TUMS: Twitter-Based User Modeling Service

Ke Tao, Fabian Abel, Qi Gao, and Geert-Jan Houben

Web Information Systems, Delft University of Technology {k.tao,f.abel,q.gao,g.j.p.m.houben}@tudelft.nl

Abstract. Twitter is today's most popular micro-blogging service on the Social Web. As people discuss various fresh topics, Twitter messages (tweets) can tell much about the current interests and concerns of a user. In this paper, we introduce TUMS, a Twitter-based User Modeling Service, that infers semantic user profiles from the messages people post on Twitter. It features topic detection and entity extraction for tweets and allows for further enrichment by linking tweets to news articles that describe the context of the tweets. TUMS is made publicly available as a Web application. It allows end-users to overview Twitter-based profiles in a structured way and allows them to see in which topics or entities a user was interested at a specific point in time. Furthermore, it provides Twitter-based user profiles in RDF format and allows applications to incorporate these profiles in order to adapt their functionality to the current interests of a user. TUMS is available via: http://wis.ewi.tudelft.nl/tums/

Keywords: user modeling, twitter, semantic enrichment, service.

1 Introduction

Applications that aim for personalization and would like to adapt their functionality to the current interests and demands of a user require information about their users [1]. User-adaptive systems suffer from cold-start and sparsity problems [2]. For example, when systems have to deal with new users or fresh content, those systems require user profile information that allows for estimating the interests of the users. In this paper, we present a Twitter-based user modeling service (TUMS) that exploits users' Twitter activities to infer semantically meaningful user profiles.

Since Twitter was launched in 2007, it became the most popular microblogging service, with 190 million users who post more than 65 million posts every day¹. As people are not limited to a certain domain, they can discuss various topics of which most of the topics are related to news which could also be found in mainstream news articles [3]. There is also an active research community studying Twitter in terms of topics like social network analysis [3,4], community and user influence mining [5,6,7], recommendations of URLs [8] or

¹ http://techcrunch.com/2010/06/08/twitter-190-million-users/

R. García-Castro et al. (Eds.): ESWC 2011 Workshops, LNCS 7117, pp. 269–283, 2012.

[©] Springer-Verlag Berlin Heidelberg 2012

sentiment analysis [9,10]. Inferring user profiles from individual Twitter activities is a hard problem as tweets are limited to 140 characters which makes the deduction of semantics difficult. Rowe et al. [11] propose to exploit the context in which tweets have been published. In particular, they propose to map tweets to events and exploit semantic descriptions of these events to clarify the semantic meaning of tweets. In previous work, we proposed strategies to automatically map Twitter messages to related news articles which allows us to exploit the news articles to enrich the semantics of individual tweets [12]. This enrichment builds the basis for the user modeling strategies [13] that are made available via the TUMS service which we present in this paper.

The TUMS service allows end-users to inspect Twitter-based profiles and enables other applications to re-use these profiles. People can overview their personal Twitter activities or profiles of other users to explore the topics those users were concerned with in the past. Entity-based, topic-based and hashtag-based tag clouds allow to further visualize the profiles. Visualizing the profiles is important for getting an insight into the Twitter activities: for example, individual people can thus become aware of what can be inferred from their Twitter activities and perhaps then reconsider how they publish tweets on Twitter. TUMS is also of interest for other services on the Social Web as it enables them to consume Twitter-based profiles in RDF format. Profiles can be used for personalization and are particularly interesting for other applications that suffer from sparsity problems (e.g. services that cannot collect sufficient data about their users) and services that are interested in "realtime" or very fresh profile information.

In the following section we summarize related work before we introduce TUMS in detail in Section 3. Then we outline the architecture of our service, the user modeling strategies featured by TUMS and present the graphical user interface as well as the service API. Insights on applying TUMS within an application that aims for personalization is given in Section 4. Finally, we conclude and give an outlook on our plans for future work in Section 5.

2 Related Work

With the advent of Semantic Web technologies and appropriate vocabularies such as FOAF [14], SIOC [15] or the Weighted Interest vocabulary [16], re-use of user profiles is becoming easier nowadays. Research on generic user modeling services [17], mediating user models [18], identifying users across system boundaries [19] and cross-system user modeling and personalization [20,21] further supports the re-use of user profiles in different application contexts. In this paper, we introduce a service that generates user profiles by exploiting Twitter and allows for applying these profiles in other applications.

Research on Twitter often focuses on analyzing large fractions of the Twitter network to study information propagation patterns [3,22,23] or to identify influential users [6,7]. Dong et al. [24] exploit Twitter to detect and rank fresh URLs that have possibly not been indexed by Web search engines yet. Lately, Chen et al. conducted a study on recommending URLs posted in Twitter messages

and compare strategies for selecting and ranking URLs by exploiting the social network of a user as well as the general popularity of the URLs in Twitter [8]. Yet, there exists little research on analyzing the semantics of individual tweets and exploiting Twitter as a source for modeling user interests. Rowe et al. [11] proposed to exploit contextual information to enrich the semantics of Twitter messages(tweets). In previous work, we followed this suggestion and linked tweets to related Web resources [12]. Based on this, we consider on the technical and semantical aspects of the services Given this semantic enrichment, we proposed user modeling strategies that allow for recommending news articles. In this paper, we make those enrichment and user modeling strategies available to the public and present TUMS, a Twitter-based User Modeling Service.

3 Architecture, Stategies, and Implementation

TUMS is a user modeling service that exploits the tweets of individual users to infer and provide user interest profiles. TUMS targets two types of consumers: (i) end-users who would like to overview and inspect their profiles (or profiles of friends) in a structured way and (ii) applications that would like to incorporate the RDF-based user profiles in order to adapt their functionality to the current interests of a user. In this section, we first present the architecture of TUMS; then we describe user modeling strategies in TUMS; finally, we present both graphical and machine-processable user profiles generated by TUMS.

3.1 Architecture

TUMS aims to to deliver both a human and a machine readable representation of the user characteristics that can be deduced from the posts a user published on Twitter. The only input required by TUMS is the Twitter username of the person that should be profiled. Figure 1 depicts the architecture of the TUMS service. TUMS is composed of three modules: (i) the *Crawler and backend*, (ii) the *User Modeling* module, (iii) the *Graphical User Interface and RDF Endpoint*.

Crawler and Back-end. To start using the TUMS functionality is easy as TUMS just requires a Twitter username as input. Given a username, TUMS initiates a pipeline for crawling, processing, storing and analyzing the Twitter data that is publicly available for the corresponding user. TUMS features a tweets and a news crawler: the tweets crawler aggregates Twitter posts of the given user and stores the tweets in the TUMS data repository; the news crawler continuously monitors traditional news media. At the moment, TUMS monitors 62 RSS feeds published by mainstream news publishers such as CNN, New York Times, and BBC and aggregates the full content of each news article by making use of Boilerpipe². Once a new username

http://code.google.com/p/boilerpipe/, a Java library that detects supplemental text (e.g. advertisements or menus) in a website and allows for the extraction of the main content.

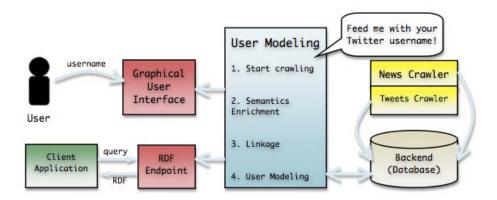


Fig. 1. The architecture of TUMS

is provided through either the graphical interface or the RDF endpoint, TUMS will continuously start monitoring the tweets of the corresponding user. Whenever a new tweet is observed, the tweet will get processed to enrich the semantics of the tweet (see below). Tweets, news articles and semantic metadata which is extracted from tweets and news, are stored in the data repository which builds the basis for user modeling.

User Modeling. Given the raw data crawled from Twitter, TUMS aims to infer user interest profiles that adhere to the Friend-Of-A-Friend (FOAF) [14] vocabulary and the Weighted Interest vocabulary [16]. Hence, given the raw text of tweets, TUMS will output profiles which specify how much a user is interested in a certain topic. The example output can be found in Section 3.4. In order to infer interests of a user from the tweets this user posted on Twitter, TUMS performs four main steps: (1) it notifies the tweet crawler to start collecting tweets, (2) it enriches the semantics of the user's tweets by categorizing the topic of a tweet and extracting hashtags and other entities (e.g. persons, locations) from the tweet, (3) if possible, it links the tweet to related news articles to further enrich the semantics of a tweet with semantics extracted from the news article and (4) it applies different user modeling strategies (see Section 3.2) to generate user profile information.

The first three steps will be executed whenever a new tweet is observed while the last step, the actual user modeling, is executed on query time, i.e. when a user (via the graphical user interface) or client application (via the Web service) requests profile information.

Graphical User Interface and RDF Endpoint. Given the data aggregated, processed and stored by the above modules, the Twitter-based user profiles can be retrieved by visiting the graphical user interface or by invoking the TUMS Web service API. The graphical interface enables end-users to overview their profiles by means of tag clouds (topic-based, hashtag-based and entity-based clouds) and diagrams (see Section 3.3). The REST-like Web service interface outputs user profiles in RDF using FOAF [14] and Weighted Interest vocabulary [16] as depicted in the above example.

design dimension	design alternatives
profile type	(i) hashtag-based, (ii) topic-based, or (iii) entity-based
weighting schemes	(i) term frequency (TF) or (ii) $TF \times IDF$
enrichment	(i) tweet-only-based enrichment or (ii) linkage and exploitation of external news articles (propagating entities/topics)
temporal constraints	(i) specific time period(s), (ii) temporal patterns (weekend, night, etc.), or (iii) no constraints

Table 1. Design space of Twitter-based user modeling strategies

3.2 User Modeling Strategies

The goal of TUMS and the user modeling strategies in particular is to generate user interest profiles that conform to the following model.

Definition 1 (User Profile). The profile of a user $u \in U$ is a set of weighted concepts where with respect to the given user u for a concept $c \in C$ its weight w(u,c) is computed by a certain function w.

$$P(u) = \{(c, w(u, c)) | c \in C\}, u \in U$$

Here, C and U denote the set of concepts and users respectively.

To generate such profiles, we developed a set of user modeling strategies [13] that vary in four design dimensions: (i) the type of profiles created by the strategies, (ii) the weighting scheme, (iii) the data sources exploited to further enrich the Twitter-based profiles, and (iv) temporal constraints that are considered when constructing the profiles (see Table 1). The generic model for profiles representing users is specified in Definition 1.

As listed in Table 1, TUMS allows for three types of profiles that differ with respect to the type of concepts $c \in C$ for which we specify an interest weight (see Definition 1): entity-based, topic-based, and hashtag-based profiles. For entity-based profiles, we differentiate the weights between the entities extracted from the tweets. The topic-based profiles are rather broad and abstract from the concrete content as we map the tweets to 19 static topics such as sports, politics, or music.

TUMS allows for different methods as weighting scheme w(u,c). For example, using term frequency (TF), the weight of a concept is determined by the number of Twitter activities in which user u refers to concept c. In a hashtag-based profile, w(u, #technology) = 5 means that u published five tweets that mentioned "#technology". Other weighting methods such as $TF \times IDF$ are possible as well and a detailed comparison of different weighting schemes is planed for future work. The resulting user profiles will be normalized so that the sum of all weights in a profile is equal to $1: \sum_{c_i \in C} w(u, c_i) = 1$. With p(u) we refer to P(u) in its vector space model representation, where the value of the i-th dimension refers to $w(u, c_i)$.



Fig. 2. The start page of TUMS in the graphical user interface

Tweets posted by a user u may refer to external resources. A user can explicitly link to other Web resources in her tweet or she could discuss topics and events that are, for example, discussed in mainstream news articles as well. TUMS aims to also take these external resources into account when constructing entity-based and topic-based user profiles (see semantic enrichment in Table 1). In particular, profiles are enriched with entities and topics extracted from news articles that are linked with tweets. In previous work, we have done the study on selecting appropriate news articles for enriching tweets [12]. We revealed that for tweets which do not explicitly link to external Web resources we can find related news articles that report about the same event as the tweet with a high precision of more than 70%.

Temporal constraints are considered as the fourth dimension of the profiles (see Table 1). By specifying temporal constraints, client applications can, for example, retrieve the latest profile of a user or a profile that is only based on Twitter activities which a user performed we revealed.

By selecting and combining the different design dimensions and alternatives, TUMS can exploit different user modeling strategies and thus generate different profiles, in the form of visualized charts and RDF triples.

3.3 Graphical User Interface of TUMS

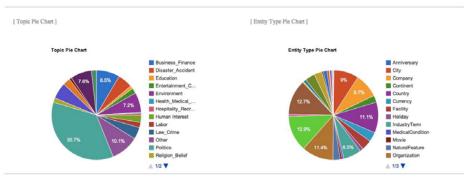
End-users can easily access TUMS via its graphical user interface which requires only a Twitter username as input. Figure 2 shows the start screen that is displayed when users are accessing TUMS. In the given example, the profile of a user named "USAGodG20" should be returned. By clicking on the "Get Tweet Profile" button the user will be directed to the next page that allows the user to overview and inspect the profile of "USAGodG20" (see Figure 3).

TUMS can represent user profiles by means of different visualizations, including hashtag-based clouds, pie charts visualizing topic-based profiles and entity types referenced by a user, and entity-based clouds (see Figure 3). In the following, we examine the implemented visualization features of TUMS and describe them by example results that we obtain for requesting the profile of the user "USAGodG20".

Hashtag-Based Clouds. The hashtag-based clouds represent the hashtag-based profile of a user and thus visualize the hashtags which the corresponding user

Hello, USAGodG20! Welcome back!

[hashing cloud]
peta est conspiracy teaparty no Globalization BUSH London patriot NWO MerryChristmas PayAt CFL nokill MerryChristmasEve NAU FF financial noki n 24 resistnet infowars insport Obama 2 toot auca 2 topp de



[Prime Minister] [President] [CNN] [United States] [USD] [China] [European Union] [Russia] [United Kingdom] [Washington] [BBC] [Japan] [White House] [New York] [Congress] [Iraq] [GBP] [Senate] [North Korea] [South Korea] [the New York] [JavaScript] [Beijing] [Official] [leader] [spokesman] [Islamic Republic of Iran] [David Cameron] [Barack Obama] [chairman] [London] [Italy] [Ireland] [Department of State] [Secretary OState] [correspondent] [Christmas] [Australia] [Pakistan] [America] [India] [California] [Obama administration] [food] [Twitter] [judge] [Middle East] [spokeswoman] [Republican Party] [New Year's Dayl [Florida] [lawyer] [football] [Captain] [Texas] [The Times] [The New York Times] [Facebook] [Google] [iPhone] [Chicago] [injury] [New York [Viv) [Qatari [Thanksaiving] [Bosson] [https://doi.org/10.1007]]

Fig. 3. Overview of user profile page of TUMS

mentioned in her Twitter messages. Different hashtags will be displayed in different font size depending on their weight within the hashtag-based profile (cf. w(u, c) in Definition 1). The higher the weight w(u, c) within the profile P(u), the bigger the font size.

For example, at the top of Figure 3 we see the hashtag cloud of "US-AGodG20". The visualization of the hashtag-based profile tells us that "US-AGodG20" is particularly interested in the news about globalization, NAFTA, and political leaders in the United States, e.g. Bush, Obama.

Topic-Based Piecharts. One of the reasons for the popularity of Twitter lies in the idea that anyone can contribute and discuss her opinion about any topic. TUMS generates topic-based profiles and the TUMS user interface provides a pie chart that displays the relative importance of a topic within a user profile. Topic-based profiles feature 19 different dimensions that correspond to broad topics like politics or education. Given the pie chart visualization of a topic-based profiles, we can see about which topics a user has published most.

The example topic-based piechart of USAGodG20 shows that 35.7% of her tweets are concerned with politics. This observation is consistent with the results we got from the hashtag-based cloud.

Entity Type-Based Piecharts. During the user modeling process, entities are extracted from tweets and external sources for enriching the semantic of tweets. TUMS extracts 39 types of the entities. The pie chart of entity types shows the importance of the entity types for a given user and can be inferred from the entity-based profiles.

The example pie chart generated based on the entity-based profile of USAGodG20 shows that USAGodG20 is often referring to persons (12.9%), organizations (11.4%), countries (11.1%) and cities (9%). Natural features (e.g. beach, ocean) or movies are less frequently mentioned.

Entity-Based Clouds. With entity-based clouds, TUMS visualizes the entity-based profile of a user and thus lists the top entities mentioned by a user. It refines the entity type based pie chart overview and shows particular persons, locations, etc. in which a user is interested. As for the hashtag-based cloud, entities are represented in different font size according to their weight w(u,c) in the entity-based profile (cf. Definition 1).

At the bottom of Figure 3 we see the most important entities within the entity-based profile of USAGodG20. Entities with larger font size in the cloud, such as "Prime Minister", "President", "United States", "China", and "European Union" are words that are quite common within the political domain which confirms again the findings from the above profile visualizations.

3.4 RDF Endpoint of TUMS

User profiles are furthermore made available via a Web service in RDF format. Via HTTP content negotiation³ or by specifying the format in the HTTP query string (format=rdf), client applications are enabled to request the different types of profiles generated by TUMS in RDF. The URL pattern that represents a query for profile information is defined as follows:

```
.../profile/[username]/[profile type]/[weighting]/[enrichment]/?[temporal constraints]
```

Hence, in addition to the *username*, which refers to the Twitter username of the person whose profile should be returned, there are four types of parameters which can be customized in the pattern. These four parameters correspond to the four dimensions in Table 1. The format and possible values of these parameters are explained below.

profile type. The profile type parameter enables applications to specify which type of profile they would like to retrieve: hashtag-based, topic-based or entity-based. Given a user profile $P(u) = \{(c_1, w(u, c_1)), (c_2, w(u, c_2)), ...\}$ where $c_i \in C$, the profile type thus specifies the type of the concepts c_i . While topics and entities are represented via a URI by nature and therefore explicitly describe the semantic meaning of a topic and entity respectively, we try to clarify the semantic meaning of hashtags by referring to $\#tagdef^4$. Given the three types of profiles, the corresponding parameter values that can be used in the HTTP request are the following: hashtag, topic, and entity. For example, the hashtag-based profile can be retrieved as follows:

.../profile/USAGodG20/hashtag/

The RDF-formatted response to this HTTP request might be the following:

³ http://tools.ietf.org/html/rfc2616#section-12

⁴ http://tagdef.com

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix wi: <http://purl.org/ontology/wi/core#> .
@prefix wo: <http://purl.org/ontology/wo/core#>
@prefix tums: <http://wis.ewi.tudelft.nl/tums/rdf/> .
@prefix dbpedia: <a href="http://dbpedia.org/resource/">http://dbpedia.org/resource/</a>> .
<http://twitter.com/USAGodG20>
 a foaf:Person;
 foaf:name "John of God";
 wi:preference [
     a wi:WeightedInterest;
     wi:topic dbpedia:Conspiracy_(political) ;
     wo:weight [
        a wo:Weight;
        wo:weight_value 0.0757 :
        wo:scale tums:Scale
     ];
  wi:preference [
     a wi:WeightedInterest;
     wi:topic dbpedia:Globalization;
     wo:weight [
        a wo:Weight;
        wo:weight_value 0.1576;
        wo:scale tums:Scale
     ] .
```

Hence, the hashtag-based profile of USAGodG20, which is in the graphical user interface represented by means of a tag cloud (see above), is represented using FOAF [14] and the Weighted Interest vocabulary [16]. For each concept(wi:topic), which is represented by a DBpedia URI, in the profile, the weight is clearly specified using $wo:weight_value$. TUMS normalizes the weights in a profile vector so that the sum of all weights in the profile is equal to 1. Correspondingly, the scale (tums:Scale) on which the weights are specified ranges from 0 to 1 ($wo:min_weight = 0$, $wo:max_weight = 1$, cf. Weighting Ontology⁵).

weighting schemes. TUMS supports different weighting schemes for computing the weights in a user profile (cf. w(u,c) in Definition 1). At the moment, TUMS supports (i) term frequency (TF) and (ii) TF multiplied by inverse document frequency $(TF \times IDF)$. The corresponding parameters are tf and tfidf so that a request may have the following format:

```
.../profile/USAGodG20/hashtag/tf/
```

enrichment. Currently, the possible options for this parameter are: (i) tweet or (ii) tweetnews. By default, TUMS extracts semantics from tweet. However, developers can specify whether external sources should also be exploited to further enrich the semantics of tweets when constructing the profile (tweetnews).

temporal contraints. Temporal constraints of profiles can be specified in the URL-based query as well. For specifying a certain time period, developers can use start and end to specify the start and end of the period that should be considered when generating the profiles. Moreover, it is possible to specify pattern-like temporal constraints like weekend or night. They can be

⁵ http://purl.org/ontology/wo/core

specified by adding further parameters to the query string. At the moment, TUMS allows for a wk parameter which can be weekend (Twitter activities on Saturdays or Sundays) or weekday (Monday till Friday) and a dk parameter which can be night (from 6pm till 6am) or day (from 6am till 6pm). An example query combing these two features look like follows:

 $\dots/hashtag/?start=2010-12-01\&end=2011-02-28\&wk=weekend\&dn=night$

4 Application of TUMS

TUMS can be adopted by developers that aim at personalized applications. As Twitter allows for capturing daily activities, TUMS is of particular benefit for applications such as news recommendation systems that require information about the current concerns of a user. Moreover, it is beneficial for systems that suffer from sparsity problems such as the cold-start problem.

4.1 Developing Personalized Applications with TUMS

The cold-start problem is one of the issues that every personalized application developers have to solve. If a new user registers to the system then there is usually no or little profile information available. Therefore, it becomes difficult to provide personalization right after a new user registers to the system. Like stumbleupon⁶, systems could decide to ask end-users to fill a long list of topic interests for starting knowledge. Contrary to this, TUMS provides a solution that enables systems to automatically obtain user interest profiles without requiring any further user interaction. Various personalized applications can be developed based on the TUMS service. For example, a recommender can easily know about a user with the generated profile in order to provide fresh contents like news articles, video clips, pictures, etc.

Liu et al. [25] analyzed a content-based recommender for Google News. They showed that interests in news topics such as technologies, politics, etc. change over time. Hence, information about the latest interests of a user is essential for personalized systems such as news recommender. As TUMS keeps collecting users' tweets in real-time, the profile can be generated based on the latest information.

End-users might ask themselves for the reason of such recommended items. Since the visualization of the profiles is already provided by TUMS, it can be applied to explain why recommendations have been generated. Such explanations may foster acceptance of recommendations. For example, for some e-shopping websites like Amazon⁷ it is common to give an explanation why a product was recommended. Using TUMS, recommender systems can link to the (particular fragment of the) user profile that they used in order to compute the recommendation.

⁶ http://www.stumbleupon.com/

⁷ http://www.amazon.com

4.2 Personalized News Article Recommendations with TUMS

Let us consider the following scenario: Bob is a PhD student majoring in Semantic Web and interested in cycling during the weekend. He is a passioned Twitter user and a news junkie. First, let's look at what Bob has posted on Twitter. Here are two of his tweets:

- Reading interesting papers about a service aiming at helping personalized application developers to do user modeling based on twitter.
- Cycling day http://bit.ly/m71kUb

It would be great if a news recommending application can know Bob's interests from these tweets. However, without further processing these tweets are still in plain text. With TUMS, personalized application can benefits from these tweets, as the information in plain text can be enriched with more concepts not only from the tweets themselves, but also the external resources which can be refered from the URLs in the tweets. The tweets enriched by TUMS look as follows:

```
@prefix sioc: <http://rdfs.org/sioc/spec/>
@prefix dcterms: <http://purl.org/dc/terms/>
@prefix dbpedia: <http://dbpedia.org/resource/> .
<http://twitter.com/#!/Bob/statuses/40435486531137536> a <sioc:Post> ;
     dcterms:created "2011-02-23T15:39:06+00:00";
     sioc:content "Reading interesting papers about a service aiming at
            helping personalized application developers to do user modeling
             based on twitter. "@en ;
     sioc:topic dbpedia:Personalization;
     sioc:topic dbpedia:Twitter;
     sioc:topic dbpedia:Service-oriented_architecture;
     sioc:has_creator <http://twitter.com/#!/Bob/>;
<http://twitter.com/#!/Bob/statuses/42513338076381184> a <sioc:Post> ;
     dcterms:created "2011-02-27T19:46:06+00:00" :
     sioc:content "Cycling day http://bit.ly/m71kUb"@en ;
     sioc:topic dbpedia:Cycling;
     sioc:topic dbpedia:Segregated_cycle_facilities;
     sioc:topic dbpedia:Mountain_bike;
     sioc:has_creator <http://twitter.com/#!/Bob/> ;
     sioc:links_to <a href="http://bit.ly/m71kUb">http://bit.ly/m71kUb">http://bit.ly/m71kUb</a>.
```

Base on these enriched tweets of Bob, the semantically meaningful profiles can be generated by TUMS:

```
@prefix foaf: <a href="http://xmlns.com/foaf/0.1/">
@prefix wi: <a href="http://purl.org/ontology/wi/core#">
@prefix wo: <a href="http://purl.org/ontology/wo/core#">
@prefix wo: <a href="http://whitp://wwi.udelft.nl/tums/rdf/">
@prefix wo: <a href="http://whitp://www.nudelft.nl/tums/rdf/">
@prefix wo: <a href="http://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://whitp://wh
```

```
wo:scale tums:Scale
    ]
];
wi:preference [
    a wi:WeightedInterest;
wi:topic dbpedia:Cycling;
wo:weight [
    a wo:Weight;
    wo:weight_value 0.1413;
    wo:scale tums:Scale
    ]
].
```

Since Bob was doing research about personlization on weekdays, and cycled during the weekend, the profile above includes the entities of *Personalization* and *Cycling*. In the case that it is Sunday evening, the Women's World Cup is dominating the news, the Wimbledon final and the Tour de France are also ongoing, a news recommender can consider to deliver some personalized news articles. Then TUMS can help as the weekend profile can be retrieved. And the profile is understandable for the application because it is described in RDF:

```
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix wi: <http://purl.org/ontology/wi/core#> .
@prefix wo: <http://purl.org/ontology/wo/core#>
@prefix tums: <http://wis.ewi.tudelft.nl/tums/rdf/> .
@prefix dbpedia: <http://dbpedia.org/resource/> .
<http://twitter.com/#!/Bob>
 a foaf:Person;
  foaf:name "Bob Green";
  wi:preference [
     a wi:WeightedInterest;
    wi:topic dbpedia:Cycling;
    wo:weight [
       a wo:Weight;
       wo:weight_value 0.1729;
       wo:scale tums:Scale
       1
    ];
  wi:preference [
    a wi:WeightedInterest;
     wi:topic dbpedia:Segregated_cycle_facilities;
    wo:weight [
       a wo:Weight;
       wo:weight_value 0.0823;
       wo:scale tums:Scale
       ]
     ] .
```

The concept *Cycling* in the weekend profile, is the word in the tweet that Bob posted on Sunday; while the concept of *Segregated cycle facilities* (a.k.a. bike lanes) comes from analyzing the external resources. TUMS analyzed the content of the blog post with the URL of "http://bit.ly/m71kUb" in Bob's tweet, and found that concept. Based on the weekend profile, recommedations can be made by computing and ranking the similarity between Bob's profile and the concepts in the news articles. Then, when Bob was going to use the news recommender application in a weekend, his weekend profile will be adopted to make the recommendation. As a result, a news article with the title of "Cycling-Tour rivals profit from Contador's misfortune", which was about the Tour de France, can be

selected for him to read. In the meanwhile, Bob can also figure out why this news article is recommended to him through the indication of the concepts matched concepts in both his tweets or external resources mentioned in them, and the content of the news.

5 Conclusion and Future Work

In this paper, we present TUMS, a Twitter-based User Modeling Service, that generates semantic user profiles by exploiting tweets. Based on functionality for enriching the semantics of tweets, TUMS features a variety of user modeling strategies that produce entity-based, topic-based or hashtag-based profiles. TUMS makes these profiles available to other applications via an RDF endpoint and allows end-users to explore profiles visually. Given a Twitter username, TUMS aggregates tweets from Twitter and starts monitoring the Twitter user. All tweets published by the user will be processed by the semantic enrichment module of TUMS which, for example, extracts entities from tweets and links tweets to related external Web resources to further enrich the semantics of a tweet. Given the semantically enriched tweets, TUMS can be queried to return semantic profiles for specific periods in time.

TUMS is available online and ready for being used by platforms that aim for personalization based on users' Twitter activities. Evaluations regarding the characteristics and quality of profiles that are generated by TUMS are reported in [13] and show, for example, that entity-based user profiles generated by TUMS allow for high precision when recommending news articles. Our ambition for future work is to further investigate for what other types of personalization tasks what type of TUMS profiles are appropriate so that we can further advice developers who are using TUMS how they can customize TUMS user modeling strategies to optimize their personalization quality.

Acknowledgements. This work is partially sponsored by the EU FP7 projects ImREAL (http://imreal-project.eu) and GRAPPLE (http://grapple-project.org).

References

- 1. Jameson, A.: Adaptive interfaces and agents. The HCI handbook: fundamentals, evolving technologies and emerging applications, pp. 305–330 (2003)
- Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.): Adaptive Web 2007. LNCS, vol. 4321.
 Springer, Heidelberg (2007)
- 3. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: Proceedings of the 19th International Conference on World Wide Web (WWW 2010), pp. 591–600. ACM, New York (2010)
- 4. Krishnamurthy, B., Gill, P., Arlitt, M.: A few chirps about twitter. In: Proceedings of the first workshop on Online social networks. In: WOSP 2008, pp. 19–24. ACM, New York (2008)

- 5. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis. WebKDD/SNA-KDD 2007, pp. 56–65. ACM, New York (2007)
- Weng, J., Lim, E.P., Jiang, J., He, Q.: Twitterrank: finding topic-sensitive influential twitterers. In: Davison, B.D., Suel, T., Craswell, N., Liu, B. (eds.) Proceedings of the Third International Conference on Web Search and Web Data Mining (WSDM 2010), pp. 261–270. ACM, New York (2010)
- Cha, M., Haddadi, H., Benevenuto, F., Gummadi, P.K.: Measuring User Influence in Twitter: The Million Follower Fallacy. In: Cohen, W.W., Gosling, S. (eds.)
 Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM 2010). The AAAI Press, Washington, DC, USA (2010)
- Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems (CHI 2010), pp. 1185–1194. ACM, New York (2010)
- 9. Gaffney, D.: #iranElection: quantifying online activism. In: Proceedings of the WebSci10: Extending the Frontiers of Society On-Line (2010)
- Diakopoulos, N.A., Shamma, D.: Characterizing debate performance via aggregated twitter sentiment. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010, pp. 1195–1198. ACM, New York (2010)
- Stankovic, M., Rowe, M., Laublet, P.: Mapping Tweets to Conference Talks: A Goldmine for Semantics. In: Passant, A., Breslin, J., Fernandez, S., Bojars, U. (eds.) Workshop on Social Data on the Web (SDoW 2010), co-located with ISWC 2010, Shanghai, China, vol. 664. CEUR-WS.org (2010)
- 12. Abel, F., Gao, Q., Houben, G.J., Tao, K.: Semantic Enrichment of Twitter Posts for User Profile Construction on the Social Web. In: García-Castro, R., et al. (eds.) ESWC 2011 Workshops. LNCS, vol. 7117, pp. 269–283. Springer, Heidelberg (2011)
- Abel, F., Gao, Q., Houben, G.J., Tao, K.: Analyzing User Modeling on Twitter for Personalized News Recommendations. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 1–12. Springer, Heidelberg (2011)
- 14. Brickley, D., Miller, L.: FOAF Vocabulary Specification 0.91. Namespace document, FOAF Project (November 2007), http://xmlns.com/foaf/0.1/
- Bojars, U., Breslin, J.G.: SIOC Core Ontology Specification. Namespace document, DERI, NUI Galway (January 2009), http://rdfs.org/sioc/spec/
- Brickley, D., Miller, L., Inkster, T., Zeng, Y., Wang, Y., Damljanovic, D., Huang, Z., Kinsella, S., Breslin, J., Ferris, B.: The Weighted Interests Vocabulary 0.5. Namespace document, Sourceforge (September 2010), http://purl.org/ontology/wi/core#
- 17. Kobsa, A.: Generic user modeling systems. User Modeling and User-Adapted Interaction 11(1-2), 49–63 (2001)
- Berkovsky, S., Kuflik, T., Ricci, F.: Mediation of user models for enhanced personalization in recommender systems. User Modeling and User-Adapted Interaction (UMUAI) 18(3), 245–286 (2008)
- 19. Carmagnola, F., Cena, F.: User identification for cross-system personalisation. Information Sciences: an International Journal 179(1-2), 16–32 (2009)
- Abel, F., Henze, N., Herder, E., Krause, D.: Interweaving Public User Profiles on the Web. In: De Bra, P., Kobsa, A., Chin, D. (eds.) UMAP 2010. LNCS, vol. 6075, pp. 16–27. Springer, Heidelberg (2010)

- Mehta, B.: Learning from What Others Know: Privacy Preserving Cross System Personalization. In: Conati, C., McCoy, K., Paliouras, G. (eds.) UM 2007. LNCS (LNAI), vol. 4511, pp. 57–66. Springer, Heidelberg (2007)
- Sakaki, T., Okazaki, M., Matsuo, Y.: Earthquake shakes twitter users: real-time event detection by social sensors. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010, pp. 851–860. ACM, New York (2010)
- Lerman, K., Ghosh, R.: Information contagion: an empirical study of spread of news on digg and twitter social networks. In: Proceedings of 4th International Conference on Weblogs and Social Media, ICWSM (May 2010)
- 24. Dong, A., Zhang, R., Kolari, P., Bai, J., Diaz, F., Chang, Y., Zheng, Z., Zha, H.: Time is of the essence: improving recency ranking using twitter data. In: WWW 2010: Proceedings of the 19th International Conference on World Wide Web, pp. 331–340. ACM, New York (2010)
- Liu, J., Dolan, P., Pedersen, E.R.: Personalized news recommendation based on click behavior. In: Rich, C., Yang, Q., Cavazza, M., Zhou, M.X. (eds.) Proceeding of the 14th International Conference on Intelligent User Interfaces (IUI 2010), pp. 31–40. ACM, New York (2010)
- 26. Passant, A., Hastrup, T., Bojars, U., Breslin, J.: Microblogging: A Semantic Web and Distributed Approach. In: Bizer, C., Auer, S., Grimnes, G.A., Heath, T. (eds.) Proceedings of the 4th Workshop Scripting For the Semantic Web (SFSW 2008) co-located with ESWC 2008, Tenerife, Spain, vol. 368. CEUR-WS.org (2008)