

Unified Semantic Blog Mining Framework and Summarized Blog Retrieval

Godfrey Winster Sathianesan and Swamynathan Sankaranarayanan

Department of Information Science and Technology,
College of Engineering Guindy Campus, Anna University, Chennai-600025, India

Received 2012-06-15, Revised 2012-06-19; Accepted 2013-04-02

ABSTRACT

In today's scenario, publishing the personal content and technical content has become very easy for the publishers, because of the readily available prominent social media called blog. Content developers, teachers, researchers post various articles relevant to their research or newly emerging topics. Blog content in one blog source may resemble the semantics of blog from other source. Readers, who are fascinated in reading the blog content, anticipate retrieving the blogs from different sources for their query. A large number of posts are available in the web. Hence the blog reader's task becomes very complex to search the relevant content for their query. This study introduces a novel idea to collect the blog using unified Semantic Blog Mining Framework (SBMF). SBFM collects blogs from different blog sources using the ontology constructed for education domain. Blogs collected from different sources are collection which contains relevant or irrelevant blogs. The new blog summarizer summarizes the blog content and ranks the blogs according to the similarity of the blog with query given by the user. The proposed blog summarizer check the similarity of each sentence in a blog, sort the order of sentence based on the similarity of the text with query word and reduces the number of sentences. The experimental results shows that the proposed unified SBFM and blog summarizer produces better relevant and summarized number of blogs compared to various search engines. SBFM confirms the hundred percentage relevancy compared to other blog search engines. Blog summarizer yields accurate summarization for the collected blogs.

Keywords: Summarization, Blogging, Ontology, Ranking, Blog Mining, Social Learning

1. INTRODUCTION

Blogs are online journals or sets of chronological news entries that are maintained by individuals, communities' news entries that are maintained by individuals, communities or commercial entities and can be used to publish personal opinions, diary-like articles or news stories relating to a particular interest or product (Breslin *et al.*, 2009). It can be used to create online encyclopedia, photo gallery and literature collections. Blog facilitates collaboration and sharing between users with low technical barriers. One outcome of such blogs

is that they can produce more valuable knowledge collectively rather than created by separated individuals.

A vocabulary is a collection of terms being used in a particular domain that can be structured hierarchically as taxonomy and combined with some relationship, constraints and rules to form ontology (Breslin *et al.*, 2009). A combination of ontology together with a set of instances of classes constitutes a knowledge base. As more and more social websites, communities and services come online, users of the web face lot of problems. Blog retrieval has various challenges, such as identifying the reliable and authoritative blogs, ambiguity in classifying

Corresponding Author: Godfrey Winster Sathianesan, Department of Information Science and Technology,
College of Engineering Guindy Campus, Anna University, Chennai-600025, India

blogs and unstructured nature. The issues are addressed by collecting the blogs from reliable sources based on the ontology, preprocessed to eliminate the irrelevant blogs and stored in the blog repository with the semantic relationship. First, the social websites are isolated from one another, so there is lack of interoperability between them. Blog are published in various blog sites like Technorati, blog.gs, wordpress. Second, the bloggers use informal text like bcos(because), b4(before), U(you). which makes the retrieval and search process complex and less efficient. Third, web contains huge amount of non-English blogs, which makes the blogger and user very difficult to read. **Figure 1** shows that the blog search engine retrieves 2638 blogs for the keyword “parallel computing” which contains non English blog as the first relevant blog. To address the issues identified in the web, the blogs are first collected from various blog sites like, Technorati, blog.gs, wordpress. Since the blog from various sites have different blog format, a common markup language called Blog Markup Language (BML) is used to generalize the blog structure. BML uses the set of tags in XML to generalize the blog information. This improves the interoperability among the blogs and makes the search process easier. Second, commonly used informal text is identified and its corresponding expansion is available as bag of words. Third, non-English blogs are identified and removed to improve the efficiency of search.

The contribution of this study can be summarized as follows:

- A three layered framework is designed and implemented for blog mining
- The domain specific ontology is created for an education domain
- Blog summarization method is proposed to summarize the blogs retrieved for the user query
- Results of proposed SBMF and blog summarizer are evaluated and compared with blog search engines

1.1. Related Work

Blog is increasingly becoming an important source of information. Blog community is a group of bloggers who have the same interest and they search for similar kind of topics on the Internet. The blog community discovery algorithm can cluster bloggers effectively (Lu and Lee, 2008). A framework for learning XML document

classification achieves good performance in both supervised and semi-supervised with the bonus of producing comprehensible learning theories (Wu, 2012). Description Logics (DLs) are a family of state_of_art knowledge representation languages and their experience power has been carefully crafted to provide useful knowledge modeling primitives while allowing for practically effective decision procedures for the basic reasoning problems. The ontology editor protégé 4 supports description graphs (Motik *et al.*, 2009).

Gold standard evaluation of ontology learning method measures take into account the lexical and the relational dimensions of the learned ontology and penalize it in properties to its differences from the gold standards (Zavitsanos *et al.*, 2011). Effective XML keyword search which includes the identification of user search intention and result ranking in the presence of keyword ambiguities. It introduces XML TF*IDF similarity ranking scheme to capture the hierarchical structure of XML data (Bao *et al.*, 2010). Blog recommendation mechanism is used to personally recommend suitable blog articles or bloggers to the users in the blogosphere of practice (Li and Chen, 2009).

Portfolio framework called blogfolio framework supports the user or teacher to create a blog for online learning. The system fails to collect blog similar to the teacher’s blogs from the social web (Lin *et al.*, 2007). Domain ontology of competitive intelligence is built, in order to support the process of semantic content analysis and the intelligence retrieving on the web (Liu *et al.*, 2011). Learning Resources Recommendation System Based on Education Blog introduces the design and implementation of interest mining and learning resources. The interest mining component and the blog recommendation module adopts document classification technology to mine learners’ interests and gets a satisfactory result (Liu, 2008).

Ranking of bloggers based on link analysis can exemplify the characteristics of blogs and reduce the influence of link spamming (Yuhanget *et al.*, 2008). Due to interactive and dynamic atmosphere of blogspace, we believe that behavioral features of users, such as commenting, blog updating rate, different types of citation and time of citation, can help in ranking the blogs according to the importance and popularity more accurate than other features like blogs similarity (Tayebi *et al.*, 2007). The user when searches for the relevant useful content may fail in their need.



Fig. 1. Blog Search Engine ICEROCKET retrieves non English blog for keyword “parallel computing

A high score is assigned to the blog entries submitted by a good blogger and not yet linked. Blogs based on acceptance of the bloggers prior work also get high score. A blog site is constructed from a set of blog entries written by a single blogger and the quality of blog entries and topics are dominated by the ability or interests of the blogger (Fujimura *et al.*, 2005).

The user may get the irrelevant and not trustable blog using EigenRumor algorithm. Many links in blogs always link to the bloggers friends. This situation represents human relationship network and social network more often than blog entries. The top results returned by search engine cannot retrieve the information that users needed exactly (Shen *et al.*, 2008). Blogrank uses both link and similarity information in order to estimate the probability of a blog surfer to follow a link to another blog (Kritikopoulos *et al.*, 2006). Blogrank prioritizes links to blogs with similar topics or contributors and assigns probability to weblogs which are strongly interconnected. Personalized ontology construction is used to collect web information. Ontologies are widely used to represent user profiles in

personalized web information gathering. Ontology mining discovers interesting and on-topic knowledge from the concepts (Tao *et al.*, 2011). Ontologies provided a shared and common understanding of a domain that can be communicated between people and across application systems (Lilac and Nadeen, 2010). An ontology based approach to implement recommendation system that involves applying innovative web usage mining on log system to discover all possible imminent navigation patterns of current user and resolve any uncertainties in discovering the navigation pattern by applying ontological concept based similarity comparison and scoring algorithm (Mohanraj and Chandrasekaran, 2011).

Document summarization techniques are one way of helping people to find information effectively and efficiently (Ouyang *et al.*, 2011). Methods like Genetic Algorithm (GA), Mathematical Regression (MR), Feed Forward Neural Network (FFNN), Probabilistic Neural Network (PNN) and Gaussian Mixture Model (GMM) are used for text summarization task. The method using feature extraction criteria and training the model lags to address

multi-document summarization (Fattah and Ren, 2009). Regression models are used to the sentence ranking problem in query-focused multi-document summarization. The regression models are implemented using Support Vector Regression (SVR). SVR is a regression type of Support Vector Machine (SVM) (Ouyang *et al.*, 2011). Usually errors occur in query-based opinionated summary for blog entries. Such errors are identified and blog summarization was compared with news texts summarization (Tao *et al.*, 2011). Fuzzy logic with swarm intelligence could play an important role in the selection process of the most important sentences to be included in the final summary.

Existing text summarization methods (Mithun and Kosseim, 2009) lags in addressing the multi-document summarization, regression models (Ouyang *et al.*, 2011) are query-focused. These problems lead to inaccurate summarization. Hence the method to summarize the blog and a unified framework called Semantic Blog Mining Framework (SBMF) is proposed in this study.

2. MATERIALS AND METHODS

Figure 3 shows the framework for collecting the blogs from the web, preprocessing and storing it for the user retrieval. The system contains blog collector, blog preprocessor, blog storage and user retriever. In Blog collection a domain specific ontology is created and the subjects from the ontology are given as the seed to the crawler to collect the blogs from the web. Crawler crawls various website like Technorati, blog.gs, wordpress to collect the blogs relevant to the subject. The blogs collected from the web are converted to a common format using BML. Then the blogs are available in a structured format for further processing. The blogs collected may contain irrelevant, non-English and blogs with inadequate content. **Figure 2** shows the irrelevant blog retrieved for the keyword “apple” from education domain ontology. The keyword actually means the brand apple for the electronic systems like laptop, iPhone and ipad, but the content retrieved is the blog about a watch with name “apple”.

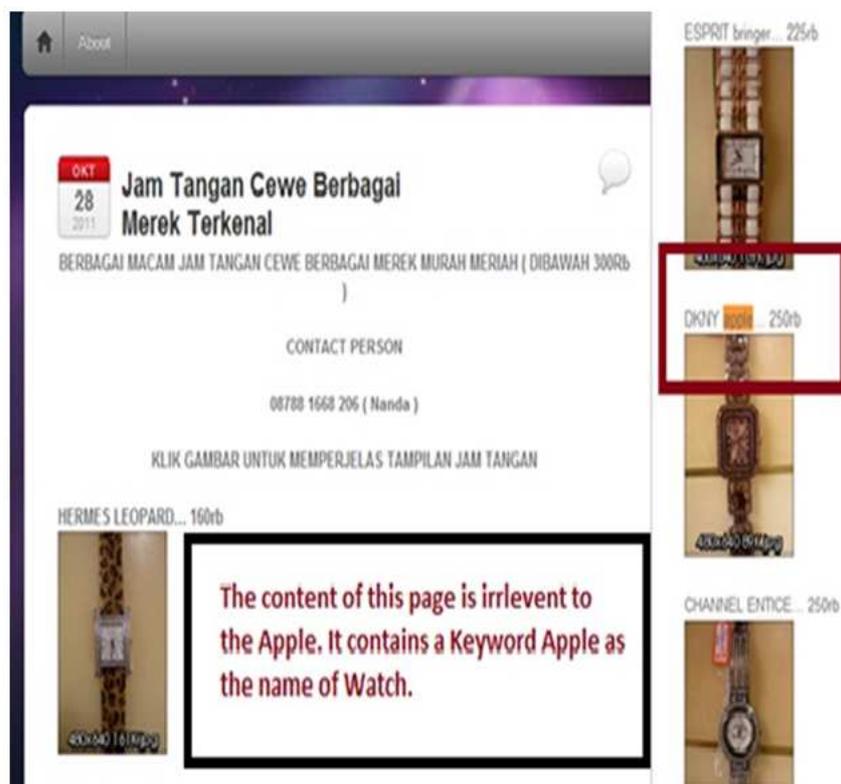


Fig. 2. Irrelevant blog for keyword “apple” in the Education domain ontology

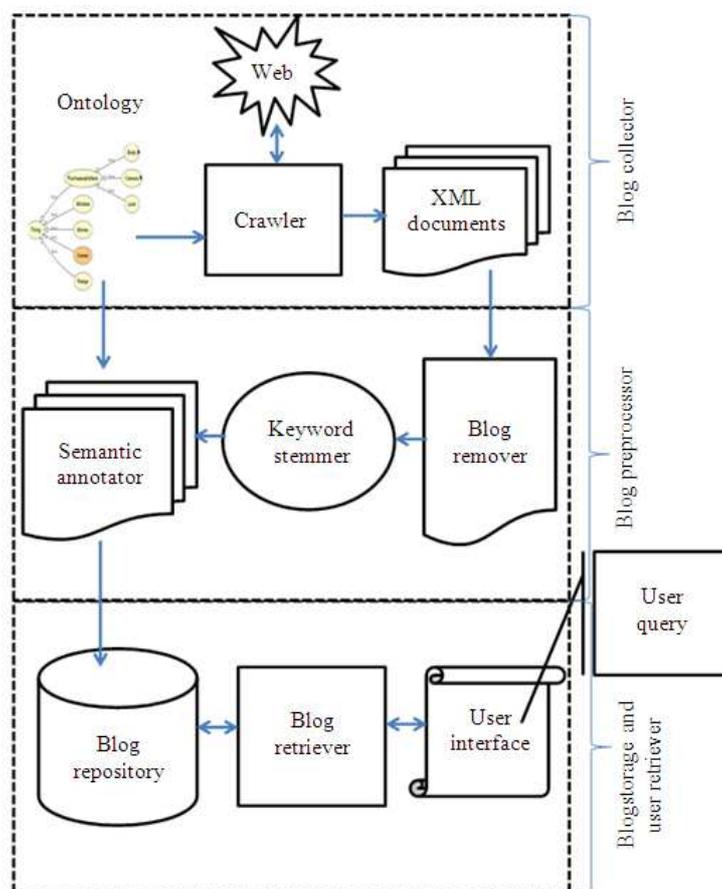


Fig. 3. Architecture of Semantic Blog Mining Framework (SBMF)

Blogs are preprocessed to improve the efficiency of search. Blog preprocessing is done to remove the non-English blogs. The blogs are checked for the percentage of relevancy to the subject from the ontology. Blog preprocessor removes the irrelevant blogs. The blogs with less content are removed to make the available blogs meaningful. The preprocessed blogs are then annotated with one another to make the semantic relationship of blogs using domain specific ontology. The annotated blogs are stored in blog repository.

Figure 4 Blog summarizing system in blog storage and user retrieval layer the blogs are stored in blog repository. The user who needs a blog relevant to his keyword, search the blog through a user interface. The search keyword is analyzed for its similarity based on the relevance and the relevant blogs are retrieved for the user. The domain specific ontology was created for

education domain. However domain specific ontologies can be created for various applications like Business, Entertainment, politics, social work and natural disaster. Blog mining is mostly used in all the domains **Table 1** shows the various applications of blog mining domains with example.

Blog summarizing system contains the blog search interface, blog ranker, blog summarizer and the SBMF as shown in **Figure 4** Blog search interface allow the user to query the system. The user can search for his required blogs from the repository through the query interface. The user registration is done to maintain the user profile to refine his search. Blog ranking ranks the blogs according to the order of relevancy using Content Based Importance (CBI). SBMF is the common framework to collect, preprocess and store the semantically annotated blogs.

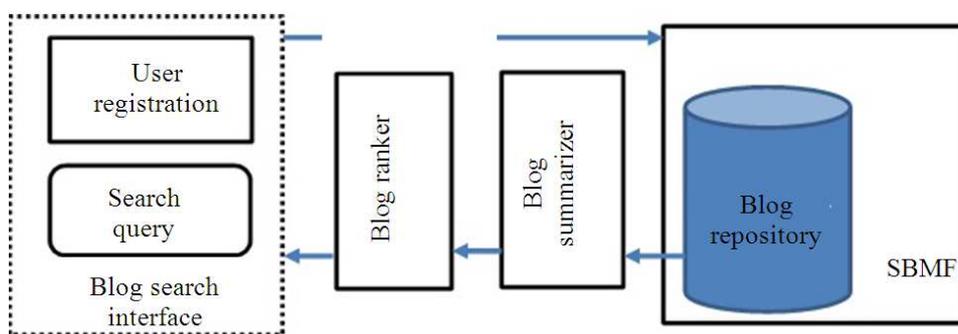


Fig. 4. Blog summarizing system

Table 1. Application of blog mining

| Domain | Description | Examples |
|------------------|--|---|
| Education | Posting articles on emerging and new topics | Oracle, IBM and Microsoft |
| Business | Product reviews, opinion about products, Advertisement | Customer review and comments, Product Advertisement |
| Entertainment | Audio, video (podcast) and sports | Audio and video collection, |
| Politics | Political Blogs show the public interest on the party or candidate | Political Party |
| Social work | Bloggers post articles related to the poverty, | Child poverty |
| Natural Disaster | Hurricane, Earthquake, Volcano | Hurricane |

Two types of summarization are introduced in this study. First, the blogs are summarized content wise. Second, blog repository is summarized to give the overview of subject content in the blog repository. To summarize the blog content this study introduces new method to summarize the minimal number of sentences with meaningful content. This study also discusses the importance of using the well-known measures like Term Frequency and Inverse Document Frequency (TF-IDF), similarity measure, Jaccard coefficient, word matching and stopwords. The summarized blogs are ranked using Content Based Importance (CBI) (Dominich, 2008). Query word is given by the user to the system through the user interface created to evaluate the proposed work. Blog summarizer yields the summarized version of multiple blogs from the blog repository. User can also get the summarized blog in the ranked order with more relevant blogs first. This will minimize the maximum time of user in searching the relevant and useful blog content.

The semantic blog mining framework uses the ontology to collect the blogs from various blog sources. The example of blog sources are technorati, blog.gs and wordpress. The collected blogs are converted to BML format and then preprocessed. The blogs are annotated to create semantic relationship

among blogs. The semantically related blogs are stored in the blog repository based on similarity between keyword and blog. The similarity between keyword and blog are analyzed for relevancy. The framework is implemented and the results obtained are efficient for personalized semantic based blog retrieval (Sathianesan and Sankaranarayanan, 2012). The framework is compared with blog search engines which searches the web for relevant blogs. The blog search engines collect a smaller amount of relevant blogs. The SBMF collect the blogs from different sources. The relationship of blogs with one another is created based on the ontology relationship. The blogs collected by this framework is preprocessed and only relevant blogs are stored in the repository.

The preprocessed blog are stored in oracle database with relationship between blog content and tags which represent the actual value of the blog. The relevance score of each tag was calculated to reflect the content relevance to the blog. A user interface was developed to interact with the user to get a query word. Blog repository was searched for the content based relevance of the query word along with the meaning of the word from wordnet. The blog content may contain the query word or the meaning of the query word.

The collected blog content was split into various sentences ($S_1, S_2, S_3 \dots S_n$). Term frequencies of all keyword are computed for each sentence. It is also necessary to compute the similarity of the sentence S_i to the query word 'q'. In some cases the sentence may not have the same word as query word but the semantics may appear in the sentence. The sentence is checked for word matching in terms of semantics, equivalence and relevance. Sentence with more than 75 percentages of stopwords doesn't yield any useful information so it will be removed from the sentence collection. Calculate the mean value of TF-IDF, cosine similarity and word matching to consider highest mean valued sentence as first sentence. To make the summarized content simple top five meaningful sentences are selected.

3. RESULTS

Figure 5 and 6 shows the summarized blogs ranked in the order of relevance to the query word. Blog summarizer also summarizes the overall blog content in

the blog repository. Figure 7 shows the summary of blogs in the repository of SBMF. The subject in the ontology and the frequency of corresponding subject blogs are summarized. The experimental results shows that the proposed work performs well compared with existing blog search engines like Technorati, blogpulse, blogscope, icrocket and regator. Blog summarizer retrieves only the relevant blog to the query with meaningful summarization and minimal number of blog. The statistics of blogs collected from various blog sources and the percentage of relevancy of blog for selected keywords using blog search engines and SBMF is shown in Table 2. Figure 8 shows the number of blogs retrieved in different search engines for selected keywords. Figure 9 shows the relevancy of blogs for the selected keyword using various blog search engines and blog search using SBMF. The blogs retrieved using search engines collect huge amount of irrelevant blogs.

Blogs found: 16
pages displayed below

| Rank | Blog Title | Score |
|------|--|--------|
| 1 | Ben's Blog World: iOS 4.1 walkthrough http://benstiffler.blogspot.com/2010/09/ios-41-walkthrough.html Date: 06-09-2010 Author: benstiffler iOS 4.1 walkthrough Complete feature guide to Apples iOS 4.1 update for iPhone and iPod touch iOS 4.1 is a relative small update to iOS 4 that fixes some important bugs, adds a few new features, and powers the all new iPod touch 4. Steve Jobs announced the headlines during Apples 2010 special music event and they include Game [1 walkthrough is a story by TiPb. This feed is sponsored by The iPhone Blog Store. TiPb - The #1 iPhone, iPad, and iPod touch Blog Complete feature guide to Apples iOS 4.1 update for iPhone and iPod touch iOS 4.1 is a relative small update to iOS 4 that fixes some important bugs, adds a few new features, and powers the all new iPod touch 4. Steve Jobs announced the headlines during Apples 2010 special music event and they include Game Center social gaming network, TV rentals, iTunes Ping social music network, HDR photography, and HD video uploads. ... feature guide to Apple's iOS 4.1 update for iPhone ... Apple.com/uk says iOS 4.1 coming September 8. Motorola Charm -- hands on with T-Mobile's tiny me... | 0.0265 |
| 2 | Photo Infos: Apple's iOS 4.2 Available Today for iPad, iPhone ... http://photoinfos.blogspot.com/2010/11/apples-ios-42-available-today-for-ipad.html Date: 22-11-2010 Author: Henrik Lehnerer | 0.0088 |

Fig. 5. Summarized top ranked blogs for keyword “apple”

15: **Rotolo Media: Well Said! Re: Apple's iOS** Score: 0.0015
<http://rotolo-media.blogspot.com/2010/11/i-couldnt-have-said-it-better-re-apples.html>
 Date: Unknown Author: Rotolo Media / Wednesday, November 10, 2010 /
 original post From commentor B@M@n at LifeHacker re: How To Google-ify Your iPhone "I'm an iPhone owner so read on b-4 you dismiss. Apple doesn't want to provide me with the luxury of choice or flexibility when it comes to my contact management and many other subjects, media, web browsing, etc. It's a great phone but the corporate attitude is clearly not in favor of empowering consumers anymore. This is further pointed out by the app store in Lion. I would hate to give up my Macbook love if Lion is the sign of things to come.

skip to main | skip to sidebar. Well Said! Re: Apple's iOS. original post ... Apple doesn't want to provide me with the luxury of choice or flexibility when ...

16: **Apple iOS 4: Meet Apple's iOS 4** Score: 0.0
<http://nugglepurl.blogspot.com/2010/06/apple-ios-4-meet-apples-ios-4.html>
 Date: 08-06-2010 Author: Unknown
 Share This Story to Your Friends. Tweet The new iOS 4 Software Update includes multitasking, folders, even better Mail, iBooks, and more. The OS, which contains around 100 new and upgraded features, will be rolled out with the new iPhone 4 next month, and it will also be available as a free software update on earlier devices, including the iPad, iPhone 3G, iPhone 3GS, and second and third generation iPod touch with 32GB or 64GB storage. Watch iPhone 4 Multitasking :iOSs 4 features:1. Create playlists - Create custom playlists right on your iPhone.

Fig. 6. Summarized bottom tanked blogs for the keyword “apple”

| Bin name (# of blogs) | |
|---|------|
| EDUCATION_DEPARTMENT | (33) |
| HIGHER_EDUCATION | (22) |
| MEDICAL | (21) |
| ENGINEERING_AND_TECHNOLOGY | (26) |
| CHEMICAL_ENGINEERING | (28) |
| ELECTRONICS_AND_COMMUNICATION_ENGINEERING | (20) |
| MOBILE_COMMUNICATION | (20) |
| MOBILE_COMPUTING | (25) |
| MOBILE_DEVICES | (25) |
| MOBILE_OPERATING_SYSYSTEM | (15) |
| RIMS_BLACKBERRY_OS | (25) |
| ANROID | (17) |
| WINDOWS_MOBILE | (17) |
| APPLES_IOS | (16) |
| MICROCONTROLLER_AND_MICROPROCESSOR | (25) |
| PROCESSOR_COMPUTER | (19) |
| AMD | (29) |
| CYRIX | (23) |
| INTEL | (30) |
| INTEL_CORE_TO_DUO_PROCESSOR | (24) |
| INTEL_QUAD_CORE_PROCESSOR | (30) |
| I3_PROCESSOR | (17) |
| I5_PROCESSOR | (24) |
| I7_PROCESSOR | (28) |
| PENTIUM_III_450_MHZ | (30) |
| PENTIUM_MOBILE_PROCESSOR | (10) |
| PENTIUM_IV_1.3GHZ | (17) |
| PENTIUM_II_233_MHZ | (27) |
| INTEL_DUAL_CORE_PROCESSOR | (18) |

Fig. 7. Blog repository summary

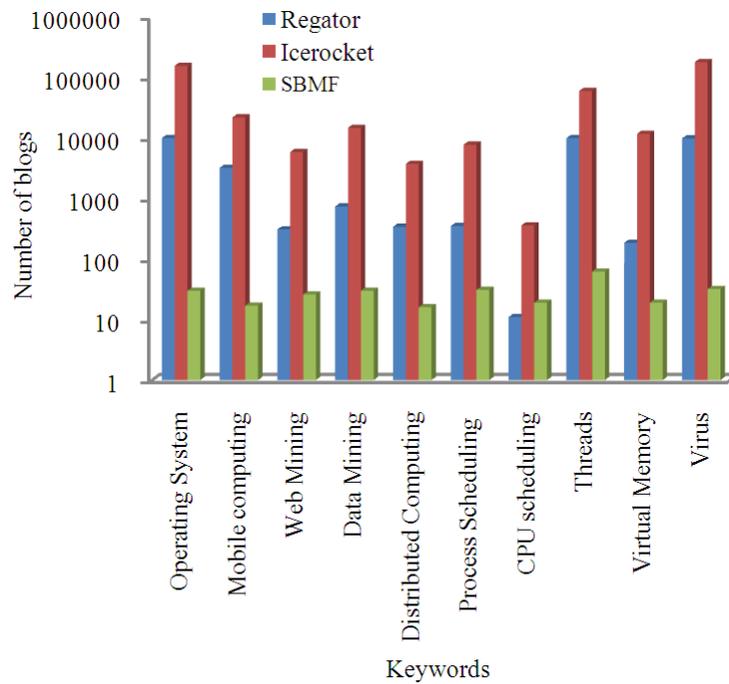


Fig. 8. Blogs collected using different search engines for Keywords

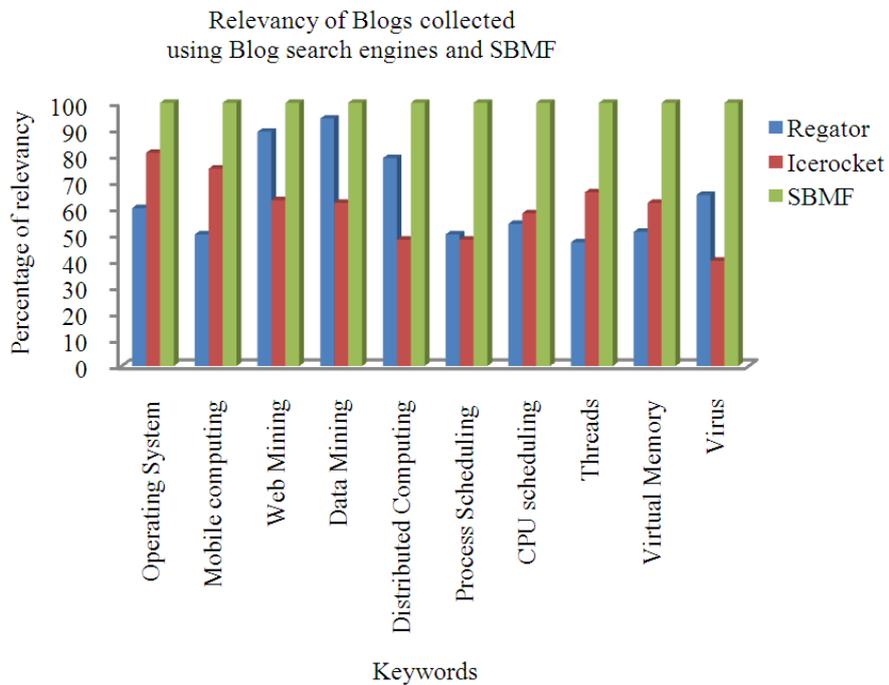


Fig. 9. The percentage of relevancy of blogs in blog search engines and SBMF

Table 2. Percentage of relevancy of blogs using blog search engines and SBMF

| Keywords | Technorati | | Regator | | Icerocket | | SBMF | |
|-----------------------|--------------|---------------------|--------------|---------------------|--------------|---------------------|--------------|---------------------|
| | No. of blogs | Blogs relevancy (%) |
| Operating system | 141 | 74 | 10000+ | 60 | 157411 | 81 | 30 | 100 |
| Mobile computing | 48 | 88 | 3229 | 50 | 22079 | 75 | 17 | 100 |
| Web mining | 4 | 75 | 309 | 89 | 5903 | 63 | 26 | 100 |
| Data mining | 13 | 100 | 742 | 94 | 14826 | 62 | 30 | 100 |
| Distributed computing | 4 | 75 | 342 | 79 | 3757 | 48 | 16 | 100 |
| Process scheduling | 2 | 0 | 352 | 50 | 7801 | 48 | 31 | 100 |
| CPU scheduling | 0 | 0 | 11 | 54 | 361 | 58 | 19 | 100 |
| Threads | 27 | 73 | 10000+ | 47 | 60451 | 66 | 62 | 100 |
| Virtual memory | 4 | 100 | 186 | 51 | 11777 | 62 | 19 | 100 |
| Virus | 67 | 31 | 9899 | 65 | 181415 | 40 | 32 | 100 |

4. DISCUSSION

The experimental result shows that the SBMF yields better results. The bar chart shows that the number of blogs retrieved for each keyword in blog search engines is very high and the relevancy is very low. Semantic blog mining framework uses the ontology to collect the relevant blogs from the web, then remove the irrelevant blog and create the relationship between blogs before storing in the blog repository. In this framework the blogs stored are relevant to the subject which makes the search process easier and reduce the search time as well as the user ambiguity. Since the collected blogs are preprocessed and semantically related, only the relevant blogs are retrieved for the user. Hence the relevancy of blog is hundredpercentage.

5. CONCLUSION

This study introduces a novel framework to collect the blogs from various blog sources. Then the collected blogs are preprocessed and semantically annotated blogs are stored in a blog repository. User searches the repository for their query to obtain the relevant result. The system is evaluated for the education domain specific ontology. The experimental result shows that this framework yields better result than the blog search engine result. This study was compared with existing blog search engines like Technorati, blogpulse, blogscope, icerocket and regator. Blog Summarizer retrieves only the relevant blog to the query with meaningful summarization and minimal number of blog. The same can be extended to any kind of applications in blog mining. In future, different domain specific

ontology can be used to collect blogs from different sources and similar relevancy can be studied.

6. REFERENCES

- Bao, Z., J. Lu, T.W. Ling and B. Chen, 2010. Towards an effective XML keyword search. *IEEE Trans. Knowl. Data Eng.*, 22: 1077-1092. DOI: 10.1109/TKDE.2010.63
- Breslin, J.G., A. Passant and S. Decker, 2009. *The Social Semantic Web*. 1st Edn., Springer, Heidelberg, ISBN-10: 3642011713, pp: 300.
- Dominich, S., 2008. *The Modern Algebra of Information Retrieval*. 1st Edn., Springer, Berlin, ISBN-10: 3540776591, pp: 327.
- Fattah, M.A and F. Ren, 2009. GA, MR, FFNN, PNN and GMM based models for automatic text summarization. *J. Comput. Speech Language*, 23: 126-144. DOI: 10.1016/j.csl.2008.04.002
- Fujmura, K., T. Inoue and M. Sugisaki, 2005. *The EigenRumor algorithm for ranking blogs*. NTT Cyber Solutions Laboratories.
- Kritikopoulos, A., M. Sideri and I. Varlamis, 2006. BlogRank: Ranking weblogs based on connectivity and similarity features. *Proceedings of the 2nd international Workshop on Advanced Architectures and Algorithms for Internet Delivery and Applications, (AAA-IDEA '06)*, ACM Press, New York, USA. DOI: 10.1145/1190183.1190193
- Li, Y.M. and C.W. Chen, 2009. A synthetical approach for blog recommendation: Combining trust, social relation and semantic analysis. *J. Expert Syst. Appli.*, 36: 6536-6547. DOI: 10.1016/j.eswa.2008.07.077

- Lilac, A.S. and A. Nadeen, 2010. Educational advertising ontology: A domain-dependent ontology. *J. Comput. Sci.*, 6: 1070-1077. DOI: 10.3844/jcssp.2010.1070.1077
- Lin, H.T., T.H. Kuo and S.M. Yuan, 2007. A web-based learning portfolio framework built on blog services. *Inform. Technol. J.*, 6: 858-864.
- Liu, C., D. Yang, Y. Wang and Q. Pan, 2011. Web intelligence analysis in the semantic web based on domain ontology. *Inform. Technol. J.*, 10: 2343-2349.
- Liu, L.Q., 2008. Learning resources recommendation system based on education blog. Proceedings of the International Conference on Computer Science and Software Engineering, Dec. 12-14, IEEE Xplore Press, Wuhan, Hubei, pp: 805-807. DOI: 10.1109/CSSE.2008.62
- Lu, Y. and H. Lee, 2008. Blog Community Discovery Based on Tag Data Clustering. Proceedings of the Pacific-Asia Workshop on Computational Intelligence and Industrial Application, Dec. 19-20, IEEE Xplore Press, Wuhan, pp: 14-18. DOI: 10.1109/PACIIA.2008.144
- Mithun, S. and L. Kosseim, 2009. Summarizing blog entries versus news texts. Proceedings of the Workshop on Events in Emerging Text Types, (ETT' 09), Association for Computational Linguistics Stroudsburg, PA, USA., pp: 1-8.
- Mohanraj, V. and M. Chandrasekaran, 2011. An ontology based approach to implement the online recommendation system. *J. Comput. Sci.*, 7: 573-581. DOI: 10.3844/jcssp.2011.573.581
- Motik, B., B.C. Grau, I. Horrocks and U. Sattler, 2009. Representing ontologies using description logics, description graphs and rules. *Artif. Intell.*, 173: 1275-1309. DOI: 10.1016/j.artint.2009.06.003
- Ouyang, Y., W. Li, S. Li and Q. Lu, 2011. Applying regression models to query-focused multi-document summarization. *J. Inform. Process. Manage.*, 47: 227-237. DOI: 10.1016/j.ipm.2010.03.005
- Sathianesan, G.W. and S. Sankaranarayanan, 2012. Personalized semantic based blog retrieval. *J. Comput. Sci. Technol.*, 27: 591-598. DOI: 10.1007/s11390-012-1246-8
- Shen, J., Y. Zhu, H. Zhanj, C. Chen and R. Sun *et al.*, 2008. A content-based Algorithm for Blog Ranking. Proceedings of the International Conference on Internet Computing in Science and Engineering, Jan. 28-29, IEEE Xplore Press, Harbin, pp: 19-22. DOI: 10.1109/ICICSE.2008.11
- Tao, X., Y. Li and N. Zhong, 2011. A personalized ontology model for web information gathering. *IEEE Trans. Knowl. Data Eng.*, 23: 496-511. DOI: 10.1109/TKDE.2010.145
- Tayebi, M.A., S.M. Hashemi and A. Mohades, 2007. B2Rank: An algorithm for ranking blogs based on behavioral features. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, (CWI' 07), IEEE Computer Society Washington, DC., pp: 104-107.
- Wu, J., 2012. A framework for learning comprehensible theories in XML document classification. *IEEE Trans. Knowl. Data Eng.*, 24: 1-14. DOI: 10.1109/TKDE.2011.158
- Yuhanget, Y., H. Yu, T. Zhao, H. Tan and Dequan Zheng, 2008. An efficient approach to important bloggersdiscovery. *J. Elect.*, 25: 218-225. DOI: 10.1007/s11767-006-0135-0
- Zavitsanos, E., G. Paliouras and G.A. Vouros, 2011. Gold standard evaluation of ontology learning methods through ontology transformation and alignment. *IEEE Trans. Knowl. Data Eng.*, 23: 1635-1648. DOI: 10.1109/TKDE.2010.195