

Utilizing Microblogs for Web Page Relevant Term Acquisition

Tomáš Uherčík, Marián Šimko, Mária Bieliková

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies
Ilkovičova 3, 842 16 Bratislava, Slovakia

uhercik07@student.fiit.stuba.sk, {simko,bielik}@fiit.stuba.sk

Abstract. To allow advanced processing of information available on the Web, the web content necessitates semantic descriptions (metadata) processable by machines. Manual creation of metadata even in a lightweight form such as (web page) relevant terms is for us humans demanding and almost an impossible task, especially when considering open information space such as the Web. New approaches are devised continuously to automate the process. In the age of the Social Web an important new source of data to mine emerges – social annotations of web content. In this paper we utilize microblogs in particular. We present a method for relevant domain terms extraction for web resources based on processing of the biggest microblogging service to date – Twitter. The method leverages social characteristics of the Twitter network to consider different relevancies of Twitter posts assigned to the web resources. We evaluated the method in a user experiment while observing its performance for different types of web content.

Keywords. Automatic Term Recognition, Keyword extraction, User-generated content, Social annotations, Microblog, Twitter

1 Introduction

World Wide Web has become one of the most important sources of information. The amount of information on the Web is huge, so a user is often overloaded and traditional information retrieval is no longer effective. Web content necessitates semantic descriptions (also referred to as metadata) that allow advanced processing such as intelligent search and personalization. This vision of machine-processable layer of metadata is for more than a decade addressed by the Semantic Web movement [3]. However, despite the increase, “semantization” of the Web is not as spread as we would like [15]. It is due to the fact, that manual creation of metadata for huge amounts of web content is almost impossible for a human being. Hence approaches for automated creation of metadata had emerged. They differ both in source of data they process and in an expressivity of semantics they produce.

The most fundamental form of metadata is term-based representation that we refer to as relevant terms (RTs). They constitute a basis for advanced semantic representations such as ontologies (either lightweight or heavyweight) [18]. With no proper terms that are relevant for a domain, construction of a more expressive semantic representation is hardly possible.

Recently there has been a huge spread of data that are an intended or not intended result of social activities and interaction on the Web. User-created annotations in various forms are produced by crowds for better creating, enriching, organizing and sharing additional information often related to existing web resources. Due to their amounts, social data represent a very promising source for mining and discovery of potentially useful semantic structures. We believe that despite the fact that they contain more noise, they can be used with an advantage to *supplement* traditional approaches to semantics discovery. In our work we particularly focus on microblogs as a phenomenon of today's web.

In this paper we propose a method for relevant domain term extraction for web resources that are referenced in the microblog Twitter by utilizing social aspect of microblogging. We selected the Twitter since it is the largest microblogging network to this date and it has less conversational focus than other social networks [9]. Other social networks like Facebook are used mostly for user interactions, which would result in metadata of lower quality.

The rest of the paper is structured as follows. In section 2 we discuss related work. In section 3 we present a method for relevant terms extraction from Twitter posts. Section 4 describes evaluation of the method. Conclusions are drawn in section 5.

2 Related Work

Approaches related to relevant terms extraction can be divided according to source of information utilized for processing. There are two basic groups of approaches in keyword extraction area – extraction directly from web page textual content (utilizing Automatic Term Recognition algorithms; ATR) and extraction from external sources.

There are two main groups of ATR algorithms: termhood and unithood algorithms. Termhood algorithms are based on term frequency, for example assuming that the term candidate will occur more often in domain specific documents than in the rest of them [1]. Unithood algorithms measure collocation strength in terms. It can be done for example by investigating mutual occurrence probability of words in term candidate [8]. Both approaches can be further extended, e.g., by processing document formatting or visual style [10]. Another method for keyword extraction from text is TextRank [13], which utilizes graph-based text representation as a source for keyword relevance calculation.

Works related to the area of rating (web) resources based on graph analysis are important to our work. Page and Brin introduced PageRank for web pages' ranking by processing topology of the Web [5]. This method was adapted to Twitter's characteristics to calculate user rankings by Tunkelang [16]. Its main idea is that user has big influence when is followed by many influential users. This recursive algorithm is

executed, until it converges. Another ranking method TwitterRank [17] is used to find topic-sensitive influential twitterers. It considers topic similarity between users to compute user’s influence to others. Unfortunately this algorithm is badly scalable for large amounts of tweets, which arise every day. TweetRank ranking was introduced to measure ranking of web resources referenced in Twitter posts [12].

Twitter as a social network and microblogging service is used for a lot of tasks nowadays, e.g., for topical news recommendation [14], for trends detection [6] or for extraction of personalized annotation tags for Twitter users [19]. To our best knowledge, there are no approaches which acquire keywords about Web resources using Twitter posts as a source of metadata extraction similarly to our concept.

3 Utilizing Twitter for Relevant Domain Term Extraction

Acquiring metadata from microblogs has a lot in common with ordinary methods, which extract keywords from documents, but it has its specifics.

3.1 Twitter – Source of Annotations for the Web Content

We can see the structure of the microblog in the context of our approach in the Fig. 1 . Users are interconnected by a followership relationship, i.e., they subscribe for content produced by others. Followers value content of followees, they can be influenced by them. We can compute authors’ rating or popularity using this kind of a relationship. There is also a relationship between a user and posts – tweets – he or she published. It is important, because author’s relevance can be a good indicator of tweet’s relevance. Another relationship is present between tweets and web resources they refer to. When the web resource is referred by the tweet, it can be described by that tweet. We believe that tweet could contain useful metadata about the resources.

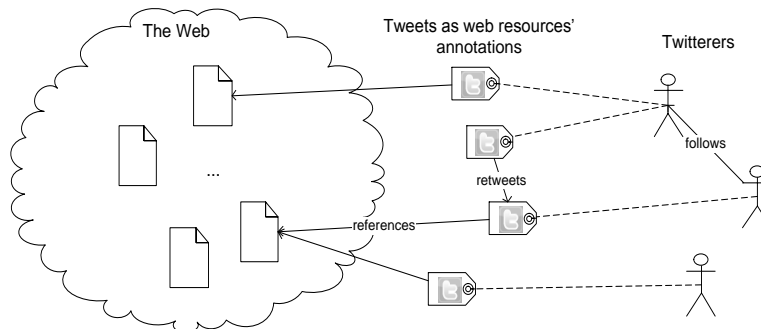


Fig. 1. Twitter posts as web resources’ annotations. We leverage Twitter graph consisting of users and tweets to extract relevant terms for a resources the tweets refer to.

The most important cons of Twitter are presence of slang, risk of user account abused for advertisement and spam spread. Specifics of microblogs that can be used

with an advantage are the hashtags (descriptive tags in tweets marked by symbol #), mentions (references to particular users marked by symbol @), retweets (re-posted tweets) and structure of implicit social network of microblog.

We see Twitter posts as annotations of web resource they refer to. Tweets' content, Twitter graph and aforementioned Twitter specifics we consider a potential source of data to mine in order to improve relevant term acquisition from web resources referenced on Twitter.

3.2 Method for Web Page Relevant Terms Extraction

We proposed a method for web page¹ relevant terms extraction utilizing Twitter posts. It consists of the following steps:

1. Tweet processing,
 1. Web resource lookup,
 2. Tweet ranking,
 3. Relevant terms extraction,
2. Web resource content processing,
3. Combining results.

In the first step we process tweets – we look up the web resource's URL in tweets and get tweets referring to the resource, we rank the tweets and extract terms with relevance weights. Then we extract relevant terms from web resource's content. Finally, we combine the results and select top- k relevant terms according to relevance weight.

Tweet Processing

Web Resource Lookup

In this step we find all occurrences of a given web resource's URL in tweets. We consider public tweets, which we access through Twitter search API. Twitter search API allows searching in tweets, which are not older than one week (approximately). This is a problem not only because we have a limited number of user annotations to exploit, but also because they are influenced by trends of that short time period. To overcome this drawback, we have to store tweets for relevant terms extraction purposes. However, the number of tweets produced daily is so big that it is very difficult to store all the Tweets. Since our method leverages social network of the Twitter, reasonable trade-off is to store only relevant tweets, e.g., tweets produced by relevant users (authorities), or tweets meeting stricter requirements related to content quality.

Tweet Relevance Ranking

In this step we rank tweets to distinguish more and less relevant ones. Our assumption here is the more relevant post, the more accurate terms it contains. The relevance of a tweet we compute by considering relevance of an author of the tweet. We utilize

¹ In this paper we use terms web page and web resources interchangeably.

TunkRank algorithm [16], which basically is PageRank [5] adapted to Twitter authors ranking. We selected this algorithm since it is scalable, simple and has reasonable results in author ranking.

Besides TunkRank, we also consider user ranking that reflects frequency of author's posts. We slightly changed the TunkRank computation to reflect frequency of posting tweets. Our change is based on the assumption that users who post tweets very often may write tweets with less valuable information, while authors, who tweets less often, write about the topics, they are highly interested in and they can contribute more valuable content [7]. We compute user ranking dependent on time gap between published tweets as follows:

$$TTunkRank(u) = \sum_{f \in \text{Foll}(u)} \frac{1 + \frac{p}{\log(TG)} \cdot TTunkRank(f)}{|\text{Foll}(f)|} \quad (1)$$

where $TTunkRank(u)$ is the time-aware TunkRank for user (author) u , $\text{Foll}(u)$ is the set of users following u , p is convergence constant and TG is time gap median, which is defined as:

$$TG = \text{Med}(TG_0, TG_1, \dots, TG_n) \quad (2)$$

where TG_i is the time gap between published tweets of the author. TG_0 is time gap between publishing first and second tweet, TG_1 is time gap between second a third tweet etc. $TTunkRank$ can be of value from an interval $(0;1)$.

For each tweet referencing a web resource we compute tweet ranking as a maximal ranking of a user who either created or retweeted the tweet.

$$TweetRelRank(t) = \max_{u \in U_t} (URank(u)) \quad (3)$$

where $TweetRelRank(t)$ is relevance ranking of tweet t , $URank(u)$ is user rating of user u . (note that we can use any known user ranking algorithm, e.g., $URank(u) = TTunkRank(u)$) and U_t is a set of all users who created or retweeted the tweet t .

In addition to user ranking tweet ranking can be determined by considering other structural properties, e.g., number of retweets, or properties of content, e.g., formal quality of tweet (less slang) or emotional characteristics derived by natural language processing. These features have not yet been incorporated to our method since we currently focus on exploring an impact of user rating.

Relevant Term Extraction from Tweets

In this step we extract the relevant terms from the gathered tweet posts. In contrast with traditional text processing methods we consider information about tweet relevance. For extraction of relevant terms we can use any known automatic text recognition (ATR) algorithm or their combination.

In our implementation we use TextRank algorithm [13], but we could use another algorithm for textual ranking of words from documents (in our case tweets), or a text processing service such as AlchemyAPI or OpenCalais².

² <http://www.alchemyapi.com/>, <http://www.opencalais.com/>

We compute final microblog-based relevance $MBRel$ of a relevant term rt for web resource d as follows:

$$MBRel(rt, d) = \max_{t \in T_d} (TweetRelRank(t)) * TRank(rt, T_d) \quad (4)$$

where $TweetRelRank(t)$ is relevance ranking of tweet t and $TRank$ is textual ranking of a relevant term rt in a set of tweets T_d referencing a web resource d . Note that for computation of $TRank$ the whole set of tweets is considered. We follow an approach of Wu et al., who showed that multiple posts processing results in better performance of term extraction [19].

Relevant Term Extraction from Content of Web Resources

In addition to microblog-based term extraction we extract terms from a web resource's content. This step is a part of our method since the content of the resource remains a very relevant source for relevant term extraction and our aim is to leverage microblog to enrich relevant domain terms extracted from content and provide more accurate results.

Similarly to tweets' content processing, we can employ any existing method to content processing in order to extract relevant terms from the content. We denote the relevance of relevant domain terms extracted from content as $CRel$. The detailed description of relevant domain terms extraction from a web resource content is beyond the scope of this paper.

Combining Results

An input of this step is set of top- k relevant terms acquired from microblogs and the set of top- l relevant terms obtained by the resource content processing. We combine relevant terms' relevancies from the both sources by computing combined ranking.

For every relevant term extracted from microblog we have relevance $MBRel$ and for every keyword gained from textual content analysis, we have relevance $CRel$. Final relevance Rel of relevant term rt in web resource d we compute as follows:

$$Rel(rt, d) = \alpha.MBRel(rt) + (1 - \alpha).CRel(rt, d) \quad (5)$$

where $MBRel(rt, d)$ and $CRel(rt, d)$ are a microblog-based relevance and content-based relevance of relevant term rt in a web resource d , respectively. α is weighting coefficient determining the importance of both rankings.

Finally, we select top- m ranking relevant terms to be a final set of relevant terms for a given resource.

4 Evaluation

We evaluated our method by conducting a user experiment, where accuracy of the results provided by the method was assessed by a selected group of users – experts in particular domains selected.

4.1 Experiment setup

We obtained more than 90 GB dataset from Twitter using its streaming API during 10 days. We selected the set of 60 web resources (URLs) for evaluation. The main criteria of choosing the web resources for evaluation were (i) our ability to judge if the keywords are appropriate, and (ii) the representativeness of the web resource set (i.e., we chose different types of resources – videos, pictures, news articles, etc.). We preferred to choose web resources with rather higher PageRank, because we were not intended to gain metadata about spam pages. We utilized TextRank algorithm for tweets’ content processing and the Web service AlchemyAPI for relevant term extraction from web resources’ content. We set the weighting coefficient α determining the importance of the both relevancies to balanced value 0.5.

Since our aim was to evaluate the method with respect to different types of web resources’ content, we classified web resources into the following groups:

- *Comprehensive text* – news articles or other types of dense text. This type represents approximately 20 % of evaluated set.
- *Brief text* – pages, which consist of different types of sparse text, page navigation etc. This type represents approximately 20 % of evaluated set.
- *Product pages* – pages, which describe products. One of the typical characteristics of this product is highlighting of products’ important features in text. This type represents approximately 20 % of evaluated set.
- *Minimal text content pages* – this type involves videos, pictures, music, radios etc. Due to big variety of particular subtypes of this category, it represents approximately 40 % of evaluated set.

We evaluated our method by an experiment in which independent experts fulfilled surveys, where they rated extracted terms as relevant or irrelevant to shown web pages. Our survey was completed by 46 active users, who produced 4400 assessments of offered terms. The assessment in this context means marking a term as relevant or irrelevant. For evaluation we considered only terms, which had three or more assessments.

We employed three measures that evaluate precision of the whole set of extracted terms, enrichment rate based on tweets and the extension proportion of relevant terms from tweets. Precision of the method in web resource d we define as follows:

$$prec(d) = \frac{|RT_{rel}(d)|}{|RT(d)|} \quad (5)$$

where $RT_{rel}(d)$ is set of all relevant terms from web resource d judged really as relevant and $RT(d)$ is set of all relevant terms extracted by our method. The second measure is the enrichment rate enr , which shows the rate of enrichment, how terms from Twitter enrich the whole set of web resource relevant terms obtained by our method.

$$enr(d) = \frac{|RT_T(d) \setminus RT_c(d)|}{|RT_T(d) \cup RT_c(d)|} \quad (6)$$

where $RT_T(d)$ is the set of relevant terms for web resource d obtained from microblog Twitter and a $RT_c(d)$ is the set of terms obtained from the content of the web resource. We consider enrichment rate very important because it shows, how much the terms from Twitter participate in the whole set of relevant terms – it shows the potential of user-generated data for relevant terms acquisition.

The third measure we observed shows the proportion of the relevant terms from microblog Twitter, which are not present in the web resource’s content.

$$ext(d) = 1 - \frac{|RT_T(d) \cap AT(d)|}{|RT_T(d)|} \quad (7)$$

where $RT_T(d)$ is the set of relevant terms obtained from microblog Twitter and $AT(d)$ is the set of all terms which are present in the web resource’s content. This measure shows the extension proportion of metadata set by terms acquired externally, i.e., those not present in the web resource’s content and relevant for the resource.

4.2 Results and Discussion

For all aforementioned measures we computed averaged values per web resource. Results obtained for the whole set of web resources, which were involved in the experiment, are shown in Table 1. We are particularly interested in precision of terms acquired from both content and microblog ($prec^{merged}$), precision of terms acquired from microblog only ($prec^{twitter}$), enrichment rate (enr) and extension proportion (ext).

Table 1. Results of experiment for the whole set of evaluated web resources.

	$prec^{merged}$	$prec^{twitter}$	enr	ext
Comprehensive text	0.700	0.760	0.387	0.770
Brief text	0.590	0.594	0.429	0.207
Product pages	0.595	0.550	0.211	0.810
Minimal text content pages	0.645	0.685	0.448	0.224
All	0.639	0.663	0.380	0.155

Average precision obtained seems to be high, despite the fact that we offered users all terms excluding only those with zero relevance. Interesting and quite surprising finding was that precision of terms extracted from microblog was higher than precision of the whole set of terms merged with terms extracted from web resource’s content. Another positive observation is that Twitter terms enrich the whole content with 38 % rate, which means that 38 % of relevant terms were present only in the set originating from Twitter. Terms extracted in duplicate by both methods were considered as obtained from content to show the improvement made by Twitter terms. To deal with the fact that terms extracted from content and terms extracted from Twitter can be similar but not equal (while representing the same thing), we measured the equality of terms with a small toleration based on *Levenshtein* distance.

Precision. We obtained the best precision for comprehensive text. It is not surprising because web resources of this type are mostly described in microblog posts using words that describe the content. This type of resource is also frequently shared by users of microblog, who tend to express to/comment the content of page *objectively*. Sometimes they are authors themselves and use microblog posts to promote the web resource they created while providing additional useful descriptions. The fact that content based method we used for term extraction is successful for this type is not surprising at all, because these methods have in general good results for comprehensive content.

The worst precision we obtained for product pages, which are in microblogs mentioned probably by persons, who are interested in sale of products and profiting from the sale. They can provide descriptions, which could be confusing. However, the merged precision is higher because the basic characteristics of product on the page are often appropriate. The precision comparison we can see in the Fig. 2.

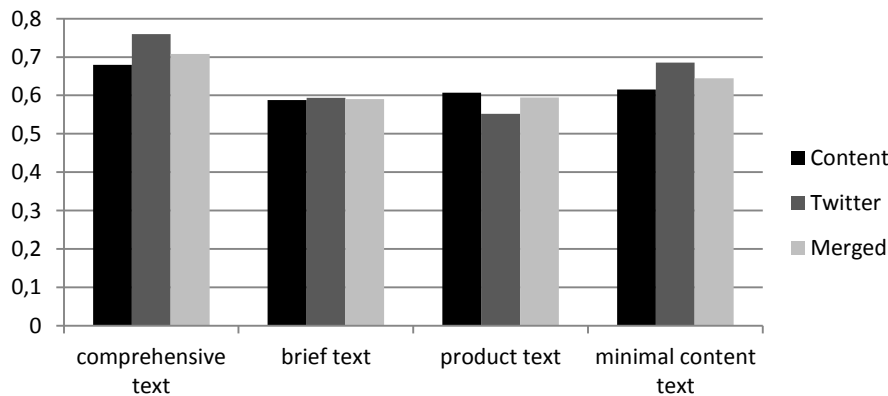


Fig. 2. Precision comparison for particular types of pages.

Enrichment rate. The highest enrichment rate *enr* was obtained for web resources with minimal text content, i.e., pictures, videos or similar content. The method was able to reasonably improve results for resources where traditional content-based methods are not very successful. Surprising is the fact, that this rate is not significantly higher than for other types of web resources. This can be explained by our focus on web resources, which contain at least few words (e.g., a picture with at least some description). We were particularly interested in improvement comparison and by considering non textual web resources only, the enrichment rate would be obviously 100 %.

The lowest enrichment rate we obtained for product pages. This is caused by the fact, that these pages are shared together with advertisement phrases and they are published in the same form by many salesmen. We obtained only narrow set of terms. Moreover, they could be even confusing. Comparison of improvement rate for individual types of pages we can see in Fig. 3.

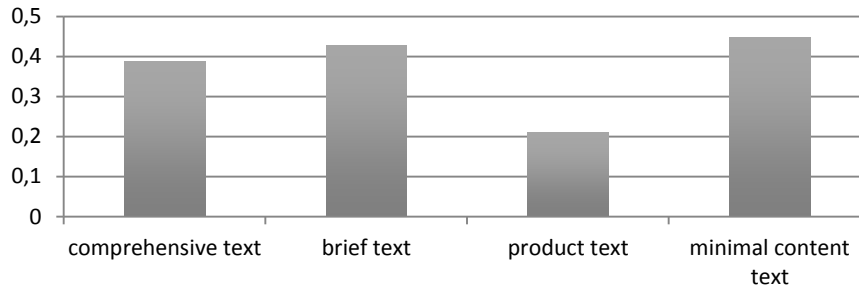


Fig. 3. Comparison of enrichment rate for individual types of web resources.

Extension rate. The highest extension rate was obtained for pages with minimal text content and brief text as we supposed, because there are only few words contained within, so Twitter terms extend the merged set a lot. On the other hand, the lowest extension rate was achieved for comprehensive text containing rich text, and product pages, which often include different phrases, trying to confuse search engines. We can see the comparison of the extension rate in Fig. 4.

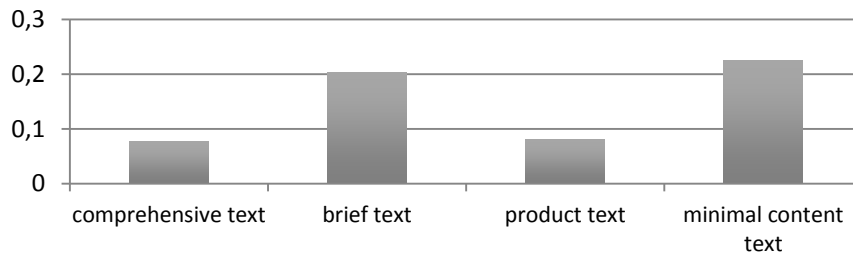


Fig. 4. Comparison of the extension rate for individual types of resources.

Finally, Fig. 5 **Error! Reference source not found.** shows the correlation of extracted terms' relevance calculated by our method and relevance assessed by human judges (we computed average judge rating for each resource). As we can see, the measures are correlated, i.e., our method succeeds in extracting relevant terms that are considered relevant also by the human judges.

5 Discussion and Conclusions

Relevant domain terms are the cornerstone of domain and user representation for advanced information processing web-based applications [4, 2]. In this paper we presented the method for web resources' relevant domain terms acquisition by utilizing microblogs created by huge masses of various users. We see microblog posts as web resources' annotation with a tremendous potential for relevant information to mine.

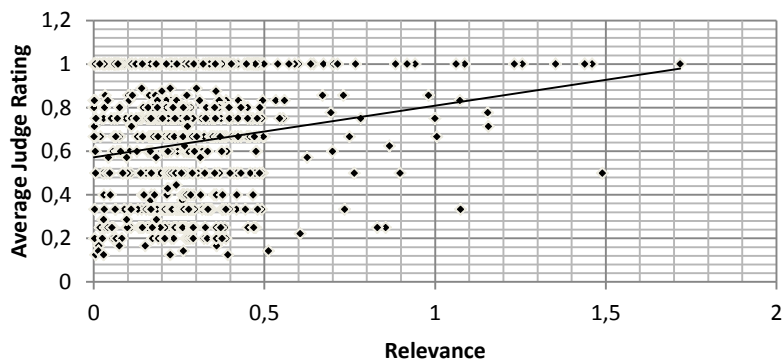


Fig. 5. Correlation of relevance computed by our method and ratings of judges.

Increasing amount of such data together with ability to select only relevant contributions makes microblog a potential source for improvement of traditional content-based approaches to relevant terms extraction.

Our method is able to enrich the set of relevant terms extracted from web resource's content, especially for resources, which do not contain a lot of self-describing text. We rely on the Twitter graph, which enables us to filter extracted terms by considering tweets relevance derived from authors' ranking. A further enhancements of the method can be done by more complex processing of tweets, e.g. by considering their language features (hence recognizing informal and emotional posts from objective ones) or employing additional data mining while considering Twitter specifics (e.g., by clustering related tweets by hashtags or using other similarity measure).

We showed that our approach significantly enriches the relevant terms extracted from web resource's content. Depending on the type of a resource to extract relevant terms for, the results differ. The potential of the method is to treat different content type differently. Our future work covers further analysis of different resource types and incorporation of resource type identification into the method. Currently we investigate setting of parameter α and its automatic derivation with respect to a web resource type.

Such massive data sources as microblogs deserve a special attention when considering scalability and efficiency of processing. User rating computation represents a particular challenge, when more advanced criteria or hybrid approaches would be considered to select the most relevant users. This also calls for further research in adopting such approaches to take full advantage of parallel computing and cloud-based services.

Acknowledgement. This work was partially supported by the grants VG1/0675/1/2011-2014, APVV-0208-10 and it is the partial result of the Research & Development Operational Programme for the project Research of methods for acquisition, analysis and personalized conveying of information and knowledge, ITMS 26240220039, co-funded by the ERDF.

References

1. Ahmad, K., Gillam, L., Tostevin, L.: University of Surrey participation in TREC 8: Weirdness indexing for logical document extrapolation and retrieval (WILDER). In: Proc. of The Eighth Text REtrieval Conference, TREC 8 (1999)
2. Barla, M.: Towards Social-based User Modeling and Personalization. In Information Sciences and Technologies Bulletin of the ACM Slovakia, 3(1), pp. 52-60 (2011)
3. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. Scientific American Magazine, May (2001)
4. Bieliková, M., Barla, M., Šimko, M.: Lightweight Semantics for the “Wild Web”. In B. White, P. Isaias and F. M. Santoro (eds.): Proc. of the IADIS Int. Conf. on WWW/Internet, ICWI 2011, IADIS Press, pp. xxv-xxxii (2011)
5. Brin, S., Page, L.: The anatomy of a large-scale hypertextual web search engine. In Proc. of the 7th Int. Conf. on World Wide Web, pp. 107–117 (1998)
6. Dong, A.: Time is of the essence: improving recency ranking using Twitter data. In: Proc. of the 19th Int. Conf. on World Wide Web, 2010, ACM, pp. 331–340.
7. Chen, J., Nairn, R., Nelson, L., Bernstein, M., Chi, E.: Short and tweet: experiments on recommending content from information streams. In Proc. of the 28th Int. Conf. on Human factors in Computing Systems, 2010, ACM, pp. 1185–1194.
8. Church, K.W., Hanks, P.: Word association norms, mutual information, and lexicography. In: Computational Linguistics, MIT Press, pp. 22–29 (1991)
9. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In: Proc. of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis, pp. 56–65 (2007)
10. Lučanský, M., Šimko, M., Bieliková, M.: Enhancing Automatic Term Recognition Algorithms with HTML Tags Processing. In Proc. of Int. Conf. on Computer Systems and Technologies, CompSysTech 2011. ACM, pp. 173–178 (2011)
11. Lučanský, M., Šimko, M.: Impact of Web Site Visual Style on Keyword Extraction. In SOFSEM 2013: Theory and Practice of Computer Science. Springer, to appear (2013)
12. Majer, T., Šimko, M.: Leveraging Microblogs for Resource Ranking. In SOFSEM 2012: Theory and Practice of Computer Science, LNCS 7147. Springer, pp. 518–529 (2012)
13. Mihalcea, R., Tarau, P.: TextRank: Bringing Order into Texts. In: Proc. Of Conf. on Empirical Methods in Natural Language Processing, ACL, pp. 404–411 (2004)
14. Phelan O., McCarthy K., Smyth, B.: Using twitter to recommend real-time topical news. In: Proc. of the 3rd ACM conf. on Recommender systems, ACM, pp. 385–388 (2009)
15. Sabou, M., Gracia, J., Angeletou, S., D'Aquin, M., Motta, E.: Evaluating the semantic web: a task-based approach. In Proc. of 6th Int. Semantic Web Conf., ISWC 2007, LNCS 4825, Springer, pp. 423–437 (2007)
16. Tunkelang D.: A Twitter Analog to PageRank, 2009. Available at: <http://thenoisychannel.com/2009/01/13/a-twitter-analog-to-pagerank/>
17. Weng, J., Lim, E., Jiang, J., He, Q.: TwitterRank: Finding Topic-sensitive Influential Twitterers, In Proc. of the 3rd Int. Conf. on Web Search and Data Mining, pp. 261–270 (2010)
18. Wong, W., Liu, W., Bennamoun, M.: Ontology learning from text: A look back and in the future. ACM Computing Surveys (CSUR), 44(4), Article No. 20 (2012)
19. Wu, W., Zhang, B., Ostendorf, M.: Automatic generation of personalized annotation tags for twitter users. In: The 2010 Annual Conf. of the North American Chapter of the Association for Computational Linguistics, 2010, ACL, pp. 689–692.