Twitter User Modeling and Tweets Recommendation based on Wikipedia Concept Graph *

Chunliang Lu and **Wai Lam**
Dept. of Systems Engineering & Engineering Management
The Chinese University of Hong Kong

Yingxiao Zhang
Dept. of Information Engineering
The Chinese University of Hong Kong

Abstract

As a microblogging service, Twitter is playing a more and more important role in our life. Users follow various accounts, such as friends or celebrities, to get the most recent information. However, as one follows more and more people, he/she may be overwhelmed by the huge amount of status updates. Twitter messages are only displayed by time recency, which means if one cannot read all messages, he/she may miss some important or interesting tweets.

In this paper, we propose to re-rank tweets in user's timeline, by constructing a user profile based on user's previous tweets and measuring the relevance between a tweet and user interest. The user interest profile is represented as concepts from Wikipedia, which is quite a large and inter-linked online knowledge base. We make use of Explicit Semantic Analysis algorithm to extract related concepts from tweets, and then expand user's profile by random walk on Wikipedia concept graph, utilizing the inter-links between Wikipedia articles. Our experiments show that our model is effective and efficient to recommend tweets to users.

Introduction

Nowadays, social networks, such as Facebook and Twitter, are becoming part of life. Since the launch in July 2006, Twitter has gained much success in past few years, with over 140 million active users as of 2012. As a microblogging service, Twitter allows user to post short messages (known as *tweets*), up to 140 characters long, to update public timeline of oneself. A user can get recently updated tweets from people who he/she is following, in the form of timeline after logging in. Generally, users would like to follow people who share common interests. Those who follow a user are commonly referred as *follower*s and those whom a user follows are called *followee*s. One can post tweets from website interface, mobile phone, email, or instance message. Users make use of Twitter to get news articles, read friends' updates, and even chat with each other. As stated in Twitter's blog, more than 340 millions of tweets are created per day by March 2012 [1]. However, one may face too many status updates from his/her followees. Although Twitter introduced hashtags to label tweets, large portion of tweets are posted without hashtags. Currently there is no effective way to filter and rank relevant tweets for each user. Effectively selecting information from large amount of tweets becomes a challenge.

Previous recommendation systems mainly focus on news recommendation. They use user browser history information and search query logs to build user profiles. These approaches face the problem of availability of past user browsing information or search log data. They also have the cold start problem. We notice that in social network, especially in Twitter, many users have created large amount of messages that imply their interest, which are precise and readily available source to construct user interest profiles. In this paper, we investigate a model to make use of this information to build user profiles. Such a user profile can help to reorganize a user's timeline by relevance to the user interest. We propose a model to map a tweet and a user to Wikipedia-based concepts. Wikipedia is currently the world's largest knowledge resource, featuring more that three million articles. It is actively updated by tens of thousands volunteers. Each Wikipedia article describes a single topic, with a succinct, well-formed title. In addition to text, Wikipedia also has rich information about the relationships between different articles, in the form of category pages, infoboxes, and inter-links. After we have constructed Wikipedia-based representation for user interests and tweets, we can measure the relevance of each tweet and then rank all tweets by the user interest.

The contributions of this paper are threefold. First, we develop a general model to extract a user interest profile given a tweet, based on Wikipedia concepts, and then expand user's interest by random walk on the Wikipedia concept graph. Second, we develop a model that can rank each tweet by relevance to interest and affinity between users. Finally, we have conducted extensive experiments and the results demonstrate the effectiveness of our model and its superiority to an existing approach using term frequency and inverse document frequency (TF-IDF), which is also em-

*The work is supported by grants from the Research Grant Council of the Hong Kong Special Administrative Region, China (Project Code: CUHK413510). This work is also affiliated with the Microsoft-CUHK Joint Laboratory for Human-centric Computing and Interface Technologies.

[1] http://blog.twitter.com/2012/03/twitter-turns-six.html

ployed in the state-of-the-art method presented in (Chen et al. 2010).

This paper is organized as follows. In the next section, we review some related work in user profiling, recommendation system, and social network analysis. Then, we present our framework in a section, followed by the experiment design and result analysis. The last section gives our conclusions and future work.

Related Work

User profiling and content recommendation system have drawn attention of researchers for many years since the emergence of the Internet. Many recommendation applications have emerged, including news, RSS feeds, social network messages, etc. (Hill and Terveen 1996) considered the problem of Usenet news messages filtering, by using the frequency-of-mention. (Sugiyama, Hatano, and Yoshikawa 2004) constructed a user profile to filter search result using a standard information retrieval system. (Billsus and Pazzani 1999; Schouten et al. 2010; Grčar, Mladenič, and Grobelnik 2005) focused on news recommendation, based on the user profile extracted from user reading history. (Kim et al. 2010) used a user profile to do personalized web search. (Tang et al. 2010) built profiles for researchers, making use of various information sources over the web.

All the profiling and recommendation systems need to gather information about users. (Grčar, Mladenič, and Grobelnik 2005) constructed a user profile based on browsing history. (Ahmad Wasfi 1999) introduced the idea of entropy of a page being accessed to determine its interestingness. (Latif et al. 2010) located the researcher information from linked data, and constructed researcher profiles combining information from several sources. (Geyer et al. 2008) used self-descriptions in online user profiles, such as "Hobbies", to construct user the interest profile. (Abel et al. 2011b; 2011a) used Twitter messages to construct a user profile and then recommend news based on that profile. They compared performance using different user modeling strategies, such as entity-based, topic-based profile, and also tried to extract information from news mentioned in Twitter message to enrich the profile.

Our study focuses on extracting user interest information from tweets. (Java et al. 2007) studied the usage of Twitter, and showed that the main usage of Twitter is to talk about daily activities and to seek or share information. (Naaman, Boase, and Lai 2010) studied the types of tweets user posted on Twitter. (Laniado and Mika 2010) explored the usage of hashtags and proposed using hashtags to organize tweets message. (Li et al. 2010) studied the problem of extracting meaningful keywords form short and informal social messages. They defined six features and used four different supervised machine learning algorithms, doing experiments on Facebook data. (Hannon, Bennett, and Smyth 2010) collected all the user's tweets and used Lucene platform[2] to measure the similarity between a tweet and a user profile. (Chen et al. 2010) is the most relevant research to ours. They implemented several models, based on content information as well as social information, to do the tweets recommendation. We also make use of this information, however, we develop a new model by considering such information to construct a large ontology graph.

Over the past years, researchers proposed various models to represent user profiles. (Gauch et al. 2007) provided a good review on profile construction. (Ramanathan and Kapoor 2009) created hierarchical user profiles using Wikipedia category information. (Nanas, Uren, and De Roeck 2003) built a concept hierarchy representation of a user profile. They tried to identify the most appropriate terms to build the concept hierarchy, and used the span of contiguous words to associate the concepts into a connected graph. (Kim et al. 2010) also constructed a concept network-based user profile, making use of Open Directory Project. (IJntema et al. 2010) compared different approaches to measure the similarity on ontology. (Ahn et al. 2007) studied the problem whether we should enable users to modify their profiles by themselves. They claimed that the ability to edit user profiles may harm the systems performance.

Besides academic research, there are some start-up companies that offer the service to filter and recommend tweets, such as my6sense.com, MicroPlasz.com and feedafever.com. Facebook also introduced the EdgeRank algorithm to rank user's status updates. However, none of them disclose their algorithms, or their performance.

[2]http://lucence.apache.org

Model Description

In this section, we present a new model to build a user profile by analyzing the tweets of the user, making use of a Wikipedia concept graph. We begin with an overview of our framework, and then describe each component in detail.

Motivation

We make use of the following observations:

- If the user A is following the user B and the user A has lots of interactions with B, then A is interested in B's interests.
- If the user A is interested in the concept α, then A will be interested in the concepts that are closely related to the concept α.

The first observation is intuitive. People follow others to get updated information (Java et al. 2007). We can measure the interest score by considering the affinity between users. The second observation considers the dynamic property of user's interests. For example, if someone is a big fan for Apple, he/she may be interested in Apple's new products, even though the name of the new product has not been in his/her previous interest profile. Wikipedia provides rich information about the relatedness of two concepts, and we will take advantage of this to construct user' profile.

Tweets Representation

To map a Twitter message to a set of concepts, we employ Explicit Semantic Analysis (ESA) model proposed in

(Gabrilovich and Markovitch 2007). ESA is currently an effective algorithm for computing the relatedness of texts. The original use of ESA is to compute semantic relatedness between two text fragments. For any given text, ESA will find a weighted Wikipedia concept vector to represent the text.

First we process each Wikipedia concept by representing it as a vector of words that are contained in the corresponding article page. The weight of each word is calculated as term-frequency inverse-document-frequency (TF-IDF) score, which is the strength of the relatedness between the word and the corresponding concept. Next we represent each tweet as a vector of words contained, also weighted as TF-IDF score. Thus we can compute the related Wikipedia concepts for a tweet by computing the similarity between two vectors. To speed up the process, we build an inverted index for Wikipedia, map each word into a list of Wikipedia concepts and the associated weights.

Besides concepts related to a tweet, we also incorporate the information about the users who are involved. This include users who post the tweet, and users who are mentioned or replied to. We treat all users as equal weight.

Thus for each tweet t, we can use the following representation.

$$T_C = \{(c_i, w_i)\}_i, T_U = \{(u_j, 1)\}_j \tag{1}$$

where T_C represents the relevant concepts, T_U represents the users involved. Each Wikipedia concept c_i corresponds the weight w_i computed from ESA algorithm, while the relevant users u_j all have equal weight.

The reason that we include the involved users in one tweet is as follows. Some topics may not be directly related to one's interests. However, as more and more friends talk about it, this may draw one's attention due to its popularity. This is especially true for breaking news.

User Profile

A user profile consists of two separate parts, namely, a concept vector representing user interested topics, and another vector representing the affinity with other users.

To construct a interest concept vector, we first get all tweets of the user and extract relevant concepts using ESA algorithm. Then we add up all the concepts and corresponding weights, resulting a large concept vector. We get user interest concept vector after normalization. For each user, we have

$$I_C = \{(c_i, w_i)\}_i \tag{2}$$

where c_i is one of Wikipedia concepts, and w_i is the corresponding weight respectively.

For the user affinity vector, we consider explicit interactions between users. Besides the follower/followee information, we measure the affinity between users by counting the number of tweets that reply, retweet, or mention between two users. We normalize the weight for each user. For each user, we have

$$I_U = \{(u_j, w_j)\}_j \tag{3}$$

where u_j and w_j represent a followee and the corresponding weight.

Random Walk on Link Graph

To make use of the Wikipedia link graph, previous research simply used the category information. However, according to our observation, the category information is not that important for constructing user interest profile. For example, if one is interested in "iPhone", he/she may also be more interested in other products from Apple, rather than various mobile phones under the same category "Smartphones".

Wikipedia articles have rich inter-link information, which is a good indicator of relatedness. In order to get the related concepts, we will run random walk on the Wikipedia graph. The intuition behind using Markov random walk is as follows. If another concept is mentioned in Wikipedia, then these two concepts have some kind of connections. For the previous example, in the Wikipedia page of *iPhone*, there are multiple links to the page of *Steve Jobs*. People interested in *iPhone* mostly will be interested in tweets that mention *Steve Jobs* as well.

We can regard all Wikipedia concepts as a link graph $G = (C, E)$, where $C = \{c_i\}_{i=1}^n$ is the concept articles in Wikipedia, and n is the total number of concepts. Let $\mathbf{W}$ represent an $n \times n$ weight matrix, in which w_{ij} is the number of link counts between the concepts c_i and c_j.

We define the transition probability $P_{t+1|t}(c_k|c_j)$ from the concept c_j to c_k by normalizing the outlinks of the concept c_j, as in the following equation

$$P_{t+1|t}(c_k|c_j) = w_{jk} / \sum_i w_{ji} \tag{4}$$

where i ranges all the concepts connecting to c_j. The one-step transition probabilities can be written as $\mathbf{P} = [P_{t+1|t}(w_k|w_j)]$, of size $n \times n$. And we can calculate $\mathbf{P}$ as

$$\mathbf{P} = D^{-1/2}S, \text{where } D = diag(WW^T) \tag{5}$$

Suppose we get an initial vector representation of user interest as $I = \{(c_i, w_i)\}_{i \in L}$, extracted from user's tweets. The initial vector v^0 is a n-dimensional vector with values

$$p_0(c_j) = \begin{cases} w_j, & \text{if } j \in L \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

Given the teleport probability $\alpha \in [0, 1)$, we repeatedly compute $v = \alpha P^T v^0 + (1-\alpha)v^0$ until convergence. In our experiments, we terminate the walk when the vector converges with an L_1 error of 0.001, and we restrict to 30 iterations since it will generally give reasonable result. We also set $\alpha = 0.15$ according to previous experience.

Ranking Algorithm

The ranking score can be regarded as the relevance measure between a specific tweet and the user interest profile. Since the user profile and the tweet are all represented as a set of weighted Wikipedia concepts and related users, we can directly compute the similarity between these two vectors.

In our model, we use cosine similarity to compute the score.

$$\text{Score} = \frac{I_C \cdot T_C}{\|I_C\| \|T_C\|} \cdot \frac{I_U \cdot T_U}{\|I_U\| \|T_U\|} \tag{7}$$

where I_C, I_U is the user interest profile, and T_C, T_U is the tweet vector representation. As seen in the above equation, we incorporate both the content relevance and affinity between users.

Experiment

For our experiments, we have implemented the following three models.

- **WikiProfile-RW**: This is our full model, which uses Wikipedia concepts to model user interests, and random walk to expand interest profile.
- **WikiProfile**: Our model without random walk.
- **TFIDF**: A method derived from term frequency and inverted document frequency (TF-IDF). This approach was also investigated in the state-of-the-art algorithm proposed in (Chen et al. 2010). It uses the TF-IDF weighting with the words in user's tweets.

For the TFIDF model, we randomly selected 10,000 tweets to compute inverse document frequency. We used all user's tweets and computed the TF-IDF vector as user interest profile. To rank the similarity between a user and a tweet, we compute the TF-IDF vector for the tweet and employ cosine similarity.

Data Sets

Using Twitter's API, we crawled over 20 thousands of users and 6 million tweets. Twitter constrains that for each user, we can only crawl his/her last 3200 tweets. However, this is sufficient for our experiments. To generate the dataset, first we randomly selected 200 users from the users we crawled. Then we manually inspected the user according to the following criteria.

- Only users who posted English tweets are considered.
- Users should have posted at least 100 tweets. We want to target relative active users.
- Contains at least 20 retweets or replies. We use this information to measure the performance of various models.
- Should be personal account, not company or organization. Since we mainly focus on user modeling, we will ignore such accounts.

After we chose the set of users in the dataset, we selected latest 1000 tweets from the user's timeline. To reduce the complexity, we only considered tweets posted by one's followees. However, our model can be applied to all public tweets as well. The later case can be used to recommend new accounts and tweets relevant to one's interest.

Twitter provides several functions to response to each tweet. One can reply a tweet, or retweet a tweet. Both will create a new tweet in user's own public timeline. Users can also mark some tweets as favorites, to recognize awesome tweets or save it for reading later. Twitter provides an API to get all these information. And in our experiments, we regarded the favorite/retweeted/replied tweets as relevant tweets. In case there were not enough relevant tweets, we included some of user's own tweets as relevant.

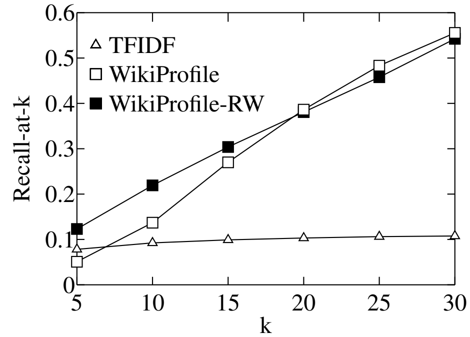

Figure 1: The recommendation performance measured by recall-at-k

For Wikipedia part, we downloaded the dump[3] of February 11, 2012, which contains 5,877,913 pages in total. After removing stop words and stemming, we imported all the necessary tables into MySQL database, and built the required inverted index using Apache Lucene.

Evaluation Metrics

We computed three quality measures: recall-at-k, precision-at-k, and average hit-rank. For a given user, we give a ranked list of tweets by the similarity between tweets and the user profile. Suppose the total number of relevant tweets is h in the top-k items, out of the total number of interest items $n_T(u)$ in test data set. Then the recall-at-k and precision-at-k for u are

$$\text{recall-at-k} = \frac{h}{n_T(u)} \tag{8}$$

$$\text{precision-at-k} = \frac{h}{k} \tag{9}$$

$$\text{average hit-rank} = \frac{1}{n_T(u)} \cdot \sum_{i=1}^{h} \frac{1}{rank_i} \tag{10}$$

where $rank_i$ is the position of the true interest tweets in the recommendation list.

Result Discussion

We varied k from 5 to 30, and collected the corresponding measures. Figures 1-3 show the averages of recall-at-k, precision-at-k and average hit-ranks.

We can see that our proposed model outperforms the TFIDF model for precision-at-k and recall-at-k. For the recall-at-k, we can see that our full model WikiProfile-RW gives more correct predictions on small k, and then more or less equal to the WikiProfile model. It means that we can get more relevant recommendations. However, for the precision-at-k, the WikiProfile model performs slightly better compared with WikiProfile-RW. This may be due to the fact that WikiProfile introduces more related concepts in user interest profiles.

[3]http://dumps.wikimedia.org/

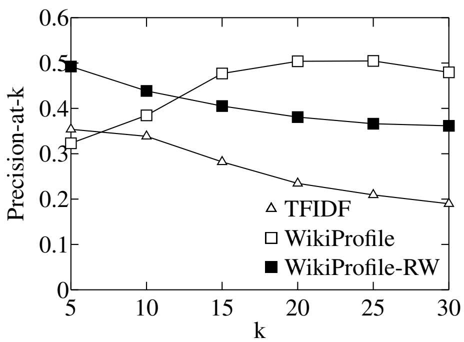

Figure 2: The recommendation performance measured by precision-at-k

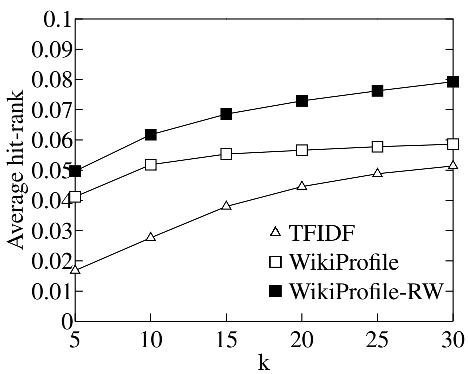

Figure 3: The recommendation performance measured by average hit-rank

For the average hit-rank plot, WikiProfile-RW performs the best, followed by WikiProfile. Both models outperform the TFIDF model, showing the effectiveness of our model.

Our model is also very efficient. After building Wikipedia inverted index and saving it to a database, the time to construct Wikipedia concept representation for each tweet is negligible. On a ordinary PC workstation, it can process hundreds of tweets per second.

One may notice that the overall performance is not very high. This is due to the constraints of the availability of the information in the data set. In fact a user typically does not label all the interested tweets as favorites or retweet them. However, in our experiment we can only use the favorite/reply/retweet information.

Conclusions and Future Work

In this paper, we have proposed a new recommendation model based on Wikipedia concepts and link structure. Using random walk on Wikipedia concept graph, we expand user's interest and improved the precision of recommendation. Our experiments with real-life data sets have demonstrated the effectiveness in tweets recommendation.

Some future research directions are as follows. For the Wikipedia part, we can further refine the set of concepts we used to do modeling. Since a large portion of tweets contain links to other web pages, we can also crawl the page information to help the extraction of related concepts in a tweet. To facilitate user reading, we can also give tags to each tweet and group tweets in different categories, making use of the Wikipedia-based representation. Meanwhile, the fast speed of information diffusion on Twitter can be a great tool for editor of Wikipedia to update/add pages. We believe that combining the rich knowledge of Wikipedia and quick information spread on Twitter will generate interesting and useful applications.

Acknowledgments

We would like to thank Prof. Cheng Hong for helpful advice and suggestions, and Dr. Bing Lidong for inspiring discussion.

References

Abel, F.; Gao, Q.; Houben, G.-J.; and Tao, K. 2011a. Analyzing user modeling on twitter for personalized news recommendations. In *Proceedings of the 19th international conference on User modeling, adaption, and personalization*, UMAP'11, 1–12. Berlin, Heidelberg: Springer-Verlag.

Abel, F.; Gao, Q.; Houben, G.-J.; and Tao, K. 2011b. Semantic enrichment of twitter posts for user profile construction on the social web. In Antoniou, G.; Grobelnik, M.; Simperl, E. P. B.; Parsia, B.; Plexousakis, D.; Leenheer, P. D.; and Pan, J. Z., eds., *ESWC (2)*, volume 6644 of *Lecture Notes in Computer Science*, 375–389. Springer.

Ahmad Wasfi, A. M. 1999. Collecting user access patterns for building user profiles and collaborative filtering. In *Proceedings of the 4th international conference on Intelligent user interfaces*, IUI '99, 57–64. New York, NY, USA: ACM.

Ahn, J.-w.; Brusilovsky, P.; Grady, J.; He, D.; and Syn, S. Y. 2007. Open user profiles for adaptive news systems: help or harm? In *Proceedings of the 16th international conference on World Wide Web*, WWW '07, 11–20. New York, NY, USA: ACM.

Billsus, D., and Pazzani, M. J. 1999. A personal news agent that talks, learns and explains. In *Proceedings of the third annual conference on Autonomous Agents*, AGENTS '99, 268–275. New York, NY, USA: ACM.

Chen, J.; Nairn, R.; Nelson, L.; Bernstein, M.; and Chi, E. 2010. Short and tweet: experiments on recommending content from information streams. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, 1185–1194. New York, NY, USA: ACM.

Gabrilovich, E., and Markovitch, S. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the 20th international joint conference on Artifical intelligence*, IJCAI'07, 1606–1611. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Gauch, S.; Speretta, M.; Chandramouli, A.; and Micarelli, A. 2007. The adaptive web. In Brusilovsky, P.; Kobsa, A.; and Nejdl, W., eds., *The adaptive web*. Berlin, Heidelberg: Springer-Verlag. chapter User profiles for personalized information access, 54–89.

Geyer, W.; Dugan, C.; Millen, D. R.; Muller, M.; and Freyne, J. 2008. Recommending topics for self-descriptions in online user profiles. In *Proceedings of the 2008 ACM conference on Recommender systems*, RecSys '08, 59–66. New York, NY, USA: ACM.

Grčar, M.; Mladenič, D.; and Grobelnik, M. 2005. User Profiling for Interest-focused Browsing History. In *Workshop on End User Aspects of the Semantic Web, ESWC-2005, Heraklion*.

Hannon, J.; Bennett, M.; and Smyth, B. 2010. Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on Recommender systems*, RecSys '10, 199–206. New York, NY, USA: ACM.

Hill, W., and Terveen, L. 1996. Using frequency-of-mention in public conversations for social filtering. In *Proceedings of the 1996 ACM conference on Computer supported cooperative work*, CSCW '96, 106–112. New York, NY, USA: ACM.

IJntema, W.; Goossen, F.; Frasincar, F.; and Hogenboom, F. 2010. Ontology-based news recommendation. In *Proceedings of the 2010 EDBT/ICDT Workshops*, EDBT '10, 16:1–16:6. New York, NY, USA: ACM.

Java, A.; Song, X.; Finin, T.; and 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*, WebKDD/SNA-KDD '07, 56–65. New York, NY, USA: ACM.

Kim, H.-j.; Lee, S.; Lee, B.; and Kang, S. 2010. Building concept network-based user profile for personalized web search. In *Proceedings of the 2010 IEEE/ACIS 9th International Conference on Computer and Information Science*, ICIS '10, 567–572. Washington, DC, USA: IEEE Computer Society.

Laniado, D., and Mika, P. 2010. Making sense of twitter. In Patel-Schneider, P.; Pan, Y.; Hitzler, P.; Mika, P.; Zhang, L.; Pan, J.; Horrocks, I.; and Glimm, B., eds., *The Semantic Web ISWC 2010*, volume 6496 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. 470–485.

Latif, A.; Afzal, M. T.; Helic, D.; Tochtermann, K.; and Maurer, H. A. 2010. Discovery and construction of authors' profile from linked data (a case study for open digital journal). In Bizer, C.; Heath, T.; Berners-Lee, T.; and Hausenblas, M., eds., *LDOW*, volume 628 of *CEUR Workshop Proceedings*. CEUR-WS.org.

Li, Z.; Zhou, D.; Juan, Y.-F.; and Han, J. 2010. Keyword extraction for social snippets. In *Proceedings of the 19th international conference on World wide web*, WWW '10, 1143–1144. New York, NY, USA: ACM.

Naaman, M.; Boase, J.; and Lai, C.-H. 2010. Is it really about me?: message content in social awareness streams. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, 189–192. New York, NY, USA: ACM.

Nanas, N.; Uren, V.; and De Roeck, A. 2003. Building and applying a concept hierarchy representation of a user profile. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, SIGIR '03, 198–204. New York, NY, USA: ACM.

Ramanathan, K., and Kapoor, K. 2009. Creating user profiles using wikipedia. In *Proceedings of the 28th International Conference on Conceptual Modeling*, ER '09, 415–427. Berlin, Heidelberg: Springer-Verlag.

Schouten, K.; Ruijgrok, P.; Borsje, J.; Frasincar, F.; Levering, L.; and Hogenboom, F. 2010. A semantic web-based approach for personalizing news. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, SAC '10, 854–861. New York, NY, USA: ACM.

Sugiyama, K.; Hatano, K.; and Yoshikawa, M. 2004. Adaptive web search based on user profile constructed without any effort from users. In *Proceedings of the 13th international conference on World Wide Web*, WWW '04, 675–684. New York, NY, USA: ACM.

Tang, J.; Yao, L.; Zhang, D.; and Zhang, J. 2010. A combination approach to web user profiling. *ACM Trans. Knowl. Discov. Data* 5:2:1–2:44.