

## PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.

<http://hdl.handle.net/2066/101833>

Please be advised that this information was generated on 2017-12-06 and may be subject to change.

# Leveraging Unscheduled Event Prediction through Mining Scheduled Event Tweets

Florian A. Kunneman

Antal van den Bosch

*Centre for Language Studies, Radboud University, P.O.Box 9103, 6500 HD Nijmegen,  
The Netherlands*

## Abstract

A considerable portion of social media messages is devoted to current events. Aside from references to events that recently happened, social media messages may also refer to events that have not occurred yet. Future events, such as football matches in the case study we present here, may be scheduled and known to happen; other future events, such as transfers of football players, may only be rumoured, and may in fact not happen in the end. We describe a news mining component that learns to identify tweets referring to scheduled and unscheduled future events, by being trained on messages referring to scheduled future events (as the latter are easy to harvest). Our results show that discriminating between tweets that refer to upcoming football matches and tweets that refer to past matches can be done relatively reliably with supervised machine learning methods. However, when these trained models are applied to unscheduled events, performance drops to near-baseline performance. We discuss how these results can be explained by the distinction between event type and event domain.

## 1 Introduction

Signalling the likelihood of impending events can be a valuable tool for journalists as well as for the news-reading public, who both wish to be on top of the news as it happens. The massive amount of short messages posted via the medium of `twitter.com`, so-called *tweets*, provide a potentially valuable source of information for this task, outperforming newswire articles in terms of dynamics and pluralism. A key step in the automation of this task is to be able to identify tweets posted to pass on information about, or state an opinion on, an event with a potentially high impact or news value that has not occurred yet. However, such tweets will only represent a small group within the total set of tweets posted at a selected moment in time, making it difficult to highlight them.

Tweets that refer to future events can be detected by training a classifier on positive and negative examples of such tweets that may be gathered from news archives with hindsight knowledge. The intuition is that tweets referring to a future event contain features distinctive from other tweets, including the closely related class of tweets referring to ongoing or past events. For example, future tense and the presence of time adverbs such as *soon* may be strong predictors for English tweets [5]. In order to create a model that captures these features and their weights, a sufficient amount of training material is needed. In this paper we set out to identify tweets referring to future scheduled and *unscheduled* events, where we collect positive cases by harvesting tweets referring to *scheduled* events we know about beforehand. Tweets of this type can often be collected with relatively little effort, as we will demonstrate for the case study domain of football<sup>1</sup>. Scheduled events are often marked by a predictable hashtag (the common way to mark an explicit keyword in a tweet by adding a '#' before a word) that is either recommended in a top-down fashion or has become conventionalized over time. In contrast, hashtags referring to unscheduled events tend to emerge during the process and can have various unpredictable forms.

Although all processed tweets will be embedded in the domain of football, it is not certain whether training on the temporal nature of tweets referring to specific football matches will be effective for the classification of another event type within the same domain such as football transfers, let alone events in

---

<sup>1</sup>We use the term 'football' as the historically accurate name for the sport that is sometimes referred to as 'soccer'.

other domains. This paper describes a case study aimed to test to what extent the similarity between tweets referring to future events can be leveraged across tweet types in the same domain. Classifiers are trained on tweets referring to football matches in the Dutch league and tested on tweets referring to other sorts of matches and unscheduled transfers of football players from one team to another that may or may not materialize. With this research we aim to find out whether the almost effortless collection of training material based on forward knowledge of scheduled events is beneficial for the detection of anticipating tweets in other domains, and ultimately the set of all tweets posted.

This paper is structured as follows. In Section 2 we provide an overview of the relatively large body of recent work on event detection in social media; we review the common trends in this field and zoom in on related work aimed at detecting future events. Section 3 introduces the domain of the case study: scheduled and unscheduled football events. In Section 4 we describe our series of experiments and their results on classifying tweets on football matches into tweets referring to future events versus present or past matches, and on classifying tweets on football transfers. We summarize, state our conclusions, and formulate points for further research in Section 5.

## 2 Related Work

The idea that messages in social media can be used as a source for the prediction of a future event or outcome has been explored in a number of studies. [2] aim to predict the commercial success of specific movies based on the number of tweets that refer to the movie from a week before the premiere. Furthermore, they perform automatic sentiment analysis on tweets posted in the first week after release. [10] perform trending news detection to improve the prediction of stock market changes. [9] aim to predict whether future events mentioned in tweets will actually occur by performing training on events that had a causal relationship. Such event pairs were mined from news archives by searching for certain lexical causality connectors in titles, and normalized by extracting verbs and nouns and connecting them to an ontology. Although these studies consider tweets that refer to future events, the automatic detection of tweets expressing the anticipation of future events has not been investigated to the best of our knowledge.

In order to collect tweets regarding events from the total stream of available tweets, irrelevant messages such as conversational tweets and tweets aimed to share personal experience should be filtered out first. [12] tackle this problem by classifying tweets as either *junk* or *news* based on training on a handlabeled set of tweets, and thereby collect suitable data for a news processing system. Instead of filtering, one can also focus on the distribution of topics discussed on twitter in time, and thereby dispose of tweets not referring to news events if they can be identified as a topic. [8] apply first story detection (the emergence of a news event from a first mention onwards) in tweets, where major events are detected as chains of tweets linked by a similarity score. This way, new topics that have a certain significance are detected online. Rather than filtering news tweets from spam or paying attention to topics, tweets linked to events could also be detected by looking at their linguistic structure. [5] try to extract future events referred to in tweets by searching for specific patterns such as phrases consisting of a verb in the future tense combined with the mention of a time expression.

The detection of tweets referring to events tends to become simpler when the domain of the events searched for is restricted. [11] are interested in tweets mentioning an earthquake in order to warn endangered residents in an early stage. The target tweets are detected by simply searching for tweets with the word 'earthquake'. [6] describe a service to monitor specific events, where the domain is based on user input. The input is enriched by semantic ontologies, thereby filtering the interesting tweets and creating a network around the event. [1] have created *twitcident*, a service to follow current emergencies. The tweets collected before additional filtering are retrieved by keyword search based on input from a police communication network on which emergency services immediately broadcast incidents.

The collection and filtering of tweets referring to *planned* or *scheduled* events is a goal in several studies. [3] aim to provide users with a service to seek information about different stages of the scheduled event (before, during and after the event). They base the keywords connected to events on information from sites such as *upcoming.com*. In order to collect the right tweets, keywords are restricted to a location and specific words describing the event. Additionally, the results from over 50 event queries were labeled by hand, and high precision tweets were used to define new queries and retrieve additional event messages. [4] retrieve tweets referring to matches in the cricket world cup during time of play, and try to extract descriptions of specific micro-events (such as a player scoring a wicket). The tweets are collected during match time using keywords based on general references to the world cup and on common terminology in

the domain of cricket. [7] are also interested in events during matches, focusing on football and rugby, and want to automatically provide the end user with highlights of a match in the form of short segments from the live coverage. Tweets referring to events in matches are collected by queries composed of keywords consisting of the first three letters of the competing teams (not concatenated) and a keyword with reference to the league or cup in which the match is played.

### 3 Case study: Football Events

The case study described in this paper concerns the classification of tweets referring to football matches as scheduled events and football transfers as unscheduled events. The goal is to test if there is an overall pattern in anticipating tweets, i.e. tweets that refer to future football events. A practical reason why Dutch tweets in the domain of football are collected as target material is that football is the number one sport in the Netherlands, and accordingly a sufficient number of people tweet about football matches and transfers. Furthermore, there is a multitude of events in the form of matches each round of the league, enabling a lot of keyword-based searching. A key advantage of tweets referring to scheduled events is that it can be established exactly, by their time stamp and the known timing of the scheduled event they refer to, whether they are posted before, during or after a match.

## 4 Experiments

Our case study consists of two experiments. In the first experiment, described in Section 4.1, we train supervised machine-learning classifiers to distinguish before-match tweets from tweets generated during or after a match. In the second experiment, described in Section 4.3, a classifier trained on the former type of tweets is applied to tweets referring to transfers, testing if tweets referring to scheduled events can be useful training material for determining whether a tweet is mentioning a future unscheduled event.

### 4.1 Football Matches: Experimental Setup

#### 4.1.1 Corpus

The corpus used in this study consists of tweets referring to football matches. The tweets are collected in an online fashion by means of selected search terms. The convention to refer to a football match by concatenating the first three characters of the home and away team respectively to serve as hashtag (for example, '#ajafey' for home team Ajax playing against Feyenoord) was used for high-precision retrieval of match tweets. All matches in the Dutch premier league, the *Eredivisie*, were harvested through these conventional match hashtags, and collected in the period from April 3 2012 until May 23 2012, the final weeks of the 2011–2012 season, including play-offs (small tournaments to settle promotion / relegation or tickets to European football). In addition we collected tweets referring to the UEFA Champions League final between Bayern Munich and Chelsea FC of May 19 2012. The retrieved tweets were restricted to Dutch user accounts in order to maintain a single language throughout the tweets as much as possible. The tweets were collected by sending queries to the Twitter API every two minutes. This short interval was applied in order to catch all the tweets posted during matches, as there is a high density of football-related tweets during game time. The keywords in the form of match hashtags were all linked to a specific timeslot in which the match was played, in order to directly label each incoming tweet based on the time at which it was posted ('before', 'during' or 'after' a match).

The resulting set of tweets was filtered by removing duplicate tweets, retweets and tweets that only consisted of a url or hashtag. This resulted in a final set of about 70 thousand tweets. These tweets were tokenized by `ucto`, a rule-based tokenizer for Dutch<sup>2</sup>. The tokenized tweets were additionally cleared of punctuation, URLs and hashtags.

#### 4.1.2 Classification

In order to get a general overview of classifier performance on the task, five typical supervised classification algorithms were applied:  $k$  nearest-neighbor (Knn) classification, Winnow, SVM, MaxEnt, and Naive

---

<sup>2</sup><http://ilk.uvt.nl/ucto>

Single words (English translation)	Expressions
morgen (tomorrow)	volgende? (dag week maand weekend)
overmorgen (the day after tomorrow)	(aan)?komende? (dag week maand weekend)
straks (soon)	(kaartje ticket)s?
binnenkort (soon)	
zo direct (soon)	
zometeen (soon)	
zin in (feel for)	
maandag-zondag (monday-sunday)	

Table 1: Words and expressions that form the basis of the vocabulary baseline

Bayes. For Knn and SVM, the *PyML*<sup>3</sup> implementation was used, while the Machine Learning for Language Toolkit<sup>4</sup> was used for Winnow, Naive Bayes and MaxEnt classifiers. SVM was applied with a second-order kernel and the  $k$  hyperparameter of Knn was set to 5.

The different event subdomains distinguished in the retrieved football tweets were league matches, play-off matches, and the 2012 Champions League final. Distinguishing features of matches in the play-offs are the definite character (there is more at stake than in the case of most league matches) and the fact that a smaller pool of clubs is involved. This might result in more emotional tweets and tweets from a more specific group of people supporting the clubs. The Champions League final shares the definite character with a high chance of emotionally loaded tweets. On the other hand, a broader public in comparison to league and play-off matches is compelled to tweet about the final. The total set of retrieved tweets contains 57,109 tweets referring to one of 86 league matches, 7,382 tweets referring to one of 20 play-off matches, and 3,404 tweets referring to the Champions League final. 10-fold cross-validation was performed on the league tweets in a first series of experiments. Then, all league tweets were used as a single training set for the classification of the tweets referring to the play-offs and the final.

The selection of features was kept to words only: from the word sequences in the tweets we derived unigram, bigram, and trigram features. By combining the three sorts of N-grams, significant longer patterns could be found while often recurring unigrams and bigrams are given an extra weight by their respective higher N-grams. Dimensionality was restricted by pruning all features occurring less than ten times. In a first classification run the removal of stopwords, stemming, lemmatization and the addition of part-of-speech tags did not lead to improvement of the results. A possible explanation for this is that tweets contain rather non-standard language and tokens, and linguistic preprocessing is therefore unreliable. For this reason the additional preprocessing was not implemented during the final classification leading to the presented results.

## 4.2 Football Matches: Results

The results of classification on tweets referring to football matches are listed in Table 4.2. Results are given as precision, recall, and F1-scores of the identification of ‘before’ tweets. The ‘before’ baseline refers to the baseline strategy of labeling all tweets with the focus category. The ‘vocabulary’ baseline consists of classifying all tweets as ‘before’ that contain one of a selected set of words or expressions (Table 4.2).

The table shows that all five classifiers obtain a good precision and recall for league match classification, scoring between 0.1 and 0.15 above the ‘before’ baseline F1 result. When classifying play-off tweets based on league training data the improvement over the ‘before’ baseline F1 is smaller, both due to a lower precision and recall. The performance on tweets referring to the Champions League final is worse, but still better in comparison to the baseline. The ‘before’ baseline for this subset is quite low as the percentage of tweets posted before the final is a lot smaller than in the case of the other subsets. The ‘vocabulary’ baseline consistently leads to the best precision, but performs bad in terms of recall. This shows that the set of ‘before’ tweets in each domain has quite some diversity, and simple future reference forms only one part. The precision scores of the classifiers indicate that the league tweets as training data do help in distinguishing tweets anticipating the final, with a markedly higher precision for both SVM and MaxEnt in comparison to the other classifiers.

Because sports matches themselves are a special kind of timed event with many particular micro-events that may be the subject of messages, the set of tweets posted during a match could hamper the task of

<sup>3</sup><http://pyml.sourceforge.net/>

<sup>4</sup><http://mallet.cs.umass.edu/>

	league matches (10-fold)			play-off matches			CL final		
	precision	recall	F1	precision	recall	F1	precision	recall	F1
‘before’ baseline	0.57	1	0.73	0.56	1	0.72	0.38	1	0.55
‘vocabulary’ baseline	0.92	0.22	0.36	0.90	0.22	0.36	0.81	0.15	0.26
Naive Bayes	0.86	0.88	0.87	0.76	0.83	0.79	0.57	0.86	0.69
MaxEnt	0.88	0.88	0.88	0.83	0.75	0.79	0.69	0.8	0.74
Winnnow	0.77	0.88	0.82	0.78	0.67	0.72	0.58	0.68	0.62
SVM	0.88	0.88	0.88	0.82	0.76	0.79	0.6	0.77	0.68
Knn	0.76	0.89	0.82	0.67	0.79	0.73	0.46	0.84	0.60

Table 2: Precision, recall and F1-scores on labeling tweets as ‘before’ in three experiments: 10-fold cross-validation on league matches (left), on the post-season playoff matches (middle), and on the 2012 Champions League final (right).

	league matches (10-fold)			play-off matches			final		
	precision	recall	F1	precision	recall	F1	precision	recall	F1
‘before’ baseline	0.75	1	0.86	0.75	1	0.86	0.45	1	0.62
‘vocabulary’ baseline	0.94	0.22	0.36	0.92	0.22	0.36	0.84	0.15	0.26
Naive Bayes	0.88	0.89	0.89	0.82	0.95	0.88	0.57	0.96	0.71
MaxEnt	0.88	0.94	0.91	0.87	0.89	0.88	0.71	0.88	0.79
Winnnow	0.84	0.90	0.87	0.84	0.85	0.85	0.66	0.75	0.7
SVM	0.91	0.94	0.92	0.86	0.86	0.86	0.58	0.85	0.69
Knn	0.84	0.96	0.9	0.79	0.91	0.85	0.48	0.94	0.64

Table 3: Precision, recall and F1-scores on labeling tweets as ‘before’ in league matches (left), playoff matches (middle), and the CL final (right), with game-time (‘during’) tweets removed.

classifying ‘before’-tweets, as these tweets do not as much refer to the event at large as ‘before’ or ‘after’ tweets do. Furthermore, tweets referring to unscheduled football events will not contain these specific game-time tweets. To measure the effect of this particular class which was included in the first experiment, we performed a second experiment on an alternate version of the league, off-season and final game tweets: without the tweets posted during matches. The results of this experiment are displayed in Table 3.

The results in Table 3 indicate that removing game-time (‘during’) tweets leads to an overall improvement in the performance of the classifiers on the ‘before’ class, while on the other hand the difference with the ‘before’ baseline score has decreased. This can be explained by the fact that the relatively higher percentage of tweets with the label ‘before’ leads to a considerable improvement of baseline precision and F1. This somewhat trivial result is furthermore colored by the fact that the removal of ‘during’ tweets can only be done in situations in which the exact game time is known, which in our training data is the case, but which may very well be unknown in another automatic news mining scenario.

When comparing the performance of the different classifiers on this dataset (both with and without tweets during matches), a number of observations can be made. In terms of F1 performance, the MaxEnt classifier has the best performance on the playoff and final subsets, suggesting that it learns the best generalizing feature weights from the league data during training. This contrasts with the SVM performance, which is strong in the 10-fold cross-validation experiments on the league data, but falls below the performance of MaxEnt on the tweets referring to the final. Knn and Naive Bayes both attain relatively high recall rates, which comes at the cost of a lower precision. With a majority of tweets in the training data labeled ‘before’, the high value of  $k$  = in the Knn classifier and the high prior probability for the class lead to a high recall and low precision on the class with both algorithms.

In sum, this first experiment showed that tweets before football matches could quite accurately be distinguished from tweets after matches based on their content, and that a reasonable performance is maintained when applying the classifiers on matches of somewhat different types, without additional training. The goal of the second experiment described in the following section is to test whether training on league matches is still valuable when applying the classifiers to the more distant event type of unscheduled transfers of football players.

## 4.3 Football Transfers: Experimental Setup

### 4.3.1 Corpus

As the first step in collecting transfer tweets a number of rumoured transfers in Dutch professional football from the summer of 2011 until the end of the 2011–2012 season were collected from the Dutch website [www.transferboulevard.nl](http://www.transferboulevard.nl). On this site visitors can post a transfer rumor, as well as assess already posted rumors on their credibility. Every transfer rumor collected from this site contains a headline, a text, its author and assessment scores. In order to formulate a query for the collection of tweets, named entities were extracted from the headlines by means of Named Entity Recognition performed by Frog, a freely available morpho-syntactic text analyzer for Dutch<sup>5</sup>. When at least a person and an organization were identified in a headline, all named entities collectively formed a query for tweets. The idea is that the combination of a player, a new club and a time frame around the moment when the transfer either happens or fails forms an accurate set of keywords via which a collection of the tweets referring to a transfer can be harvested.

Before collecting tweets based on the formulated queries, the transfer events on which the queries were based were hand labeled as leading either to a positive or negative outcome and with the date of this outcome, based on fact checking in reliable news sources. Rumours of transfers that still ‘slumbered’ (i.e. were not resolved at the time of writing) were removed from the set. This resulted in 90 transfer events with the label ‘occurred’ and 192 transfer events with the label ‘not occurred’.

The transfer events from which queries were formulated dated back to July 2011. The API offered by [twitter.com](http://twitter.com) does not go back this far. In order to collect all tweets in time referring to a transfer, Topsy search<sup>6</sup> with a searchable collection of past tweets from May 2008 onwards was queried using the Otter API<sup>7</sup>. This resulted in 3,852 tweets in the category ‘occurred’ and 3,731 tweets in the category ‘not occurred’, resulting in a set of 7,583 tweets in total.

### 4.3.2 Classification

In order to evaluate the automatic classification of tweets in the collected set by the classifiers described in the previous section, all transfer tweets are labeled ‘before’ or ‘after’ by their known date of the actual occurrence or failure of the transfer. Tweets posted on the same date as a transfer outcome are given the label ‘after’, because they mostly are a reaction to the outcome of the transfer it refers to. As the outcome of a transfer might influence the tweets, this categorization is maintained for classification. Tweets referring to rumoured transfers that neither have a positive or negative outcome (slumbering rumours) are withheld from the corpus. This results in three sets of tweets on which classification is performed: tweets referring to transfers that occur, transfers that fail, and the former sets combined.

The tweets are preprocessed in the same way as the match tweets, and again the unigrams, bigrams, and trigrams from each tweet are retrieved as features. The five classifiers applied in the former experiment are trained on the league training data without game-time tweets.

## 4.4 Football transfer Results

The results of the classification are given in Table 4.4. The baseline score is computed on grounds of classification of all tweets as ‘before’. For the ‘vocabulary’ baseline, the same list presented in tabel 4.2 is used.

The results show a marked decline in comparison to the classification of tweets referring to matches reported in the previous section. The classifiers do not outperform the ‘before’ baseline in terms of the F1 score. The generalization performance of the MaxEnt classifier is now the lowest, while it was the best generalizing classifier of tweets regarding the final or play-off matches. In terms of recall, the Knn and Naive Bayes classifiers still retain a good performance, at the cost of a near-baseline precision.

As an extra analysis with hindsight knowledge, the classification on the different outcomes of a transfer (success or failure) are displayed in Table 4. A main difference between these two outcomes is the percentage of tweets before the conclusion of a transfer: 39% for transfers that did materialize, versus 62% for transfers that did not. Apparently, successful transfers evoke more reactions afterwards than transfers that are cancelled. When looking more closely at classifier performance, there is somewhat of a split between Naive Bayes, MaxEnt and Winnow on the one hand, performing reasonably well on occurred transfer tweets,

---

<sup>5</sup><http://ilk.uvt.nl/frog>

<sup>6</sup><http://topsy.com/>

<sup>7</sup>[otter.topsy.com/](http://otter.topsy.com/)

	<b>all transfer tweets</b>			
	precision	recall	F1	accuracy
'before' baseline	0.5	1	0.67	0.5
'vocabulary' baseline	0.63	0.02	0.04	0.5
Naive Bayes	0.42	0.48	0.45	0.41
MaxEnt	0.39	0.34	0.4	0.39
Winnnow	0.46	0.58	0.51	0.45
SVM	0.54	0.56	0.55	0.53
Knn	0.51	0.92	0.66	0.51

Table 4: Performance scores, including accuracy, on labeling transfer tweets as 'before', by the five machine-learning algorithms.

	<b>successful transfers</b>			<b>failed transfers</b>		
	precision	recall	F1	precision	recall	F1
'before' baseline	0.39	1	0.56	0.62	1	0.77
'before' baseline	0.57	0.04	0.07	0.78	0.01	0.03
Naive Bayes	0.56	0.76	0.65	0.41	0.7	0.52
MaxEnt	0.56	0.61	0.58	0.4	0.57	0.47
Winnnow	0.53	0.68	0.6	0.38	0.63	0.48
SVM	0.41	0.51	0.46	0.62	0.6	0.62
Knn	0.39	0.93	0.55	0.62	0.92	0.74

Table 5: Performance score on labeling transfer tweets as 'before', split on successful transfers and failed transfers.

and SVM and Knn on the other hand, doing well in the case of failed transfer tweets. Of course, the outcome of a transfer is not known in advance, so it is hard to make any conclusions based on this difference.

On the whole, the generalization performances of the classifiers applied to transfer tweets are quite low in terms of precision, recall and accuracy, underlining the difficulty of the task to classify the state of a tweet linked to an event type different from the training data, even though they are all football events.

## 5 Discussion

The case study presented in this paper shows that the period in which a tweet is posted related to an event, when discretized into 'before' and 'not before', can be classified reasonably accurately on the basis of training data with the same event type, regardless of slight event type variations, as we showed with the league matches, the play-off matches, and the Champions League final, which were all classified well when trained just on league matches. However, classifying tweets on another event type (transfers of football players) based on the same training data leads to a poor performance. Thus, the presumed similarity between anticipating tweets regardless of the event type is not so apparent. This shows that the domain has an influence on the tweets equal to their anticipating nature.

As the identification of anticipating tweets in general is an interesting task for news mining systems, more research could be undertaken starting from our current experimental setup. Instead of a case study in one domain, a more general approach may be followed in which the overall anticipating pattern is sought by collecting and performing training on tweets from many domains and event types mixed together. Alternatively, more generic classifiers could be trained by explicitly filtering away event-specific features such as named entities and other content words, while selecting or placing more weight on tense markers and time expressions.

Another research question to be pursued in further research would be what the most discriminative factors are that characterize tweets referring to scheduled events versus those referring to unscheduled events. That it is unknown whether the latter event will happen or not is likely to add a speculative aspect to the tweets anticipating such events. The difficulty remains, however, that it takes considerably more effort to accurately harvest and label tweets anticipating unscheduled events than the virtually effortless harvesting



and labeling carried out in our study.

## References

- [1] F. Abel, C. Hauff, G. Houben, K. Tao, and R. Stronkman. Semantics + Filtering + Search = Twitcident, Exploring Information in Social Web Streams. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media, HT 2012*, 2012.
- [2] S. Asur and B. A. Huberman. Predicting the Future with Social Media. 2010.
- [3] H. Becker, F. Chen, D. Iter, M. Naaman, and L. Gravano. Automatic Identification and Presentation of Twitter Content for Planned Events. 2011.
- [4] S. Choudhury and John G. Breslin. Extracting Semantic Entities and Events from Sports Tweets. May 2011.
- [5] Alan Jackoway, Hanan Samet, and Jagan Sankaranarayanan. Identification of live news events using Twitter. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, LBSN '11*, pages 25–32, New York, NY, USA, 2011. ACM.
- [6] P. Kapanipathi, C. Thomas, Pablo N. Mendes, and A. Sheth. Continuous Semantics: Dynamically Following Events. In *Proceedings of Manufacturing & Service Operations Management (MSOM), 2011*, 2011.
- [7] J. Lanagan and A. F. Smeaton. Using Twitter to Detect and Tag Important Events in Live Sports. In *Fifth International AAAI Conference on Weblogs and Social Media*, 2011.
- [8] Saša Petrović, Miles Osborne, and Victor 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, HLT '10*, pages 181–189, Stroudsburg, PA, USA, 2010. Association for Computational Linguistics.
- [9] K. Radinsky, S. Davidovich, and S. Markovitch. Learning causality for news event prediction. In *Proceedings of the 21st international conference on World Wide Web*, 2012.
- [10] J. Ritterman, M. Osborne, and E. Klein. Using Prediction Markets and Twitter to Predict a Swine Flu Pandemic. In *1st International Workshop on Mining Social Media*, 2009.
- [11] Takeshi Sakaki, Makoto Okazaki, and Yutaka Matsuo. Earthquake shakes Twitter users: real-time event detection by social sensors. In *Proceedings of the 19th international conference on World wide web*, WWW '10, pages 851–860, New York, NY, USA, 2010. ACM.
- [12] Jagan Sankaranarayanan, Hanan Samet, Benjamin E. Teitler, Michael D. Lieberman, and Jon Sperling. TwitterStand: news in tweets. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, GIS '09*, pages 42–51, New York, NY, USA, 2009. ACM.