challenging task. The tweets about Pukkelpop show that it is almost impossible to predict what keywords and phrases will be used most. However, a part of the discussion can be retrieved easily. This part is made up of messages that contain highly related words or are written by known users. We refer to these messages as the core stream of a discussion.

For example, to collect the tweets about Lowlands, another Dutch music festival, the core stream could contain the tweets with the word *lowlands* or tweets written by `@Lowlands_12`, the festival's official user account screen name in 2012. Note that the included words should be chosen carefully, so `#ll` should not be included, because then the core stream would also contain the tweets about the Spanish football league (in Spanish it is referred as *La Liga*, thus the hash tag `#ll` is used). However, it does not matter how the core stream is retrieved, ultimately it can be manually built.

4.1 Methodology

The Twitter Streaming API offers two ways of data collection[4].

The **GET statuses/sample** API resource provides approximately 1% of all public tweets. This resource might be useful for exploring global trends, though is not promising for analysis of smaller local events, which most probably are not included in the sample.

The **POST statuses/filter** entry point returns public statuses that match one or more filtering predicates. There are three types of filtering predicates: follow, track and locations. They allow one to get tweets from certain users, tweets with certain words and tweets from certain locations. The entry point fits the discussion retrieval task, since it can be queried for tweets with desired properties.

To retrieve the rest of the discussion the system has to query `POST statuses/filter` entry point with some predicates. The predicates could be the words that are trending in the core stream. However, it is not guaranteed that all incoming tweets will be relevant. Thus after consuming more tweets, the input stream can be split to three sub streams. **The core stream** is a stream of tweets that is guaranteed to be a part of the discussion. **The noisy stream** contains the tweets that come from the entry point. **The discussion stream** is made of the tweets in the core stream and the relevant tweets from the noisy stream.

[4] https://dev.twitter.com/docs/streaming-apis/streams/public

Definition 1. Let $P = \{p_1, \ldots, p_n\}$ be a set of n distinct filtering predicates. Then $S(P) = t_1, \ldots, t_m$ is a filtered stream of m tweets, such that every tweet in the stream satisfies at least one filtering predicate in P.

Definition 2. Given a core stream of tweets S and a filtered stream $S(P)$, the sequence of tweets $D(S, S(P))$ is the discussion stream of S with $S(P)$ which consists of tweets that are relevant to the discussion that is described by S.

In the case of collecting Lowlands tweets, the set up could be the following: S would be the core stream and it would contain the tweets with the word *lowlands* and the tweets that are written by the user `@Lowlands_12`. The predicates for the noisy stream could be $P = \{\texttt{wombats}, \#\texttt{ll}\}$, because it is known that The Wombats performed at the festival and `#ll` is used to mark Lowlands related tweets. The discussion stream would consist of the tweets in S and some tweets in $S(\{\texttt{wombats}, \#\texttt{ll}\})$.

For successful discussion retrieval, the system has to be able to *a*) decide whether a tweet from the noisy stream is relevant and should be included to the discussion stream. In addition, since in the beginning P is empty the system has to *b*) update P with words that relevant tweets are likely to contain.

Discovering potential keywords.

Suppose we are given a core stream S and we need to provide a list of keywords P that the relevant tweets are likely to contain.

The discussion in the stream S is going to evolve through time: new topics will appear, other topics will not be discussed anymore. The core stream will mirror this development through the change of word frequencies. Emergence of a new topic will give rise to the usage of certain words, making them trending.

We follow the proposal in Formula (1) to calculate the mean of a term in an effective way without iterating over occurrences of it in each window, the system stores the total number of occurrences of the term and the number of windows it occured in. For the standard deviation, the sum of frequency squares has to be stored in addition to the information about the mean.

Including a tweet to a discussion.

The task is to distinguish relevant tweets from irrelevant tweets regarding a discussion. The system needs to decide whether a tweet $t \in S(P)$ in the noisy stream belongs to the discussion $D(S, S(P))$.

A relevant tweet shares one or more topics with the discussion of interest. To induce topics in the core stream, topic models in general [5] and LDA in particular can be applied.

For every tweet in the noisy stream, its topic distribution is predicted. So, the model returns a probability distribution $p_1, \ldots, p_n$ where n is the number of topics and for $0 < i \leq n$: p_i is the probability that the tweet belongs to the ith topic. Then the entropy of the prediction is:

$$-\sum_{i=1}^{n} p_i \log p_i$$

If the model is confused, it will have a uniform prediction with high entropy. If the model is confident, it will return a skewed distribution with low entropy. Then the decision of whether a tweet belongs to the discussion can be based on the entropy: if it less than a predefined threshold, then the tweet belongs to the discussion.

5. EVALUATION

The system consumes two input streams — the core stream of only relevant tweets, and the noisy stream, which contains both relevant and irrelevant messages. The streams are disjoint, there is no message that belongs to both streams simultaneously.

However, to evaluate the system, we need to know what tweets are contained in the discussion stream. To build the discussion stream, we collected the tweets that contain a word which is related to the discussion. For example, to retrieve a discussion about a music festival, its name is used.

We extracted the core stream from the discussion. The core stream tweets mention or a written by some Twitter user, who is known to be related to the discussion. For a music festival it might be an official Twitter account of the festival. We refer to the relevant tweets which are not in the core stream as **the unseen stream**.

We formed the noise by the tweets that belong to a similar discussion. For example, another music festival which happens at the same time with the festival of interest.

5.1 Metrics

Precision and recall are widely adopted measures for document retrieval tasks.

Definition 3. Let A be the set of relevant documents and B be the set of retrieved documents. The precision is $p = \frac{|A| \cap |B|}{|B|}$. The recall is $r = \frac{|A| \cap |B|}{|A|}$.

Since the system includes two stages — filtering and classification — it is worth measuring the performance of each stage. Filtering aims to retrieve as many tweets as possible maximizing recall. The second stage, classification, filters out unrelated tweets and maximizes precision. Thus precision and recall were measured after filtering and classification. The performance of the classifier is measured in isolation, the set of relevant documetns is based on the tweets that are returned by filtering.

In the example on Figure 2, there are three tweets in the core stream, and two tweets in the unseen stream. Together they make up the discussion stream of five tweets. The tweets "1", "2" and "3" do not belong to the discussion and are considered to be noise.

The filtering step retrieves four tweets, but only the tweet "b" is retrieved correctly. The tweets "1", "2" and "3" are from the noise stream and should not be retrieved.

The tweet "b" was classified correctly, the tweets "1" and "2" were classified incorrectly.

Precision and recall of filtering (p_f, r_f), classification (p_c, r_c) and the whole system (p, r) are:

$$p_f = \frac{1}{4} = 0.25 \qquad p_c = \frac{1}{3} = 0.3 \qquad p = \frac{4}{5} = 0.8$$
$$r_f = \frac{1}{2} = 0.5 \qquad r_c = \frac{1}{1} = 1 \qquad r = \frac{4}{6} = 0.67$$

5.2 Data sets

We prepared several datasets with various features to see how the system performs in different situations.

5.2.1 Primavera Sound and Rock am Ring

This dataset tests how the system distinguishes two unrelated discussions. For this task, tweets about two rock festivals were collected: German "Rock am Ring", and Spanish "Primavera Sound". Both events took place in the first weekend of June 2012. The tweets were collected from March to September in 2012. Event discussions share the same topic — music. The fact that the festivals happened in different countries, and different artists performed there, makes the discussions unrelated.

To be a part of the Primavera Sound discussion, a tweet has to contain the phrase *primavera sound*; or has to be written by or mention the user `@Primavera_sound`, there are 35,644 such tweets. To belong to the Rock am Ring discussion, a tweet has to contain the phrase *rock am ring*; or be written by or mention `@rockamringblog`. 63,124 tweets formed the Rock am Ring discussion.

The Primavera Sound core consists of 626 tweets that mention or are written by `@Primavera_sound`. The Rock am Ring core is built of 331 tweets that are written by or mention `@rockamringblog`.

5.2.2 Pentathlon

When the discussion of interest is a part of some more global and diverse discussion, it might be the case that the overgeneralization is not desirable. In this experiment, the discussion of interest is about the modern pentathlon competition at the Olympic games on August 11–12 2012.

The core stream contains the tweets that mention or are created by `@UIPM_HQ`, the official Twitter account of Union International de Pentathlon Moderne. There are 1,078 such tweets. 22,594 tweets that contain the words *pentathlon* or *modpen* build the rest of the discussion.

1,996,422 tweets that contain the words *Olympic* (*Olympics* is matched by the Twitter API in this case as well), and *london2012*, but do not belong to the discussion are noise.

5.2.3 Pinkpop, Pukkelpop and #pp11

The case study in Section 3 showed meaning shift for the hash tag `#pp11`. This data set contain the tweets from the Summer 2011 with the words *Pinkpop*, *Pukkelpop* and the hash tag `#pp11` and targets the evaluation of the system adoption to the dynamic change of hash tag meaning.

We sampled 1,486 tweets with the word *Pinkpop* and without the hash tag `#pp11` from the all tweets we got from 2011.

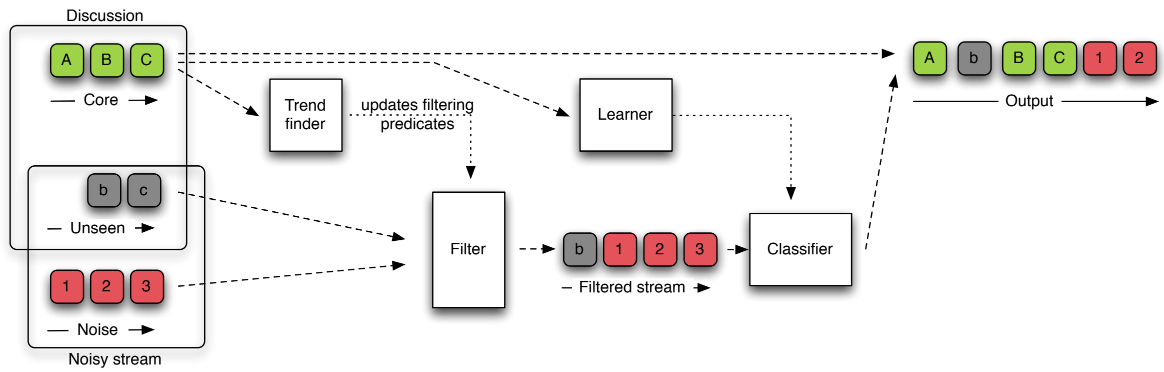

Figure 2: Experiment setup. Rounded rectangles represent tweets. Tweets marked with a capital letter are in the core stream. Tweets marked with a lower case letter are in the unseen stream. Tweets labeled by a number are noise. Rectangles are components of the system. Dashed arrows are tweet flows, so the core stream is consumed by the trend finder and the model learning component. Dotted arrows represent other information exchange.

The unseen stream is made up of 53,100 tweets which contain both *Pinkpop* and `#pp11`; or contain just `#pp11` and are written before July 1st 2011.

Finally, the noise are the 99,390 tweets that are not included in the streams above.

5.3 Baseline

We compared our method to an open source information retrieval system Whoosh[5].

Each dataset was indexed separately. An indexed document includes the text of a tweet appended with its author screen name. Tweet's ID is stored in a separate field for the evaluation purposes.

For each experiment the index is queried with the terms that define the core stream. Precision and recall is based on the returned IDs.

5.4 Results

We compared our method with the baseline on the data sets described above.

The system managed to achieve high precision (0.83) and recall (0.85) scores for the Primavera Sound task outperforming the baseline (see Table 1 for the complete results). Filtering got all the relevant tweets, while the classification step improved the precision. The baseline retrieved the tweets with perfect precision, but with low recall.

The filtering step did not perform well on the Rock am Ring data set. Filtering recall is only 0.08. This can be explained by the fact that there are many tweets that someone is listening to a song, or watching a video from previous Rock am Ring festivals. This makes the stream rather monotonous, because the number of songs people mention in their tweets is rather limited. However, classification precision is at the same level as on the Primavera sound data. Again, the baseline gave the perfect precision, but low recall.

The system performed less well on the Pentathlon data. About 20% of the relevant tweets were retrieved at the filtering step, which is better than baseline's recall of 0.04. The classification step could not improve filtering precision yielding very low performance. The reason is that the Pentathlon discussion is a tiny part of the Olympics discussion, while the core (the tweets that mention or are written by `@UIPM_HQ`) is even smaller. For this scenario, usage of keyword based search is suggested, since the expansion of the topic is not desired. The online classifier could be replaced to an offline version to improve on classification precision and recall.

For the `#pp11` data set, the filtering step gave very robust output, but the classification did not yield competitive results. The baseline could not recognize the meaning shift of `#p11` giving very low precision. On this data set our system F-score is 0.05, while the baseline gave the F-score of 0.03.

6. CONCLUSION

At the beginning of the Games of the XXX Olympiad in London, spectators are believed to disturb television coverage of a cycling road race[6]. People tweeted so much that the mobile network could not handle the load. In the vast amount of produced tweets, there is a certain number of discussions that mirror the event and reveal spectators' attitude to it, which is very fruitful to analyze. However, the standard search functionality is rather limited and does not allow for building the complete XXX Olympics tweet collection.

One of the reasons the standard approach is that the new hash tags pop up as the discussion evolves and that the standard approach is concentrated on high precision, rather than the result completeness (recall).

This work proposes a discussion retrieval method that is based on keyword search and aims to collect as many relevant tweets as possible. The task is split into two sub-tasks: trending word identification to query the Twitter streaming API for most likely relevant tweets, and the tweet classification stage that determines whether incoming tweets belong to the discussion or not.

The filtering step recall in our experiments shows that trends are good indicators of the discussion content. Entropy based classification yielded low precision and recall for events that share the same topic. The system behaved better on data sets where topics overlapped less. The over-

[5] http://whoosh.readthedocs.org

[6] http://gu.com/p/39c79

Experiment	Filtering		Classifier		System		Baseline	
	Precision	Recall	Precision	Recall	Precision	Recall	Precision	Recall
Primavera Sound	0.60	1.00	0.82	0.85	0.83	0.85	1.00	0.02
Rock am Ring	0.68	0.08	0.96	0.86	0.97	0.07	1.00	0.01
Pentathlon	0.01	0.21	0.01	0.12	0.02	0.06	0.98	0.04
Pinkpop and `#pp11`	0.87	0.95	0.07	0.03	0.14	0.03	0.02	0.05

Table 1: Experiment results

all system performance highly depends on the input data and should be investigated further. No human evaluation was performed, it might be the case that similarly to [4] the output of our system is meaningful and accurate.

Future work should concentrate on classification recall improvement. One way is to process tweet content: label named entities and normalize them. Another way could apply a more sophisticated model instead of the entropy based decision. The only step that has to be done in real-time as the data is consumed is filtering. Classification can be done after the tweets are collected, so an offline method or a more resource consuming approach can be used. In addition, one can use additional data sources, such as news portals or the official web pages of the events of interests.

7. ACKNOWLEDGMENTS

We would like to thank the anonymous reviewers for their comments. Most of the work has been done as a part of the European Master program in Language and Communication Technologies and during an internship at Paylogic Nederland.

8. REFERENCES

[1] H. Amiri, Y. Bao, A. Cui, A. Datta, F. Fang, and X. Xu. Nusis at trec 2011 microblog track: Refining query results with hashtags. In *The Twentieth Text REtrieval Conference Proceedings*, 2011.

[2] J. Benhardus. Streaming trend detection in twitter. *National Science Foundation REU for Artificial Intelligence, NLP and IR*, 2010.

[3] G. Berardi, A. Esuli, D. Marcheggiani, and F. Sebastiani. Isti@ trec microblog track 2011: exploring the use of hashtag segmentation and text quality ranking. In *The Twentieth Text REtrieval Conference Proceedings*, 2011.

[4] M. Bernstein, B. Suh, L. Hong, J. Chen, S. Kairam, and E. Chi. Eddi: interactive topic-based browsing of social status streams. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 303–312. ACM, 2010.

[5] D. M. Blei. Probabilistic topic models. *Commun. ACM*, 55(4):77–84, Apr. 2012.

[6] Y. Chen, Z. Li, L. Nie, and X. Hu. A semi-supervised bayesian network model for microblog topic classification. In *Proceedings of COLING*, 2012.

[7] N. Dai, M. Shokouhi, and B. D. Davison. Learning to rank for freshness and relevance. *Proc. 34th ACM SIGIR*, 2011.

[8] E. Diaz-Aviles, P. Siehndel, and K. D. Naini. Exploiting social#-tagging behavior in twitter for information filtering and recommendation. In *The Twentieth Text REtrieval Conference Proceedings*, 2011.

[9] 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*, pages 57–64. ACM, 2010.

[10] A. Marcus, M. Bernstein, O. Badar, D. Karger, S. Madden, and R. Miller. Twitinfo: Aggregating and visualizing microblogs for event exploration. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 227–236. ACM, 2011.

[11] M. Naaman, H. Becker, and L. Gravano. Hip and trendy: Characterizing emerging trends on twitter. *Journal of the American Society for Information Science and Technology*, 62(5):902–918, 2011.

[12] A. Saha and V. Sindhwani. Learning evolving and emerging topics in social media: A dynamic nmf approach with temporal regularization. In *Proceedings of the 5th International Conference on Web Search and Data Mining (WSDM)*, 2012.

[13] E. T. K. Sang. Het gebruik van twitter voor taalkundig onderzoek. *TABU: Bulletin voor Taalwetenschap*, 39(1/2):62–72, 2011.

[14] K. Tao, F. Abel, and C. Hauff. Wistud at trec 2011: Microblog track. 2011.