

Blog Designing and Searching Methodologies: A Review

Harsh Khatter^{*}, Brij Mohan Kalra^{**}

(Department of Computer Science, Ajay Kumar Garg Engineering College, Ghaziabad, India)

ABSTRACT

Now, Blogs are getting popular day by day. Blogs are like an online dairy created by individuals and stored on the internet. As Blog is a type of website, various blogging sites can provide excellent information on many topics, although content can be subjective. Blogs are one of the main components of Web 2.0. Paper consists of description of various Blog site designs and searching methods with their research gaps. The major characteristics and features of blogs are also highlighted.

Keywords – Blogs, Internet, Searching, Web 2.0, Web Tools.

I. INTRODUCTION

In this growing world, Web services are the part of everyone's life. From the traditional Web 1.0, read only web, which only includes chat, email, instant messaging, now switches to a new Web, named Web 2.0. Web 2.0 consists various tools and services which provides read write interface to their users. There are a large number of Web 2.0 tools: Blogs, Discussion Forums, Wikis, Social Networks, Social Bookmarking sites, Podcasting, Online Communities, RSS and Atom feeds, and many more. But, apart from all these tools, Blogs are the only tool whose intent is personal, even a lot of expertise are also present there to share their ideas and views with other persons of similar interest.

Blogs are websites that allow one or more individuals to write about things they want to share with others. The universe of all blog sites is referred to as Blogosphere [1]. Blog, a contraction of the term "web log" is a personal online diary that is frequently updated and intended for public consumption. Now to some extent it is a type of websites. People usually create a blog as a hobby to share their information and experience on a particular subject. Entries are commonly displayed in a reverse-chronological order [2]. Blogging software allows users to publish opinions, views, and ideas on any topic. Analysis of linkage between blogs has indicated that community forming in blogosphere is not a random process but is a result of shared interests binding bloggers together.

Learning, analysis and usage of the user's interest and social linkage from the blog is therefore necessary to provide useful search faculty on the blogosphere to bloggers and revenue generation opportunities like advertising to the blog service providers [3]. The act of posting to a blog is called blogging and the distributed, collective, and interlinked world of blogging is the blogosphere [4].

II. BLOGS

Blogs are the type of websites. Personal interests create Blogs. Based on the working and designing of blogs, numbers of characteristics are defined below. Users can create a new blog post, add blog post, share, rate, and comment the blog posts. For all these operations, user has to login first. Purpose of Blog is to share ideas and views among a group of people all around the world. Intent of Blog is personal. Discussions are done in the form of comments. All the posts are shown in reverse chronological order i.e. latest blog post shown on top. There is a list of potential benefits of blogs, which is mentioned below:

- Can promote analogical thinking.
- Potential for increased access and exposure to quality information.
- Combination of solitary and social interaction.
- Can promote critical and analytical thinking.
- Can promote creative, intuitive and associational thinking (creative and associational thinking in relation to blogs being used as brainstorming tool and also as a resource for interlinking, commenting on interlinked ideas).

III. REVIEW TO BLOG DESIGNING AND SEARCHING METHODOLOGIES

Beyond serving as online diaries, weblogs have evolved into complex social structures. Blogging software allows users to publish opinions on any topic without any constraints on the predefined schema.

3.1 Designing of Blogs

Blogs might be of many types. Personalized Blog is one of the most impressive categories of Blogs where the blog posts shown to the user are of his own interest. Some major works in this area are discussed below.

R. Adhikari et al. mentions that it is easy and simple to create blog posts and their free form and unedited nature have made the blogosphere a rich and unique source of data, which has attracted people and companies across disciplines to exploit it for varied purposes. The valuable data contained in posts from a large number of users across geographic, demographic and cultural boundaries provide a rich data source not only for commercial exploitation but also for psychological & sociopolitical research. Basically researchers tried to demonstrate the plausibility of the idea through clustering and opinion mining experiment on analysis of blog posts on recent socio-political developments in the new democratic republic of Nepal; and to elaborate the broader technical framework & tools required for this kind of analysis [1].

Similarly CHENG Tao et al. discusses about the Virtual enterprise (VE), which is an effective and collaborative way to jointly face the great pressures from quickly growing globalization and world-wide market competition. Furthermore, a wiki & blog-based knowledge-sharing mechanism and its prototype system are designed for supporting enterprises to inter-communicate, share knowledge and manage knowledge within a VE environment [5].

Various models are already evolved related to the blogs and blogosphere. Tse-Ming Tsai et al. recommends applying the three dimensions of value, semantic, and the social models to the emerging Blogosphere and improving the user experience for the bloggers in gathering the featured items. As per the previous works done, approaches discussed may not be comprehensive enough since the way people use blogs continues to evolve [6].

Bi Chen et al. proposed three models by combining content, temporal, social dimensions: the general blogging-behavior model, the profile-based blogging-behavior model and the socialnetwork and profile-based blogging-behavior model. These models are based on two regression techniques: Extreme Learning Machine (ELM), and Modified General Regression Neural Network (MGRNN). In paper, the empirical evaluation is done on DailyKos, a political blog, one of the largest blogs, which produce good results for the most active bloggers and can be used to predict blogging behavior [7].

Yin ZHANG et al. discussed that Clustered Web pages, such as blog posts, could be used to improve Web search. In the paper, authors proposed an extending framework using relations in the

Blogosphere and demonstrate how the framework could be used to help clustering blog posts. Evaluation of the framework with content-based extending approach is done. Experiment results show that the framework does help the clustering process [8].

ZHOU Ping proposed an algorithm of personalized blog information retrieval based on user's interest model. The paper discusses the system architecture of personalized blog information retrieval and studies the identification module of blog webpage [2].

As per Michael Chau et al., blogs are very dynamic, so it isn't as straightforward to apply traditional Web mining techniques to them. They suggest that a general blog framework created for different tasks must consists of a blog spider, a blog parser, a blog content analyzer, a blog network analyzer, and a blog visualize [9].

And a framework, BlogHarvest, for blog mining and search is demonstrated by Joshi et al. This framework extracts the interests of the blogger, finds and recommends blogs with similar topics and provides blog oriented search functionality [3].

3.2 Searching the blog posts

Over the past decades, various searching techniques are come into existence with the growth of World Wide Web. From the starting of the Web era, various searching methods, techniques and types of searching algorithms are introduced and as per searching requirement, they are used. There are lots of searching methods in which search can be done on keywords, on queries, on topics, on phrases, on pages, etc. The query based and topic based search is used in forums, whereas the page search or phrase search is used in search engines where the exact finding is required. Rest of all websites use keyword based searching. Keyword based searching provides an easier way to search the contents on internet. In the same way, maximum number of websites use keyword based searching. A review of searching algorithms and methods in brief is given below.

Initially, the searching is done using the Query Tree. Top-down Approach is followed to search the results. But it is a traditional method where indexing is use to Reduce the Complexity. In this, A* Graph algorithm is used which keeps track on visited node and distance travelled [10]. After this method, next approach was Proximity Search. In this method, analysis of textual proximity of keyword is done. Focus is on queries based on general relationship

among objects where proximity is defined based on shortest paths between objects. Figure 1 shows the working based on this proximity approach [11].

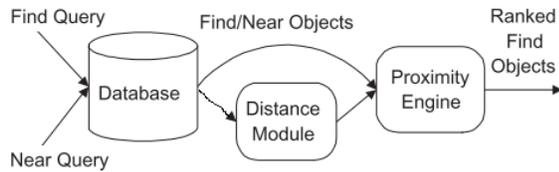


Fig. 1: Proximity Search Design [11].

To search the content on internet, the method introduced was “BANKS: Browsing & Keyword Searching” [12]. It is best to use for Relational databases and static data. In this method keyword search is used. When we talk about searching the content on web, the term semantics of data comes in mind. A semantic web portal for ontology searching, ranking and classification is the next approach. Chintan Patel et al. discuss a model for this. Model consists of crawling & classification of content. Then on the basis of page rank, Ranking has to be done. Searching is based on Context Oriented Query Language and a Machine Interface is well defined in the model [13]. The model has been implemented using statistics, recall numbers, etc. Some minor changes has been done in this model which was discussed by *Xing Jiang and Ah-Hwee Tan*. They introduced Description Logic and Fuzzy Description Logic based on the queries [14].

In previous methods and approaches to search the keyword data, the problem was “keyword queries are weak to express”. Gjergji Kasneci et al. discussed a framework which consists of Data Model, Query language, and Ranking Model. They called it NAGA, Network Assisted Genetic Algorithm. This performed both searching and ranking on the data [15]. The major thing is to understanding the user goals for Web Search. Daniel E. Rose and Danny Levinson discussed three parameters which concentrate on what the user exactly wants. Parameters are Navigation, Informational, and Resource [16].

Keyword search queries might be in structured, unstructured or semi structured form. For unstructured queries, Pavel Calado et al. suggest Bayesian Networks. They suggest a Bayesian network approach to searching web databases through Keyword-based queries [17]. **Guoliang Li** et al. suggested an efficient 3-in-1 keyword search method which works for all types of data i.e. Unstructured, Semi-structured and Structured. Indexing & querying

of large collections of heterogeneous data is used in this method. Authors implemented it using graphs and graph indices. They conclude that this is an efficient & adaptive keyword search of all kinds of data words [18].

Georgia Koutrika et al. discusses the searching of keywords over Structured Data in a cloud. Method uses a coupling of keywords. Tags used in this method are unstructured, whereas, clouds of data contains structured data, say, Data Cloud [19]. In 2011, an effectively interpreting keyword queries on RDF databases is discussed by *Haizhou Fu and Kemafor Anyanwu*. Before this method, heuristics were used for interpreting the keyword queries. But heuristics fails to capture user dependent queries. Here, the sequences of structured queries are used and the main work is done by query interpretation. Only the Top-k-aware queries are considered and discussed approach is called context aware approach [20].

IV. FINDINGS

Blogs are a source of enormous information. For a user it is very hard to get the relevant information from the huge network of World Wide Web. For bloggers and frequent blog readers, it is virtually impossible to keep track of the growing blogosphere and hence a service recommending the blogs matching their interests will seek high value. Blogs are the important source of information, but to get the relevant information in an efficient time is a typical task. Blog mining is an important way for people to extract useful information. Blogs are very dynamic, so it isn't as straightforward to apply traditional Web mining techniques to them. The goal is to provide the user with reliable and accurate blog information conveniently.

After taking a complete review, the gaps and problems are discussed in two parts. First is related to the designing and the architecture and working of Blog. Second is related to the problems and gaps in searching of content/blog posts in Blogs.

4.1 Based on Blog designing

Based on the design and the architecture of the blog, the efficient way to use blogs are as personalized blogs where the blog posts shown to the user are as per his own interest, irrelevant posts are not shown to the user. There is no as such system, which integrates both, an individual blog as well as a blog search engine. This kind of integration provides an additional facility to the user, which improves the knowledge and searching experience of the user. The

model of the blog must be easier to operate and handle by both, user and the developer.

4.2 Based on Searching Methods

Based on searching of blog posts, the searching method must be appropriate to search the results from all types of data i.e. structured, unstructured, and semi-structured. If a search method will search the results, only of single data type then user will not be able to fetch all relevant blog posts. The method discussed by **Guoliang Li** et al. is a better option to use in Blogs [18]. Search will be efficient and there must be an optimized query to search the Blog posts. Some minor changes and some add on services, will make the best searching results.

V. COLLABORATION OF BLOG WITH OTHER WEB TOOLS

Blog is a web tool handled by an individual. There are various other web tools like Social Networking Sites (SNS), Discussion Forums, Wikis, Online Communities, etc. Each tool has its own important and provides better results as per user's field of interest. Integration of these tools will provide an ease to the user to use best services of Web 2.0. Web 2.0 is a term used for read – write Web. User can read as well as write the content on the World Wide Web (WWW). The proper collaboration of these Web 2.0 tools will provide a new platform to its users to learn things more easily, to search things, to communicate with others i.e. friends, expertise, guides, or people of similar interests. In present scenario, various RSS and Atom feeds are available to collaborate these external links to any other site, may be a Blog, Wiki, Discussion Forum, Online Community, or Social Networking Site. As Web 2.0 came as an evolution in internet world, the collaboration of its tools will be an evolution in informal eLearning world in the same way.

VI. CONCLUSION

The major part of knowledge and recent activities are shared using blogs. After taking a review of designing and searching methods of Blogs, various research gaps and their respective findings are well discussed in the section IV. An innovative idea of collaboration of various Web 2.0 tools with Blogs is given. These new ideas and suggestions will surely improve the knowledge and searching experience of bloggers. As Blogs is getting popularity day by day, so, in future Blogs will play an important role in increasing the informal learning. Moreover, the collaboration with RSS and Atom feeds, the power of blogs will become more than twice. Therefore, there is a need to improve the Blog designing and searching

methodologies, so that the user can get what he exactly wants in an efficient manner and with ease of operability.

REFERENCES

- [1] V. K. Singh, D. Mahata and R. Adhikari, Mining the Blogosphere from a Socio-political Perspective, *International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, 2010, 365 – 370.
- [2] Zhou Ping, Research on Personalized Blog Information Retrieval, *International Conference on Web Information Systems and Mining (WISM)*, 2010, 289 – 292.
- [3] Mukul Joshi and Nikhil Belsare, BlogHarvest: Blog Mining and Search Framework, *International Conference on Management of Data COMAD*, 2006.
- [4] Peter Duffy and Axel Bruns, The use of blogs, wikis and RSS in education: A conversation of possibilities, *Learning and Teaching Conference*, 2006.
- [5] Cheng Tao, Peng Xiaobo, Feng Ping, Du Jianming, Research on Design of A Wiki & Blog-based Knowledge-sharing Mechanism for Virtual Enterprise, *Third International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, 2011, 1133 – 1137.
- [6] Tse-Ming Tsai, Chia-Chun Shih, Seng-cho T. Chou, Personalized Blog Recommendation Using the Value, Semantic, and Social Model, *International Conference on Innovations in Information Technology*, 2006, 1 – 5.
- [7] Bi Chen, Qiankun Zhao, Bingjun Sun, Mitra P. , Predicting Blogging Behavior Using Temporal and Social Networks, *Seventh IEEE International Conference on Data Mining, ICDM*, 2006, 439 – 444.
- [8] Yin Zhang, Kening Gao, Bin Zhang, Jinhua Guo, Feihang Gao, Pengwei Guo, Clustering Blog Posts Using Tags and Relations in the Blogosphere, *1st International Conference on Information Science and Engineering (ICISE)*, 2010, 817 – 820.
- [9] Chau, M., Lam, P., Shiu, B., Xu, J., Jinwei Cao, A Blog Mining Framework, *International Journal of IT Professional*, vol. 11, 2009, 36 - 41.
- [10] ennis Shasha, Jason T. L. Wang, Rosalba Giugno, Algorithmics and applications of tree and graph searching, *twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, 2002, 39 – 52.

- [11] Roy Goldman, Narayanan Shivakumar, Suresh Venkatasubramanian, Hector Garcia-Molina, Proximity Search in Databases, *24th International Conference on Very Large Data Bases*, 1998, 26 – 37.
- [12] B. Aditya, Gaurav Bhalotia, Soumen Chakrabarti, Arvind Hulgeri, Charuta Nakhe, Parag Parag, S. Sudarshan, Keyword searching and browsing in databases using BANKS, *28th international conference on Very Large Data Bases*, 2002, 1083 – 1086.
- [13] Chintan Patel, Kaustubh Supekar, Yugyung Lee, E. K. Park, OntoKhoj: a semantic web portal for ontology searching, ranking and classification, *5th ACM international workshop on Web information and data management*, 2003, 58-61.
- [14] Xing Jiang, Ah-Hwee Tan, OntoSearch: a full-text search engine for the semantic web, *21st national conference on Artificial intelligence*, vol. 2, 2006, 1325-1330.
- [15] Gjergji Kasneci, Fabian M. Suchanek, Georgiana Ifrim, Maya Ramanath, Gerhard Weikum, NAGA: Searching and Ranking Knowledge, *24th IEEE International Conference on Data Engineering*, 2008, pp. 953 – 962.
- [16] Daniel E. Rose, and Danny Levinson, Understanding user goals in web search, *13th international conference on World Wide Web*, 2004, 13-19.
- [17] Pavel Calado, Altigran S. da Silva, Alberto H.F. Laender, Berthier A. Ribeiro-Neto, Rodrigo C. Vieira, A Bayesian network approach to searching Web databases through keyword-based queries, *International Journal of Information Processing and Management on Bayesian networks and information retrieval*, vol. 40, 2004, 773-790.
- [18] Guoliang Li, Beng Chin Ooi, Jianhua Feng, Jianyong Wang, Lizhu Zhou. EASE: an effective 3-in-1 keyword search method for unstructured, semi-structured and structured data, *International conference on Management of data ACM SIGMOD*, 2008, 903-914.
- [19] Georgia Koutrika, Zahra Mohammadi Zadeh, Hector Garcia-Molina, Data clouds: summarizing keyword search results over structured data, *12th International Conference on Extending Database Technology: Advances in Database Technology*, 2009, 391-402.

- [20] Haizhou Fu, and Kemafor Anyanwu, Effectively interpreting keyword queries on RDF databases with a rear view, *10th international conference on semantic web*, 2011, 193-208.



Harsh Khatter is a postgraduate student, pursuing Master of Technology in Computer Science and Engineering from Mahamaya Technical University, Noida, India. He received his Bachelor's degree in 2010. As thesis subject, He is working on Web 2.0 tool, Blogs. His research interests include Web Services, Data Mining and Databases. He has published a research paper on databases and data mining in Elsevier international journal, one paper on Informal eLearning in ICIAICT international conference, and one in national conference. He is also a member of IEEE Society.



Brij Mohan Kalra is currently working as a Professor and Head in the Department of Computer Science and Engineering at Ajay Kumar Garg Engineering College, Ghaziabad, India. He has done his B.Tech. from Delhi College of Engineering, Delhi in 1977 and completed his M.Tech from IIT, Delhi in 1991. He has vast experience of 35 years of academia and industry in CSE and IT fields. He is pursuing his Ph.D in CSE from Gautam Buddha University, Greater Noida, India. His research interests include eLearning, Computer Networks, and Digital Logic Design. He is also a member of several professional bodies: IEEE, CSI, and IETE.