

Mapping Tweets to Conference Talks: A Goldmine for Semantics

Milan Stankovic^{1,3}, Matthew Rowe², Philippe Laublet³

¹ Hypios Research, 187 rue du Temple, 75003 Paris, France

² Knowledge Media Institute, The Open University, Milton Keynes MK7 6AA, UK

³ STIH, Université Paris-Sorbonne, 28 rue Serpente, 75005 Paris, France

milan.stankovic@hypios.com, m.c.rowe@open.ac.uk, philippe.laublet@paris-srobonne.fr

Abstract. The short message service Twitter has gained significant popularity and uptake among participants of conferences and organized events as a backchannel for intra-event communication. Information that is exchanged explicitly through such tweets, or that is implicitly present in them, remains mostly hidden and undecipherable to machines. In this paper we propose a framework for extracting valuable information from conference tweets, enabling its publication as Linked Data. We introduce the concept of mapping tweets with the talks and subevents that they refer to, in doing so gaining access to additional information about the users, talks and dynamics of the event. We present preliminary results of our work towards tweet-talk mappings and motivate our current and future work by giving several use cases for such extracted data.

Keywords: Semantic Web, Social Web, Twitter, User Profiling, Linked Data

1 Introduction

In recent years Twitter¹ communication among participants has become an important element of conferences and participatory events. Rather than whispering during talks, scientists, researchers and professionals of all kinds now use Twitter to pass their thoughts, exchange impressions about the events they are attending, share links and useful information, and interact in real-time, and at some events Twitter is used as a viable medium to collect audience questions. Many conferences now publish a Twitter #hashtag in their official program, allowing the participants to easily follow what other participants are tweeting. Given this context of application, it is clear that a lot of information about an event is captured in Twitter streams. Extracting such information requires techniques able to interpret and leverage content from a tweet, enabling analyses to be performed automatically – given that the scale and uptake of Twitter hinders manual analysis. Furthermore, tweets at conferences often discuss sub-events such as talks, presentations and keynote speeches. The alignment of tweets with such events would enable conference feedback: by assessing the sentiment of tweets and identifying the most popular/unpopular aspects of a conference; and user

¹ <http://www.twitter.com>

profiling: by modeling a user based on their area of interest and suggesting talks/users to speak with based on shared contextual information.

In this paper we propose a solution to mapping tweets with conference talks. Our solution enriches tweets with semantics using the SIOC ontology, identifies DBPedia concepts which the tweets mentions, and maps tweets with talks using an automated technique. We have collected a dataset of tweets produced during the Extended Semantic Web Conference 2010, from which we have preliminary results following the conversion of tweets into triples and identifying DBPedia concepts which the tweets mention. As this paper is presenting a work in progress, we demonstrate the utility of human-provided mappings to highlight the benefits of this data for providing feedback to conference organizers, and to compile user interest profiles.

We have structured this paper as follows: section 2 presents the key issues associated with the extraction of data from Twitter, explaining our general approach to overcoming such issues. In section 3 we present concrete use-cases that motivate our work on tweet-talk mappings and the exploration of Twitter conference data in general. Section 4 presents our framework for information extraction from Twitter conference archives and for the semantic enrichment of the data. We demonstrate its utility on several concrete examples in Section 5. Section 6 lists some of the related research initiatives, and finally in section 7 we conclude the paper and give future work directions.

2 Issues with Twitter

2.1 Amount of Tweets

Although all tweets may be considered as a generally useful source of information, the growing rate of new tweets (more than 50 million²) per day combined with Twitter API limitations makes this source almost unprocessable in real-time, therefore requiring the adoption of a strategy to filter out the tweets before processing. Another issue is the amount of tweets that are poor in information content. Those tweets are mostly expressing emotions, phatic communication or similar content. Our approach for overcoming this issue is to focus on tweets that are created during conferences and professional events. Apart from the intuition that those tweets might be richer in meaning, there are studies that support such a choice (e.g., [1] and [2]). For instance, authors of [1] claim that only 15% of conference tweets are of a personal nature, indicating that they must contain fewer phatic communication messages that naturally occur with personal communication. Given the property of containing many relevant tweets, conferences represent a convenient filter for getting to information-rich tweets. Section 3 presents the possible uses of conference tweets and motivates our focus on this particular type of Twitter content.

² <http://mashable.com/2010/02/22/twitter-50-million-tweets/>

2.2 Swift Disappearing of Tweets

Due to scalability problems and the enormous amount of tweets that the service hosts, Twitter is providing only a limited access to tweet history. Tweets disappear from public searches after 4-7 days – a time frame that is getting lower and lower, which makes treating past tweets difficult. In order to overcome this issue archiving systems for Twitter have appeared. The most ambitious is Topsy³, a service trying to recreate the archive of the most *important* tweets. Although the archive goes far back in history, it contains only a portion of tweets published in the total superset of the Twittersphere. An alternative approach is offered by Twitter archiving services (e.g., TweetBackup⁴, TwapperKeeper⁵) and desktop tools (e.g., Archivist⁶, Twinbox⁷). On TwapperKeeper in particular, once the archive is created for a particular #hashtag, it becomes publically available. We have chosen to rely on Twitter archives for our extraction process, and to use TwapperKeeper because the service contains valuable archives for over 600 conference #hashtags, created by conscious participants.

3 Use-cases for Twitter Conference Data

Several use-cases exist that motivate our work for leveraging semantics from conference tweets based on the assumptions that: (1) conference tweets can be accessed; (2) topics of tweets can be identified; (3) conference subevents are provided; and (4) correspondences can be drawn between tweets and conference subevents (e.g., a talk or a workshop) that tweets refer to. Those tweet-talk mappings represent the key point of our approach and make a true difference in the following scenarios.

3.1 Digital Memory

Many users use Twitter to share links and other useful observations during conferences. According to studies in [3], information sharing represents an important portion of tweeting activities. Tweets remain as a trace of such exchanges and allow the user to find the shared information later. The conference in question provides a context for the user's memory, allowing information to be retrieved more easily. Furthermore, many tweets from conferences are about a particular talk or subevent that the user has attended, providing an additional level of digital memory.

3.2 Sentiment Analysis, Conversational Aspects and Conference Feedback

Twitter has already been the subject of sentiment analysis research, and general systems for analyzing emotions of tweets have already been proposed [4]. Since in the case of conference tweets, there may be correspondences between tweets and

³ <http://topsy.com>

⁴ <http://tweetbackup.com/>

⁵ <http://twapperkeeper.com>

⁶ <http://visitmix.com/labs/archivist-desktop/>

⁷ http://www.techhit.com/TwInbox/twitter_plugin_outlook.html

conference subevents it is clear that such tweets could be used to measure the overall opinion of the talks and other similar presentations, thereby providing useful feedback for presenters and conference organisers alike. Another interesting direction is the study of conversational aspects of tweets, like discussions formed through chains of *@replies* and the memes formed by retweets – aspects of Twitter that have already been the subject of substantial studies [5], [6]. Intensity of conversations and memes produced around a particular conference talk could help spot the most influential speakers and the most appealing topics. Furthermore a key benefit of this approach is the contrast in implicit and explicit feedback collection: in the latter conference attendees are provided with a form through which they express their perspective on the conference, whereas the former captures the implicit opinion, which is divulged and shared between attendees.

3.3 User Profiling and Expert Finding

Based on tweet-talk mappings we could easily infer that a user having created a tweet about a given talk, and showed interest in the event, would also be interested in the topic – this is particularly pertinent at conferences given the requirement for the selection of talks amongst parallel tracks. Abstracts of talks are usually available online and contain useful topics that can be understood as identifiers of user interest. In our research we will focus on conferences where metadata is already available in a semantic format on the Semantic Web Dog Food⁸ website. We aim to demonstrate that tweet-talk mappings can help us improve the completeness of user interest profiles through the propagation of topics from talks to tweets, and by transitivity, the alignment of users with interests.

3.4 Interest Dynamics and Trending Topics

Based on the previous considerations of tweet-talk mappings and the propagation of topics from talks to tweets, we could easily imagine another use, which is the following of popular topics and the change in their popularity throughout the conference. Relations between topics that are present in DBPedia⁹ could make this functionality even more sophisticated by taking into account topic proximity in the DBPedia graph.

4 Rich Activity Twitter/Event Data (RATED) Framework

We now describe the central contribution of this paper: a framework, known as RATED, for processing archives of conference tweets and extracting useful event-related semantics from a given corpus. Our approach is divided into three distinct stages, as shown in Figure 1: the first stage is the production of Linked Data from tweets. This involves taking the basic structure and tweet metadata, and publishing this data as triples using the available ontologies and according to the Linked Data principles – we describe this stage in greater detail the following section. In the

⁸ <http://data.semanticweb.org>

⁹ <http://dbpedia.org>

second stage of our approach, we enrich the tweets with information about the topics that they discuss. This information is in the form of concepts defined by dereferenceable URIs, this associates a given instance of a tweet with one or more DBPedia concepts - describing the implicit semantics of the tweet is important for being able to search for tweets and analyse the topical dynamics of conferences. We detail this stage in the section 4.2.



Figure 1. The core process of the RATED framework

Finally, in the final stage of the approach we address the biggest challenge needed to make most from conference tweets by being able to draw correspondences from a tweet to a particular talk or subevent that the tweet refers to (wherever a tweet refers to a talk or subevent). This is the mapping stage of the approach, where we propose to train a machine learning classifier for a multi-class classification problem, using talk URIs and then use the classifier to label the corpus of tweets. We explain our proposed method for the mapping task within section 4.3. We have applied our approach on the tweets and event data from the recent Extended Semantic Web Conference 2010. Statistics and discussion about the concrete data are provided in Section 5.

4.1. Tweet Data Extraction

We have built a Java-based parser that can process TwapperKeeper archives in comma separated values format. Once the tweets are imported, we use the official Twitter API¹⁰ to retrieve the user and user account data (e.g., name, biography). The system later saves all this information in a triple store. At present the system is capable of using Jena¹¹ and Talis¹² triple stores. An example tweet is shown below in the RDF snippet. We have tried to make the most general representation of tweets possible in order to maximise the potential use of published data. Tweets are therefore represented as MicroblogPosts in the sense of the SIOC ontology¹³, the maker of the tweets is described using the FOAF ontology¹⁴, and the actual act of tweeting is wrapped into an instance of OnlinePresence described using the OPO ontology [7]. General properties like titles are represented using Dublin Core vocabulary¹⁵. Apart from tweets the data set contains full descriptions of tweet authors and their Twitter user accounts.

¹⁰ <http://apiwiki.twitter.com/>

¹¹ <http://jena.sourceforge.net/>

¹² <http://www.talis.com/platform/>

¹³ <http://rdfs.org/sioc/spec/>

¹⁴ <http://xmlns.com/foaf/spec/>

¹⁵ <http://dublincore.org/documents/dcmi-terms/>

```

<rdf:Description rdf:about="http://data.hypios.com/tweets/tweet-15162225891">
  <dc:description>A tweet :Noshir Contractor at #eswc2010 speaking of data-
  driven social network analysis of MMORPG such as WoW and EQ2</dc:description>
  <rdf:type rdf:resource="http://online-presence.net/opo/ns#OnlinePresence"/>
  <opo:customMessage>
    <sioc:MicroblogPost
      rdf:about="http://data.hypios.com/tweets/tweet-15162225891-cm">
      <sioc:content>Noshir Contractor at #eswc2010 speaking of data-driven social network
      analysis of MMORPG such as WoW and EQ2</sioc:content>
      <sioc:id>15162225891</sioc:id>
      <dcterms:language>en</dcterms:language>
      <foaf:maker rdf:resource="http://data.hypios.com/tweets/user-ciro"/>
      <dcterms:date>2010-06-01T09:16:46+0200</dcterms:date>
      <dcterms:subject
        rdf:resource="http://dbpedia.org/resource/Social_network"/>
      </sioc:MicroblogPost>
    </opo:customMessage>
    <opo:declaredBy rdf:resource="http://data.hypios.com/tweets/user-ciro"/>
    <opo:startTime>2010-06-01T09:16:46+0200</opo:startTime>
    <opo:publishedFrom>
      <opo:SourceOfPublishing
        rdf:about="http://data.hypios.com/tweets/tweet-15162225891-source">
        <opo:sourceName>http://twitter.com</opo:sourceName>
      </opo:SourceOfPublishing>
    </opo:publishedFrom>
  </rdf:Description>

```

4.2 Topic Enrichment

In order to make tweets more searchable and make them useful for the described scenario one would need to have represent them in a structured format and associate those tweets with the topics that refer to. As a basic approach we process the text of the tweets using Zemanta¹⁶ keyword extraction API. This approach gives DBPedia concepts related to a tweet, but is useful only on a limited number of tweets directly mentioning a particular topic. In the example data set we produced from ESWC2010, tweet topics are represented using *dc:subject* property.

4.3 Mapping Tweets to Conference Talks

Mapping conference tweets to the corresponding parts of conference events performs reference reconciliation: i.e., associating a tweet URI with a talk URI. Our work on an automated mapping solution is still a work in progress, where we regard the problem as a multi-class classification task: given a collection of tweets T our goal is to choose the most appropriate class label from a set of URIs describing the conference events Y . As training data we will use properties and attributes associated with each event URI, by dereferencing each URI and converting the instance description into a bag-of-words model. This provide a dataset of the form: $D = \{(x, y)\}$ where $x \in X$ denotes the collection of features for each URI, and $y \in Y$ are the class label for each talk. Our intuition is that by converting each tweet into a similar bag-of-words representation as each talk – dereferencing the tweet URI and compiling a set of features – we can then train a classifier using D and classify each tweet, choosing the class label – and therefore URI – with the highest classification confidence.

Existing evaluation metrics are applicable in our scenario such as precision and recall, and therefore f-measure with a suitable setting for β . However evaluation

¹⁶ <http://developer.zemanta.com/>

requires a gold standard against which the mappings generated from our automated technique can be compared. In the remainder of this section we describe the process by which we created this gold standard, and to provide contextual evidence of the usefulness of tweet-to-talk mappings, we present several use cases based on human-generated mappings in the following section.

Table 1. Rater Agreement Confusion Matrix

		Rater 1	
		Positive	Negative
Rater 2	Positive	a	b
	Negative	c	d

4.3.1 Gold Standard Creation

To create a reliable set of correct tweet-talk mappings for evaluating our automatic mapping technique, we conducted human classification of tweets with related talks. We used three evaluators who attended the ESWC2010 conference, to perform the human classification, and took a random subset of 200 tweets for human classification. Each evaluator mapped each tweet with one of the ESWC2010 talks available on the Semantic Web Dog Food website, or in some cases stated that the tweet does not correspond to any talk. The evaluation was done using an iterative process inspired by the Delphi method [8] to achieve agreement of experts about a certain opinion (most commonly a prediction). Experts give their predictions and justify them. In the second round they read the predictions and justifications of other experts, and can change their mind and thus agree on a common prediction. We used the same process to reach an agreement about tweet-talk mappings. To calculate the agreement levels between raters we used the kappa (k) agreement metric, defined in [9], to calculate the evaluator agreement in each phase. The k -statistic measures the chance corrected agreement between raters, using the confusion matrix shown in Table 1. Using these set definitions the k -statistic is calculated using the following formula:

$$\kappa = \frac{2(ad - bc)}{(a + c)(c + d) + (b + d)(a + b)}$$

The average agreement between raters following the first iteration of the Delphi method was 0.328, which indicates a very low level of agreement, as 0.6 is generally agreed to be acceptable value (see Table 2. ‘‘I round’’). We then conducted a second round in which evaluators could see the mappings given by others and change their mind, or argue in favour of their previous mappings. After the second round, a satisfactory level of agreement ($k = 0.820$) had been reached. Therefore we took mappings from the second round as our gold standard.

Table 2. Reaching evaluators’ agreement about tweet-talk mappings

	Agreement Function k	
	I round	II round
evaluator1 : evaluator2	0.330	0.905
evaluator1 : evaluator3	0.307	0.795
evaluator2 : evaluator3	0.348	0.761
Average Agreement	0.328	0.820

5 The Produced Dataset Statistics

To contextualize the benefits of our approach, we took the TwapperKeeper #eswc2010 hashtag archive for the Extended Semantic Web Conference (ESWC) 2010. This conference was chosen because of our ability to engage with the participants of the conference in the evaluations. All the Twitter data extracted is published in an online data store in accordance with the Linked Data principles [10], and a SPARQL¹⁷ endpoint is accessible at the following location: <http://data.hypios.com/tweets/sparql>. In total we extracted 1082 ESWC tweets. 213 tweets had DBpedia topics directly associated with them using Zemanta, making a total of 252 connections of a tweet with a DBpedia topic. The most used topics are presented in the Table 3. In our generation of human mappings, we produced a total on 89 mappings between tweets and ESWC talks. This is based on our subset of 200 tweets as explained in the previous section. All of the mappings are available through the above SPARQL endpoint¹⁸.

Table 3. Most popular topics in the data set of ESWC2010 tweets

<i>Topic URI</i>	<i>Count</i>
http://dbpedia.org/resource/Semantic_Web	30
http://dbpedia.org/resource/Linked_Data	24
http://dbpedia.org/resource/SPARQL	20
http://dbpedia.org/resource/Resource_Description_Framework	11
http://dbpedia.org/resource/Social_network	9
http://dbpedia.org/resource/Uniform_Resource_Identifier	8
http://dbpedia.org/resource/Radio-frequency_identification	7
http://dbpedia.org/resource/Hypertext_Transfer_Protocol	6

The remaining 111 tweets were either not related to a talk, or related to a talk/subevent for which there was no information on the Semantic Web Dog Food website (such as panel discussions or informal gatherings). The talks themselves are also enriched with topics (by submitting their abstracts and titles to Zemanta) and the 89 tweet-talk bindings lead to a total of 255 topic bindings, which is more than the total of topic bindings generated using simple tweet text for the whole dataset. Table 4 shows the most used topics that appear with talks from the tweet-talk mappings.

Table 4. Most popular topics in the data set of ESWC2010 talks related to tweets

<i>Topic URI</i>	<i>Count</i>
http://dbpedia.org/resource/Data	15
http://dbpedia.org/resource/Semantics	13
http://dbpedia.org/resource/Semantic_Web	12
http://dbpedia.org/resource/Resource_Description_Framework	6
http://dbpedia.org/resource/Graph_%28mathematics%29	5
http://dbpedia.org/resource/DBpedia	5

Based on these tables as well as the histogram in Figure 2, it is clear that topics based on tweet text tend to be slightly more centered, i.e., some topics dominate more

¹⁷ <http://www.w3.org/TR/rdf-sparql-query/>

¹⁸ a tweet is mapped to its talk using s custom defined property <http://ontologies.hypios.com/rated#refersTo> that is a subproperty of skos:related.

significantly. This is expected bearing in mind that the retweet chains which emerge during conferences and the meme phenomenon that is inherent to Twitter.

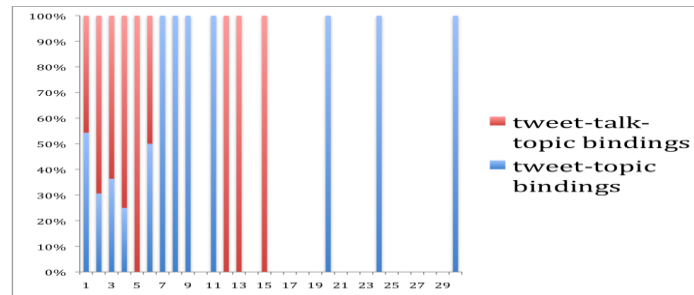


Figure 2. Histogram of number of tweets per topic when using different types of bindings

The histogram shown in Figure 2 visually sums up the above two tables. It shows how the distribution of topics cannot be bound to tweets directly for the tweet text (blue line), and that, conversely, the distribution of topics can be bound to tweets over the tweet-talk mappings (based on talk texts).

6 Application Scenarios

6.1 Conference Feedback

It is easy to imagine a system that provides feedback on the most popular topics based on those mined from tweets and also from topics mined from the talks that tweets refer to. In our case, those topics would correspond to those in Table 3 and Table 4. It is interesting to go further than this simple way of providing feedback by looking at the difference in topics that dominate tweets at a conference, and those that dominate in talks. An example in our case would be the topic Radio-frequency Identification that is popular in tweets, but not represented in talks. Such topics are the ones that people talk about in couloirs but are not represented in the official curriculum. We believe that in some cases such kinds of topic might indicate the need to extend the curriculum of future conferences with the topics people talk about, or detect future trends of interest based on regression analysis.

Another interesting possibility would be to look at the topics that appear in talks and rank them based on how many times the talks that mention them have been tweeted about. This way we could identify topics that drew attention and those that did not. Because of the incompleteness of our current human mappings we couldn't deliver such an analysis at this stage, but this feature is planned for the next version of our system that will perform the mappings automatically. If we look at the two tables of popular topics it is clear that some topics are similar and may be considered almost the same in a given context (e.g., Semantics and Semantic Web). Therefore it is also interesting to explore the semantic proximity of topic concepts and include this information in the analysis of popularity. This is one of the challenges for our future

work. Other future challenges include: associating sentiment information with the tweets and using it for the analysis of talk/presenter popularity, and mining conversations and memes in order to spot the most provocative talks.

6.2 Improving User Interest Profiling with Tweet-Talk Mappings

We believe that tweet-talk mappings might enable us to gain a greater insight into the real interests of users. Since tweets are short, it is difficult to really extract topics from them. On the other hand if we can detect the talks that the tweets refer to, the talk abstracts would enable more substantial topic extraction and thus allow user inference. To demonstrate this, we have used the human-generated mappings to connect users to more topics by propagating topics of the talks to the users over the tweets that serve as a glue. Table 5 demonstrates the case for the user Matthew Rowe, for whom we had 4 tweet-talk mappings. By accessing additional information in the form of talk abstracts, additional topics of interest are associated to the user. To gauge the accuracy of this propagation we asked the user to mark those topics that were found to be irrelevant – therefore enabling precision of the topic lists to be measured. For this single participant we yielded precision of 0.714 for topics derived from tweets and 0.632 derived from talks. These results of our first experiments motivate us to continue research in this direction as they show that there might be an additional value of tweet-talk mappings for the user interest profiling.

Table 5. Topics from tweets and talks that tweets are about for Matthew Rowe

<i>Topics from tweets</i>	<i>Topics from talks that the tweets are about</i>
SPARQL	Information
Linked_Data	Web_application
Russia	Walled_garden_(technology)
PageRank	Triplestore
Japan	Social_web
IPad	Social_network
England	Semantic_Web
	Resource_Description_Framework
	Relational_database
	Probabilistic_analysis_of_algorithms
	Privacy
	News_agency
	Nature
	Graph_(mathematics)
	Granularity
	Data

7 Related Work

Twitter has inspired the Social Semantic Web research community to seek methods to extract semantics from tweets. For instance Shinavier [11] proposes a framework that extracts and unifies the semantics from different nanoformats – syntaxes that allow embedding of meaning and references into short tweet messages. If talks and subevents had their specific references, such syntaxes might be applied for our use-cases related to conferences. An inherent disadvantage in comparison to our method is

the significant number of references that users have to use to refer to particular subevents, as well as the requirement to enforce their disciplined use.

Wagner [12] proposes the KASAS framework for analyzing awareness streams (including Twitter) and the contained concepts. This system is more general and does not intend to generate mappings from tweets to content and real life events in such a contextualized way as we do. Nevertheless this approach could nicely complement our system by providing the functionality that is currently covered by Zemanta, and hopefully generate more related concepts than those from DBPedia. Similarly, Twarq¹⁹ [13] provides an extraction facility for Twitter streams, providing basic Twitter metadata in RDF, with the addition of rich semantics of Twitter #hashtags. It would be interesting to see how this system could be combined with our tweet-talk mapping in the future. SMOB [14], the semantic microblogging tool, represents an alternative to Twitter that publishes tweets directly in semantic (RDF) form. It also provides DBPedia topics for all #hashtags used in the tweet. We could not use SMOB in our experiments because its user base is still insufficient, but in general this tool could be coupled with our tweet-subevent mapping method. Twitter analytic systems like 140kit.com are starting to emerge, but to our knowledge they still do not provide event/conference-oriented services.

8 Conclusions and Future Work

In this paper we have described motivations for extracting structured data from conference tweets. We have presented concrete use cases, and proposed a framework for Twitter conference data extraction. The key concept of our framework is to map tweets to the talks and subevents that they refer to, thereby gaining access to additional knowledge about event dynamics and user activities. Our initial experiments with human-provided tweet-talk mappings indicates that a mapping-based approach will provide significant benefits for the given use-cases (i.e., providing conference feedback, user profiling, digital memory and conference topic dynamics). However, the enormity of the Twittersphere, and the rate at which tweets are published, requires an automatic tweet-talk mapping technique. We have outlined how our technique will work, in an abstract sense, and have defined the task as a multi-class classification problem. The human-generated mappings, produced using the Delphi method, provide the necessary means through which we can assess the performance of our approach. In addition to tweets already published, we plan to make this gold-standard available to the community for use in the future.

One limitation of our approach is the reliance on existing Twitter archives, and the assumption that Linked Data is available about an event and its subevents. The former assumption is quite realistic, as people are already incentivised to create and share Twitter archives. The latter however is more difficult to achieve, but we believe that our work shows the numerous benefits of utilising Linked Data, and would motivate more event organisers to provide their event data as Linked Data. One also needs to acknowledge that our approach is not limited to the particular service Twitter.com, and is actually generalisable to other microblogging services.

¹⁹ <http://twarql.sourceforge.net/>

Acknowledgments. The work of Milan Stankovic has been partially funded by ANRT (French National Research Agency) under the grant number CIFRE N 789/2009. We are also grateful to kind people from Talis who have provided us with a free Linked Data triple store to put the data online. Zemanta was also kind to provide us with a generous API call limit that enabled us to do the topic enrichment of the tweets and talks.

References

1. Reinhardt, W., Ebner, M., Beham, G., & Costa, C. (2009). How people are using Twitter during conferences. *5th EduMedia conference*.
2. Letierce, J., Passant, A., Breslin, J., & Decker, S. (2010). Understanding how Twitter is used to spread scientific messages. Proceedings of the Web Science Conference, 26-27 April 2010, Raleigh, NC, USA,
3. Java, A., Song, X., Finin, T., & 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* (p. 56–65). ACM.
4. Asur, S., & Huberman, B. (2010). Predicting the Future With Social Media. *Arxiv preprint arXiv:1003.5699*. Retrieved from <http://arxiv.org/pdf/1003.5699>.
5. Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In *hicss-43. IEEE: Kauai, HI*.
6. Honeycutt, C., & Herring, S. C. (2009). Beyond Microblogging : Conversation and Collaboration via Twitter. *Methodology*.
7. Stankovic, M. (2008). Modeling Online Presence. In *Proceedings of the First Social Data on the Web Workshop* (Vol. 401). Karlsruhe, Germany
8. Rowe and Wright (1999): The Delphi technique as a forecasting tool: issues and analysis. *International Journal of Forecasting*, Volume 15, Issue 4, October 1999.
9. Fleiss, J. L. (1981) *Statistical methods for rates and proportions*. 2nd ed. (New York: John Wiley)
10. Berners-Lee, T. (2006). Linked Data Desing Principles. Retrieved from <http://www.w3.org/DesignIssues/LinkedData.html>.
11. Shinavier, J. (2010). Real-time #SemanticWeb in <= 140 chars. In *Proceedings of Linked Data on the Web 2010, on WWW2010*.
12. Wagner, C. (2010). Exploring the Wisdom of the Tweets: Towards Knowledge Acquisition from Social Awareness Streams. In *Proceedings of Extended Semantic Web Conference 2010, Heraklion, Crete*.
13. Mendes P.N., Kapanipathi, P., Passant A." Twarql: Tapping into the Wisdom of the Crowd,". Triplification Challenge 2010 at 6th International Conference on Semantic Systems (I-SEMANTICS), Graz, Austria, 1-3 September 2010.
14. Passant, A., Bojars, U., Breslin, J. G., Hastrup, T., Stankovic, M., Laublet, P., et al. (2010). An Overview of SMOB 2: Open, Semantic and Distributed Microblogging. In *4th International Conference on Weblogs and Social Media, ICWSM* (pp. 303-306).