

BlogGrid: Towards an Efficient Information Pushing Service on Blogspace

Jason J. Jung^{1,2}, Inay Ha¹, and Geun-Sik Jo¹

¹ School of Computer Engineering, Inha University,
253 Yonghyun-dong, Incheon, Korea 402-751
jjjung@intelligent.pe.kr, inay@eslab.inha.ac.kr,
gsjo@inha.ac.kr

² INRIA Rhone-Alpes,
655, avenue de l'Europe, Montbonnot St. Martin,
38334 Saint-Ismier cedex, France

Abstract. With increasing concerns about the personalized information space, users have been posting various types of information on their own blogs. Due to the domain-specific properties of blogging systems, however, searching relevant information is too difficult. In this paper, we focus on analyzing the user behaviors on blogspace, so that the channel between two similar users can be virtually generated. Eventually, social grid environment can be constructed on blog organization. We therefore propose a BlogGrid framework to provide the information pushing service without any user intervention.

1 Introduction

Many kinds of personalized information spaces have been developed. As a representative example, a blog (originally Weblog) is a Web application presented as a web page consisting of periodic posts, normally in reverse chronological order [8]. Since Jorn Barger coined this idea in 1997, the usage of blogs has been spread from individual diaries to arms of political campaigns, media programs and corporations, and from the writing of one occasional author to the collaboration of a large community of writers [4]. More importantly, each blogger can make explicit connections with others such as families, friends, and colleagues. Through these social activities, indeed, bloggers can organize communities in a form of blogs. In summary, we note two main features of blogs, as follows.

- **Personal content management.** Bloggers can create and manage various forms of contents such as personal history, commentaries, and photos.
- **Information propagation by social activities.** Along with the social links, they can have access to the other's blogs. It is similar to browsing the hypertext documents.

However, we have faced on two serious problems. Firstly, a large amount of information on blogspace is overwhelming to users. It means that bloggers are getting more overloaded to search relevant information. The blog-tracking company Technorati, Inc., reports almost 4.2 million Weblogs worldwide as of October 2004, up from about one million a year earlier. And a 2003 Pew/Internet survey estimated that more than 53 American adults, or 44% of Internet users, had used the Internet to

publish their thoughts, respond to others, post pictures, share files, and otherwise contribute to the content available online [8].

Second problem, which is much more serious than a former, is that most of communities tend to be organized by the private relations between bloggers, e.g., nepotism. Not only the community policy for protecting the privacy of members but also the isolation of social structure causes the information flows on blogosphere to be limited.

In this paper, we exploit the grid computing paradigm, which is capable of supporting a efficient framework of information sharing between heterogeneous sources, to blogspaces and bloggers. Thereby, each blogger's behaviors should be captured and analyzed to extract what topics he is (or they are) interested in and, more exactly, what he has been trying to search. On this grid environment for information sharing on blogosphere, called BlogGrid, each user can be provided with information pushing service. This service consists of three steps, which are construction of latent channels, organization of transparent communities, and information pushing. The goal of this service is efficient information diffusion on blogosphere. It means that a certain information should be delivered to the bloggers who have been looking for that information as quickly as possible, regardless of the social distances and topologies.

2 Modeling Behaviors on Blogspace

We assume that each action taken by bloggers have implicit meaning and be able to be applied to extract useful information. In order to model behaviors on blogspace, we need to thoroughly note the roles of blogspaces and bloggers. Blogspaces are divided into personal and community space [7]. Each user is able not only to possess his own personal blogspace, but also to take part in community blogspaces as a member. Due to the particular purposes, bloggers conduct the following behaviors.

- Linking with others. It is to establish a social network such as families, friends, and colleagues. More importantly, in order to share information about a particular topic, people can organize a community and actively participate it.
- Posting articles. It is the most basic action on blogspaces. As taking this action, bloggers can input various types of information and enrich blogspaces. This action explicitly represents the corresponding blogger's preferences. For example, he is probably interested in "music", because he has been posting articles about recital schedules and musical instruments.
- Navigating. In order to get relevant information within a blog system, people can visit other blogspaces. Bloggers can navigate the other blogspaces by the following two methods. i) Random browsing. Bloggers can "randomly" jump into other blogspaces. In fact, this has been the only way to deal with nepotism problem. ii) Accessing to neighbors on social network. By referring to the list of neighbors, bloggers easily move into their blogs.
- Responding. Bloggers can respond to a certain article while navigating blogspace. Their responses like personal opinions, sympathies, antipathies, or even apathy are express as the following two ways: i) *Comment*. A blogger can leave his responses which consist of short text sentences. More than a comment can be serially and continuously attached to each post article. ii) *Trackback*. In contrast to comments,

it allows a blogger to see who has seen the original post and has written another entry concerning it. It works by sending a ‘ping’ between the blogs, and therefore providing the alert. Trackback typically appears below a blog entry and shows a summary of what has been written on the target blog, together with a URL and the name of the blog. Since implemented by Movable Type, most blogging systems have adopted trackback mechanism as generic function.

Not only free-text sentence but also numeric rating format (e.g., from 0 to 5) and voting format (e.g., “Yes” or “No”) can be applied to reflect the corresponding blogger’s interests and opinions.

Another interesting feature is that the responding can be nested. It means that bloggers can respond to a certain comment already attached to articles.

Above all, a community blogspace is supposed to utilize categories and manage them. This category is very similar to simple taxonomies and topic hierarchies. Examples of such categories are “Music> Blues>“, “Computer Science> Artificial Intelligence>“, and “Conference> Grid Computing> GCC>“. It makes every articles posted in his blogspace well-organized and semantically annotated.

Most importantly, bloggers have concerned on dealing with the difficulties of navigating blogspaces. Thereby, in this paper, we have focused on extracting behavioral patterns of responses, rather than analyzing free texts posted on the user’s blogspace.

A blogger finally is modeled with his own preference information extracted from his behaviors on blogspaces. It is represented as a set of hierarchical topics, which is derived from particular bloggers’ categories. Suppose that user A attach his comments to blogspaces B which are annotated to a set of categories C_A

$$C_A = \{[c_1, w_{c_1}], [c_2, w_{c_2}], \dots, [c_n, w_{c_n}]\} \quad (1)$$

$$w_{c_i} = \frac{|c_i|}{n} \quad (2)$$

where n is the total number of different categories and w_{c_i} is the weight value of i -th category c_i . The summation of weight values is equal to one for normalization.

3 Information Pushing on BlogGrid

Information should be pushed to the bloggers who are looking for it. This service is simply composed of two main processes; organization of latent communities and recommendation by facilitator agent.

3.1 Transparent Organization of Latent Communities

In order to construct latent communities, a group of bloggers who are interested in same topics should be obtained. As measuring the similarity between two bloggers, we can infer their relationship and make a decision whether they should be participated in a same community together or not. Especially, this task should be

transparently done, which means it has to be performed without user’s intervention and realization. Simple scheme for measuring the similarity between two users A and B is based on the equation

$$Sim(A, B) = H[C_A, C_B, \cap_{k=1}^K \{\tilde{\sigma}(C_A, C_B)\}] \tag{3}$$

where the function $\tilde{\sigma}$ is to obtain the K common elements from both sets. And the notation H indicates several heuristic functions that systems can apply to quantitating the similarities between two users.

In this paper, we want to utilize three difference heuristic functions, mentioned in [5], to compare two random sets. Derived from the previous equation, we formulate these heuristic ways, as shown in the following equations

$$H[A, B, C] = \frac{|C|}{\max(|A|, |B|)} = \max_{i=1}^{|C|} (W_{c_i}) = \frac{\sum_{i=1}^{|C|} W_{c_i}}{|C|} \tag{4}$$

where C is $A \cap B$. While the first one in Equ.4 simply expresses the minimum ratio of the common categories, the others use the weight value of each category. Second one in Equ.4 simply chooses the category of which weight value is maximal, and the last computes the average weight value of common categories.

3.2 System Architecture of BlogGrid

Along with the estimated information channel, the relevant pieces of information should be pushed actively. Information pushing service proposed in this paper is remote and asynchronous because this is based on web environment and information about a participant’s interests extracted from his own behaviors. The whole system architecture consists of two main parts, which are a facilitator located between the users and the client-side blogspace browser that communicates with the facilitator. We embed autonomous and proactive agent module into this system. Every communication between agents is conducted, regardless of user’s interventions. Also, while browsing blogspaces to search information, users can be “implicitly” recommended from the facilitator in the following two ways:

- By querying specific information for the facilitator. After the information about a particular topic is requested, the facilitator can determine who has the maximum weight value of that topic by scanning his yellow pages.
- By broadcasting new information of like-minded bloggers from the facilitator. Every time a user responds a new comment, this fact, after normalization, is sent to the facilitator. Users thereby can obtain information related to the common concepts in their own preferences from neighbors.

Each blogger needs personal agent module. This agent initializes and manages the preference of the corresponding blogger based on blogspace repository. Through personal agents’ reporting responding activities of bloggers, the facilitator agent can automatically generate queries and recommendations.

4 Experimental Results

For evaluating our system, we have conducted preliminary experimentation. First of all, ten people were invited to participate to navigate BlogGrid environment. Before starting experiment, they claimed their own preferences, by referring to “Entertainment” directory in Yahoo directory service (<http://dir.yahoo.com/>). We asked them to post information relevant to their preferences.

After testing bed was prepared, these ten bloggers were linked with at most two acquaintances for constructing social networks, and they conducted two-week BlogGrid experimentation. During first one week, bloggers were allowed to browse other blogspaces and respond posts by leaving comments and trackbacks without BlogGrid’s recommendation. Next one week, BlogGrid supported them to easily exchange information.

The issue for evaluation is how accurately BlogGrid recommended information to each blogger. We measured the ratio of the number of posts by BlogGrid’s recommendation to the total number of posts. With human evaluation by interviewing with bloggers,

- Six bloggers, User₂ (35.3%), User₃ (31.7%), User₄ (28.1%), User₆ (37.5%), User₈ (36.7%), and User₁₀ (23.9%) were fully satisfied
- Two bloggers, User₁ (26.2%) and User₉ (15.6%) were partially satisfied
- One blogger, User₅ (28.1%) were partially unsatisfied
- One blogger, User₇ (7.5%) were fully unsatisfied

with BlogGrid’s recommendations. Roughly, 80% of bloggers rated BlogGrid as a useful system to them. User₆ has shown the largest ratio level, which means he was been most effectively helped by BlogGrid. Particularly, User₅ judged BlogGrid to be negative. We found out his preference is very diverse, so that his ratio showed relatively high.

5 Concluding Remarks and Future Work

As the concerns about blogspace is ever-increasing, efficient information delivering function is desperately necessary. We were motivated to enhance this primitive blogspace with cooperative computing paradigm.

Therefore, this paper proposes that BlogGrid system is capable of helping bloggers, by analyzing and recognizing user activity and interests. Bloggers can get information pushing service with BlogGrid’s recommendation, which consists of mainly two steps; i) organization of transparent communities and ii) information pushing by facilitator.

We developed BlogGrid system by using Blojsom libraries and Borland Delphi software. BlogGrid browser’s graphic user interface can visualize the communications and relationships between bloggers. More importantly, facilitator of BlogGrid is capable of

- 1) Investigating the factors which influence relationships.
- 2) Drawing out implications of the relational data.
- 3) Making recommendations to improve communications.

Through simple experimentation, we have shown that a relatively large part of bloggers were satisfied with this service.

References

1. Athanasiadis, I.N., Mitkas, P.A.: Social influence and water conservation: an agent-based approach. *Computing in Science & Engineering* **7**(1) (2005) 65-70
2. Blojsom: <http://wiki.blojsom.com/wiki/display/blojsom/About+blojsom>.
3. Cayzer, S.: Semantic blogging and distributed knowledge management. *Communications of ACM* **47**(12) (2004) 47-52
4. Dekker, A.H.: A Category-Theoretic Approach to Social Network Analysis. *Electronic Notes in Theoretical Computer Science* **61** (2002) 21-33
5. Jung, J.J., Lee, K.-S., Park, S.-B., Jo, G.-S.: Efficient Web Browsing with Semantic Annotation: a Case Study of Product Images in E-Commerce Sites. *IEICE Transactions on Information and Systems* **E88-D**(5) (2005) 843-850
6. Kleinberg, J.: Authoritative sources in a hyperlinked environment. *Journal of the ACM* **46**(5) (1999) 604-632
7. Nardi, B.A., Schiano, D.J., Gumbrecht, M., Swartz, L.: Why We Blog. *Communications of ACM* **47**(12) (2004) 41-46
8. Rosenbloom, A.: The Blogosphere. *Communications of ACM* **47**(12) (2004) 31-33
9. Zhuge, H.: Semantics, Resource and Grid. *Future Generation Computer Systems* **20**(1) (2004) 1-5