

# Social grid platform for collaborative online learning on blogosphere: A case study of *eLearning@BlogGrid*<sup>☆</sup>

Jason J. Jung

*Department of Computer Engineering, Yeungnam University, Dae-Dong, Gyeongsan, Republic of Korea*

## Abstract

The aim of this study is to distribute relevant contents into users (or personal information spaces) by organizing context-based communities. In this paper, we propose a novel grid platform to socialize blogs by comparing the contexts extracted from online learning blogosphere. We are especially focusing on how to apply *context matching* process to collaborative learning process on the blogosphere. Thereby, personal blogs are contextualized, i.e., all possible contexts on personal blogs are formalized by collecting social behaviors of bloggers. Such behaviors are (i) posting articles (or images), (ii) tagging the posted resources, (iii) linking to neighbor blogs, and (iv) interacting with the neighbor blogs. These blogger models are capable of being compared with others to quantify contextual similarities  $Sim_c$  between the corresponding blogs by using co-occurrence analysis method. Especially, as maximizing the contextual similarity between two blogs, the corresponding sets of tags are contextually aligned with each other. In this study, for implementing this platform, we deploy *BlogGrid* platform to support information pushing service to students. Through our experimental results, we found out that average weighting measurement scheme with co-occurrence patterns from responding (e.g., comments and trackback) activities is the most significant patterns for information pushing on collaborative learning.

© 2007 Elsevier Ltd. All rights reserved.

**Keywords:** Social grid; Online learning; Content distribution; Blogs

## 1. Introduction

As many information systems are interconnected with each other, it is difficult to efficiently support collaborations between people who are not known each other. There have been various forms of services for supporting collaborations between them. Especially, content delivering services are the most important but difficult facility in the distributed information systems. So far, this kind of delivering services has been simply based on a yellow page, which is an explicit member list registering their individual information (e.g., mailing lists). The list is able to be regarded as a community to get to know each other, and

to interact with others by communicating with them. In other words, it is playing a role of a *social hub*. Therefore, when people are given a certain task, they need to (i) search for the social hubs to gather as much information as possible, (ii) aggregate some pieces of information relevant to the task, and (iii) refine it to carry out the task. For example, if a user is assigned to survey about “Mouse”, we can say that he might be interested in obtaining ‘Animal’. If any contents (or persons) which are related to the topic are newly detected, he should be notified about the information including the contents as well as the content providers.

However, a problem in this service is that the social hubs are impossible to carry out *personalized content delivery*. It means that most of the contents received from these services may be irrelevant to their own tasks, even though they believe that the social hubs are appropriate. Here, we realize that it is because the social hub is not taking into account which contextual information is in the contents

<sup>☆</sup> This paper is significantly revised from an earlier version presented at the International Workshop on Advances in Web-based Learning affiliated by International Conference on Web Information Systems Engineering in October 2006.

E-mail address: [j2jung@intelligent.pe.kr](mailto:j2jung@intelligent.pe.kr)

and the tasks. This problem is similar in context-based information retrieval, regarding that the social hub (or mediator) has to be capable of comparing the contexts between contents and user tasks.

More seriously, another problem is *semantic heterogeneity* between people, e.g., personal information systems. It makes them more difficult to automatically collaborate, i.e., their interoperability becomes decreased. After a potential relationship between two users is discovered by the social hub, both semantic knowledge structures should be efficiently aligned. Contrary to local information systems, it is not guaranteed that every information system in the open network has identical data representation strategies (e.g., XML schema, database schema, ontologies, and so on). It means that most of their contents might be differently encoded by not only their own database schema but also their own contexts. For example, in case of semantic heterogeneity, when some people can put a query ‘Mouse’, which is a part of ‘Computer Hardware’, the information systems would return the content about a kind of ‘Animal’.

Basically, it is difficult to measure the relevance between both contexts of a content and a user, i.e., determine whether a certain content is contextually relevant and useful to some of people. Thus, in this paper, we want to identify a group of users who are contextually like-minded at a moment, instead of analyzing a content to compare it to all users. We refer to the set of linkages as a *social grid*. Formation of the social grid is dynamically changed, as the contexts of user tasks are dynamic over time. Consequently, contents newly generated by a user can be disseminated to others in a same group through the social grid.

Particularly, in this paper, we have been focusing on collaborative online learning systems on blogosphere. They have been concerning about information (or knowledge) sharing between students (Downes, 2005; Hiltz & Turoff, 2002). Most of these systems have to structurize the participating students’ contexts from their activities (e.g., social interactions) during collaborative learning tasks (Keith Edwards, 2005). Once having captured the students’ contexts, we can figure out whose contexts are most similar and interrelated to the corresponding student’s through the social grid.

In summary, we note two main problems on context-based content dissemination in online learning systems:

- context mismatching between students and between learning contents and students, and
- semantic heterogeneities between students for sharing the learning contents.

In order to solve these problems, we regard user activities as implicit evidence representing user context, and especially, we apply a weblogging (blogging) systems to keep track of these activities. Recently, Web 2.0 applications (e.g., blogs and wikis) have been getting more popular, because they are simplifying publication processes of

contents on the web (Blood, 2004). One can be provided with efficient platforms to collaboratively share information with each other. A blog (originally Weblog) is a web application presented as a web page consisting of periodic posts, normally in reverse chronological order (Rosenbloom, 2004).

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 (Blood, 2004). 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. Hence, in terms of online learning, we note two important features of blogs, as follows:

- *Personal content management*: Students can create and manage various types of contents including personal information (e.g., personal history, commentaries, photos and hyperlinks of their classmates) as well as learning materials from (e.g., presentation files, examples, and webpages).
- *Information propagation by social activities*: More importantly, along with the social links (i.e., hyperlinks), they can have access to the other classmate’s blogs. Contrast to simple browsing of a bunch of hypertext documents, they can promptly take some actions within those blogs (Higgins, Reeves, & Byrd, 2004). A representative example of such activities is leaving comments against certain resources (e.g., questioning and replying) (Xu, Kreijns, & Hu, 2006).

There have been several technology-enhances online learning systems based on Web 2.0 applications (e.g., blogosphere and wikis) (Efimova & Fiedler, 2004; Ras, Avram, Waterson, & Weibelzahl, 2005) and ubiquitous devices. However, most of current online learning solutions are not sufficiently aware of the context of the learner, that is the individual’s characteristics and the organizational context (Schmidt & Winterhalter, 2004). In order to model user context, this study formulates the user activities on blogosphere, and represent an activity as a set of features. Then, as aggregating the sets of features extracted from a sequence of activities, the user context is efficiently adapted over time. More importantly, in this paper, we propose a novel user clustering method to organize communities as a set of like-minded students during collaborative learning on BlogGrid platform.

The outline of this paper is as follows. Section 2 explains how to model students with their structural positions and activities on online learning blogosphere. In Section 3, we will describe system architecture and information pushing service of BlogGrid for collaboration between students. Section 4 will show simple example and its experimental results. In Section 5, we want to discuss our system with

some previous work. Finally, Section 6 will draw a conclusion and explain our future plan to improve our system.

## 2. Modeling students on collaborative learning

We want to model students on blogosphere in two aspects; structural patterns and behavioral patterns. As a result, in this paper, we assume that the actions taken by students are indicating explicit and implicit ways of their own context representation. Thus, we have to be able to extract useful information from both kinds of patterns about their context. In order to model behaviors on blogosphere, we need to note thoroughly the roles of blogosphere and students. Blogosphere are divided into *personal* blogs and *community* blogs (Nardi, Schiano, Gumbrecht, & Swartz, 2004). In the context of online learning domain, we can say that the community blogs are exactly replaced with *class* blogs.

### 2.1. Structural patterns

Firstly, we have to take the structural patterns into account. An online learning blogosphere is organized as an *extended* bipartite graph between two disjoint sets. Such sets are represented as both of a student layer  $\mathcal{L}_S$  and a class layer  $\mathcal{L}_C$ . Different from original bipartite graphs, the proposed blogosphere allows the vertices to be connected with each other in a layer, as shown in Fig. 1.

**Definition 1 (Student layer).** A student layer  $\mathcal{L}_S$  is defined as

$$\mathcal{L}_S = \langle \mathcal{S}, \mathcal{N}_S \rangle \quad (1)$$

where  $\mathcal{S} = \{s_1, \dots, s_{|\mathcal{S}|}\}$  is a set of student blogs, and  $\mathcal{N}_S \subseteq \mathcal{S} \times \mathcal{S}$  is a set of links between the students.

**Definition 2 (Class layer).** A class layer  $\mathcal{L}_C$  is defined as

$$\mathcal{L}_C = \langle \mathcal{C}, \mathcal{N}_C^{\leftarrow}, \mathcal{N}_{S \times C} \rangle \quad (2)$$

where  $\mathcal{C} = \{c_1, \dots, c_{|\mathcal{C}|}\}$  means a set of classes in which the students can take part. Again,  $\mathcal{N}_{S \times C} \subseteq \mathcal{S} \times \mathcal{C}$  and

$\mathcal{N}_C^{\leftarrow} \subseteq \mathcal{C} \times \mathcal{C}$  are two link sets between classes and participating students, and between classes, respectively.

Here, in this blogosphere, we have to focus on structural patterns which can be extracted from three different networks. (In fact, the networks are represented as adjacency matrices. We will explain them later.)

1. A social network between students  $\mathcal{N}_S \subseteq \mathcal{S} \times \mathcal{S}$  is explicitly built by the corresponding students' acquaintanceships. Basically, social network analysis (Wasserman & Faust, 1994) is applicable for various measures on the networks between students (note that these measures apply only if the network is connected):<sup>1</sup>

[*Closeness*:] The inverse of average length of the shortest path between a student  $s$  and any other student in the network:

$$\text{Closeness}^i(s) = \frac{|\mathcal{S}| - 1}{\sum_{s' \in \mathcal{S}} \text{spd}^i(s, s')} \quad (3)$$

[*Betweenness*:] (Freeman, 1979) The proportion of shortest paths between two students which contain a particular node (this measures the power of this student):

$$\begin{aligned} \text{Betweenness}^i(s) &= \sum_{s', s'' \in \mathcal{S}} \frac{|\{p \in \text{sp}^i(s', s), p' \in \text{sp}^i(s, s'') \mid p \cdot p' \in \text{sp}^i(s', s'')\}|}{|\text{sp}^i(s', s'')|} \end{aligned} \quad (4)$$

[*Hub and authority*:] There are different but interrelated patterns of power: Authorities that are referred to by many and hubs that refers to many. The highest authorities are those which are referred to by the highest hubs and the highest hubs that those which refers to the highest authorities. Kleinberg (1999) proposes an iterative algorithm to measure authority and hub degree of each student in interlinked environment. Given initial authority and hub degrees of 1, the degrees are iteratively computed by

$$\begin{aligned} \text{Hub}_{t+1}^i(s) &= \sum_{s': (s, s') \in \mathcal{S}^i} \text{Auth}_t^i(s') \quad \text{and} \quad \text{Auth}_{t+1}^i(s) \\ &= \sum_{s': (s', s) \in \mathcal{S}^i} \text{Hub}_t^i(s') \end{aligned} \quad (5)$$

Similarly to betweenness, the hub weight indicates the structural position of the corresponding user. It is a measure of the influence that students have over the spread of information through the network.

2. A class relations  $\mathcal{N}_C^{\leftarrow} \subseteq \mathcal{C} \times \mathcal{C}$  means relations between classes provided from each institutional curricula. Thus, we have derived some important relations from DCQ

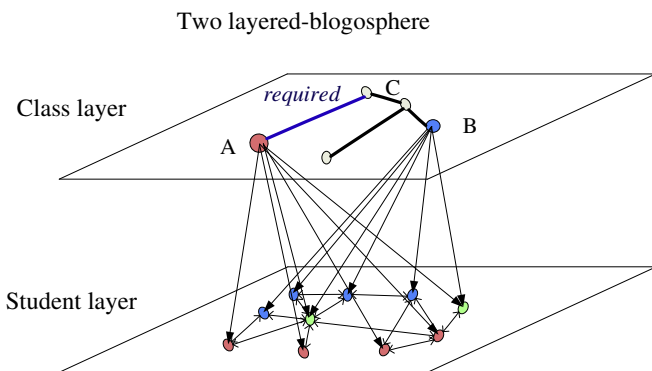


Fig. 1. Collaborative online learning environment by two-layered blogosphere; interlinked class and student layers.

<sup>1</sup> These measures are often normalized (between 0 and 1) but we present their simplest form.

(Dublin Core Quantifiers)<sup>2</sup> and IEEE LOM (Learning Object Metadata standard),<sup>3</sup> as follows. Additionally, “−1” denotes inverse relations.

[*hasPart*] (= isPartOf<sup>−1</sup>) It means the structure of a course, i.e., a set of subjects and topics. Furthermore, we can employ “equivalent” or “disjoint”.

[*requires*] (= isRequiredBy<sup>−1</sup>) It can express the prerequisite constraint. If the content from a certain course cannot be understood without knowledge of another course. The course receives an “requires” entry, while the courses with the background information receives a “isRequiredBy”.

For example, in Fig. 1, the students who want to attend class A is obliged to complete class C successfully, in advance.

3. A participant network  $\mathcal{N}_{\mathcal{S} \times \mathcal{C}} \subseteq \mathcal{S} \times \mathcal{C}$  is a set of edges between both of the layers, i.e., which student is joining which classes. Mainly, we can discover the following knowledge from this network.

[*Consensus*] With regards to the number of attending students, the consensus (or popularity) of classes can be quantified as

$$\text{Consensus}(c_i) = \frac{|\{s_k \mid \langle s_k, c_i \rangle \in \mathcal{N}_{\mathcal{S} \times \mathcal{C}}\}|}{|\mathcal{S}|} \quad (6)$$

where  $|\mathcal{S}|$  is total number of students.

[*Participation similarity*] Given two students, we can investigate co-occurrence between classes that they are participating together. The number of classes they are attending in common can be used for measuring the similarity  $\text{Sim}_p$  between their interest during collaborative learning. It is formulated by

$$\text{Sim}_p(s_i, s_j) = \frac{|\{c_k \mid c_k \in c(s_i) \cap c(s_j)\}|}{\max[|c(s_i)|, |c(s_j)|]} \quad (7)$$

where  $c(s_k) = \{c_k \mid \langle s_i, c_k \rangle \in \mathcal{N}_{\mathcal{S} \times \mathcal{C}}\}$  is a set of classes attended by student  $s_k$ .

It is noted that by analyzing the topological features on blogosphere, we can extract the structural patterns (i.e., similarity and centrality) of the students and expect them to efficient support the collaboration between student for fulfilling their learning processes.

## 2.2. Behavioral patterns

Another evidence that we have to take into account is the students’ activities on blogosphere. Basically, we want to emphasize that the patterns are dynamically changeable over time. In order to model the behavioral patterns of a

students  $B(s_i)$ , we note the activities that students can conduct on blogosphere, as follows:

- *Enlarging a social network*  $\mathcal{E}$ : It is to enlarge the established social network, as making connections with other people such as families, friends, and colleagues. In the context of collaborative learning, the neighbors might be classmates in the student layer  $\mathcal{S}$ . More importantly, in order to share information about a particular topic, students can link to classmates who are relatively closer with each other than others. Easily, a set of neighbor students of  $s_i$  is given by

$$\mathcal{E}_i = \{s_x \mid \text{ADJ}_{(ix)} = 1\} \quad (8)$$

where symmetric adjacency matrix ADJ (i.e.,  $\mathcal{N}_{\mathcal{S}}$ ) of which size is  $|\mathcal{S}| \times |\mathcal{S}|$ . If user  $s_i$  has a link to  $s_j$ ,  $\text{ADJ}_{(ij)} = 1$ . Otherwise,  $\text{ADJ}_{(ij)} = 0$ . For simplicity, the attached weight  $W_i(u_x \in \mathcal{L}_i)$  is given by either 1 or 0 from ADJ, instead of strength of social tie between nodes.

- *Posting articles*  $\mathcal{P}$ : It is the most basic action on blogosphere. As taking this action, students can input various types of information and enrich blogosphere. This action explicitly represents the corresponding blogger’s preferences. A set of articles posted a student  $s_i$  is given by

$$\mathcal{P}_i = \{p_1, \dots, p_N\} \quad (9)$$

and the corresponding weight  $W_i(p_x)$  is given by

$$W_i(p_x) = \frac{|\{p_\beta \mid \text{Category}_i(p_\beta) = \text{Category}_i(p_x)\}|}{|\text{Categories}_i|} \quad (10)$$

where function  $\text{Category}_i(p_\beta)$  is to return the corresponding category<sup>4</sup> labeling  $p_\beta$ . The weight means how much a user is interested in a certain topic (i.e., category). For example, he is probably interested in “music”, because he has been posting articles about recital schedules and musical instruments. In this paper, we choose the user-defined category information, because content-based analysis methods (e.g., keyword-based feature extractions) are rather efficient but very expensive, especially, within the information overwhelming spaces like blogosphere.

- *Navigating*  $\mathcal{N}$ : In order to get relevant information within a blog system, people can visit other blogosphere. Students can navigate the other blogosphere by the following two methods:

1. *Random browsing*: Students can randomly jump into other blogosphere. In fact, this has been the only way to deal with nepotism problem.
2. *Accessing to neighbors on social network*: By referring to the list of neighbors, students can easily move into their blogs.

<sup>2</sup> Dublin Core Metadata Initiative. Available online from: <http://dublincore.org/>.

<sup>3</sup> IEEE Learning Technology Standards Committee, Standard for Learning Object Metadata (LOM) IEEE 1484.12.1-2002. Available online from: <http://www.ieeeltsc.org/standards/1484-12-3-2005/>.

<sup>4</sup> It is also called as a “tag” in web 2.0 applications.

These activities by a student  $s_i$  can be simply accumulated as

$$\mathcal{N}_i = \{s_\alpha, s_\beta, \dots\} \quad (11)$$

where  $s_x$  means the personal blogs of the corresponding student. For frequency analysis of access patterns, we employ sessionization method proposed in Jung (2005). Thus, the attached weight  $W_i(s_x \in \mathcal{N}_i)$  is given by

$$W_i(s_x \in \mathcal{N}_i) = \text{occur}(s_x) \times \frac{1}{|\text{Session}(\mathcal{N}_i, T)|} \quad (12)$$

where  $T$  is a time span (e.g., a hour, a day, and a week), and function `Session` returns a set of sessions  $\{\text{session}_t \mid \text{session}_t = \langle s_x, s_\beta, \dots \rangle\}$  by segmenting  $\mathcal{N}_i$  with time span  $T$ . Function `occur` is able to check if a given blog address  $s_x$  is located in a session or not, and count all the sessions occurred by  $s_x$ .

We do not need to take care of navigation on class blogs.

- *Responding  $\mathcal{R}$ .* Students can respond to a certain article while navigating blogosphere. Their responses like personal opinions, sympathies, antipathies, or even apathy are expressed as the following two ways:
  1. *Comment:* A student can leave his responses which consist of short text sentences. More than a comment can be serially and continuously attached to each post article.
  2. *Trackback:* In contrast to comments, it allows a student 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,<sup>5</sup> most blogging systems have adopted trackback mechanism as generic function.

These activities by student  $s_i$  can be simply accumulated as

$$\mathcal{R}_i = \{p_{\alpha,a}, p_{\beta,b}, \dots\} \quad (13)$$

where  $p_{\alpha,a}$  means  $a$ th post in the personal blogs  $s_x$ . We can easily realize that  $\mathcal{R}_i$  is a subset of  $\mathcal{N}_i$ . The weight of each responding is given by

$$W_i(p_{\alpha,a}) = \frac{|\text{colocated}(p_{\alpha,a})|}{\max(|\text{colocated}(p_m)|_{p_m \in \mathcal{R}_i})} \quad (14)$$

where `colocated` is a function for counting comments and trackbacks colocated in the same responding.

Moreover, 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 students’ interests and opinions. Another interesting fea-

ture is that the responding can be nested. It means that students can respond to a certain comment already attached to articles.

Overall, given a set of a student’s behaviors on blogosphere, his behavioral model is represented as

$$B(s_i) = \langle \mathcal{E}, \mathcal{P}, \mathcal{N}, \mathcal{R} \rangle \quad (15)$$

where each element is assumed to be mutually exclusive with each other. More importantly, this model can change over time, as the student activities can continue to be detected.

### 3. Information pushing on eLearning@BlogGrid

Here, we want to compare the two types of user models, in order to organize communities. Each community consists of only like-minded students, e.g., collaborative networks (Newman, 2004). This community organization plays an important role of efficiently providing information to the students participating our system. Thereby, we apply co-occurrence analysis methods to all possible pairs of user models in a blogosphere for measuring the context similarities  $\text{Sim}_\mathcal{C}$  between students. In this paper, we extend a generic BlogGrid architecture (Jung, Ha, & Jo, 2005) to eLearning@BlogGrid for handling collaboration problem between students by using the similarity measurement.

#### 3.1. Community organization-based on similarity measurement

So far, we have introduced several similarities meaning possible chances to compare the students. Simple scheme for measuring the contextual similarity between two students  $s_i$  and  $s_j$  is given by

$$\text{Sim}_\mathcal{C}(s_i, s_j) = \max_{C \in \{\mathcal{E}, \mathcal{P}, \mathcal{N}, \mathcal{R}\}} H[C_{s_i}, C_{s_j}, \tilde{o}(C_{s_i}, C_{s_j})] \quad (16)$$

$$= \text{Sim}_p(s_i, s_j) \quad (17)$$

where the function  $\tilde{o}$  is to obtain the  $K$  common elements from both sets. The notation  $H$  indicates several heuristic functions that systems can apply to quantify the similarities between two users. In this paper, we want to utilize three difference heuristic functions, mentioned in Jung (2007), to compare two random sets. Derived from the previous equation, we formulate these heuristic ways, as shown in the following equations:

$$H[A, B, K] = \frac{|K|}{\max(|A|, |B|)} \quad (18)$$

$$= \max_{i=1}^{|K|} W(k_i) \quad (19)$$

$$= \frac{\sum_{i=1}^{|K|} W(k_i)}{|K|} \quad (20)$$

where  $K$  is  $\tilde{o}(A, B)$ , and  $k_i \in K$ . While Eq. 18 simply expresses the minimum ratio of the common categories, the others use the weight value of each category. Eq. 19 simply chooses the category of which weight value is maximal, and Eq. 20 computes the average weight value of common

<sup>5</sup> Movable Type. <http://www.sixapart.com/movabletype/>.

categories. Thus, we want to empirically compare the proposed contextual similarity  $\text{Sim}_{\mathcal{C}}$  to the participation similarity  $\text{Sim}_p$ .

Based on the similarities between two arbitrary students, we can apply a non-parametric approach, e.g., nearest neighborhood method (Chidananda Gowda & Krishna, 1978). However, we have to consider physical social links ADJ built by  $\mathcal{E}$  in the student layer. As extending (Newman, 2004), this task is to maximize “contextual” modularity function  $Q^\diamond$  on social network. Given the number of communities  $k$  from a set of students, a social network  $\mathcal{S}$  can be partitioned into a set of communities (or subgroup)  $\mathcal{G} = \{g_1, \dots, g_k\}$ . The users can be involved in more than one community. It means that a student in  $g_i$  can also be taken as one of members of  $g_j$ . The modularity function  $Q^\diamond$  is formulated by

$$Q^\diamond(\mathcal{S}) = \sum_{i=1}^k \frac{\sum_{s_a \in g_i, s_b \in g_i, \text{ADJ}(ab)=1} \text{Sim}(s_a, s_b)}{|g_i|} \quad (21)$$

where all possible pairs of bloggers should be considered. Thus,  $\mathcal{G}(\mathcal{S})$  is discovered when  $Q^\diamond(\mathcal{S})$  is maximized. For computing this, in this paper, we applied  $k$ -nearest neighborhood methods.

Fig. 2 shows a simple example. The links between two nodes are weighted by the similarity between the corresponding students’ behaviors. Firstly, we can find the  $k$  nodes whose similarity summation is highest (here,  $k = 2$  and the highest nodes are  $c = 3.2$  and  $f = 1.9$ ). From these centers of communities, the rest of members are searched incrementally, until all nodes get involved at least one community.

### 3.2. Proactive information pushing service

Once we found out the communities with respect to the contexts, we want to refer to the boundary between the communities as a threshold to keep contextually irrelevant information from being flown from other communities. In other words, the students in a same community can automatically share their learning objects with each other. We call the linkages (i.e.,  $\mathcal{N}_{\mathcal{S}}$ ) between students in a same community as information channels flowing into the students’ blogs.

Along with the estimated information channel, the relevant pieces of information should be pushed actively.

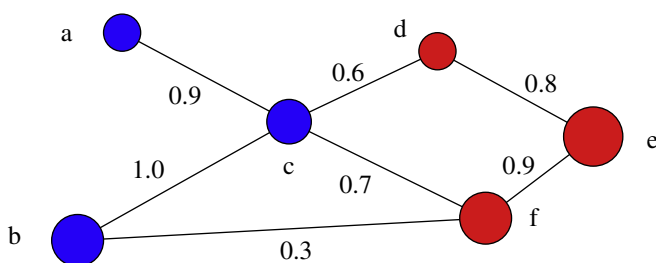


Fig. 2. Example of community organization.

Information pushing service proposed in this paper is remote and asynchronous because this is based on web environment and information about a student participant’s interests extracted from his own behaviors.

We have embedded autonomous and proactive agent module into this system. Every communication between agents is conducted, regardless of students’ interventions. Moreover, while browsing the blogosphere to search for 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 student bloggers from the facilitator:* Every time a student responds a new post or comment, this fact, after normalization, is sent to the facilitator. The students within a same community, thereby, can obtain information related to the common concepts in their own preferences from neighbors.

### 3.3. System architecture of eLearning@BlogGrid

As shown in Fig. 3, the whole system architecture of eLearning@BlogGrid consists of two main parts; (i) in the middle of blogosphere, BlogGrid server containing a facilitator (or mediator) with data repository and (ii) client-side blogging browser which is capable of communicating with the facilitator.

Through personal agents’ reporting responding activities of student bloggers, the facilitator agent can automatically generate queries (e.g., SQL queries) by using the similarity between the users in the same community (Eq. 21), and recommendations for providing the corresponding students.

Each student blogger needs personal agent module. This agent initializes and manages the corresponding blogger’s preference. Thereby, it has to monitor the actions, and inform them to the facilitator for storing blogosphere repository. Particularly, graphic user interface of the blogging browser is shown in left-bottom on Fig. 3.

## 4. Experimental results

In this section, we want to explain two main experimental issues in this paper. Firstly, we evaluated our community organization mechanism from a set of students participating in our experiments, with respect to three different heuristic functions (Eqs. (18)–(20)) and four different behaviors (Eq. (15)). In addition, they were compared to simple community organization-based on structural pattern (i.e.,  $\text{Sim}_p$ ). As second experiment, we verified the performance of the recommendation generated for collaborative learning among students, by interviewing the students for their experiences.

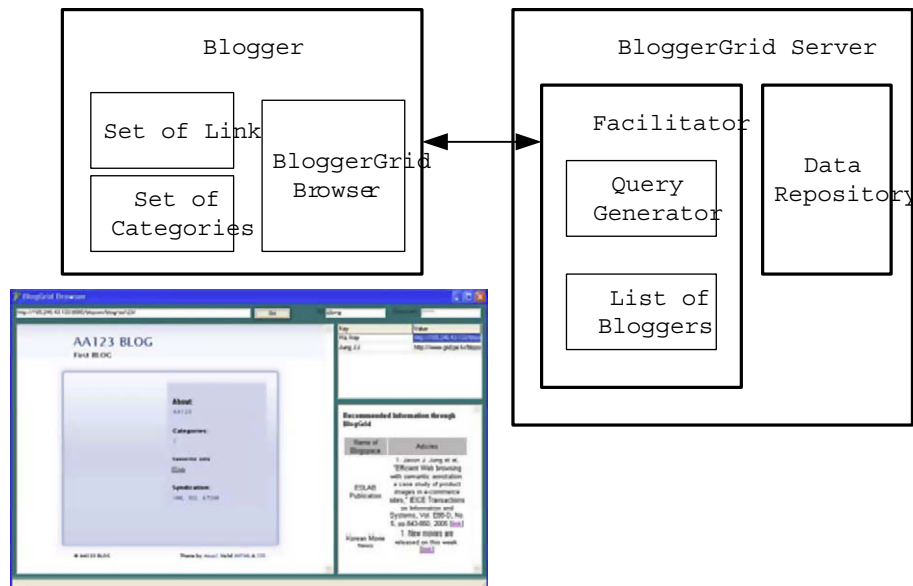


Fig. 3. System architecture and graphic user interface of BlogGrid.

In order to conduct experiments on the proposed system, we have built a testing environment.<sup>6</sup> We invited 19 graduated students attending five courses (i.e., Semantic Web, Data Mining, Database, Neural Network, and Web Services) in Inha university, Korea. Initially, they had to build their own blogs by using Blojsom platform,<sup>7</sup> and for posting articles, collect some research papers which they are interested in from the selected conference proceedings by looking at four fields, e.g., title, authors, keywords, and abstract. Such proceedings (22 conferences) are shown, as follows:

- European Semantic Web Conference (ESWC) 2004, 2005, 2006
- International Semantic Web Conference (ISWC) 2004, 2005, 2006
- IEEE International Conference on Data Mining (ICDM) 2004, 2005, 2006
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) 2004, 2005
- ACM SIGMOD International Conference on Management of Data (SIGMOD) 2004, 2005, 2006
- International Conference on Very Large Data Bases (VLDB) 2004, 2005, 2006, 2007
- IEEE International Joint Conference on Neural Networks (IJCNN) 2004, 2005
- IEEE International Conference on Web Services (ICWS) 2004, 2005

Additionally, in order to build a social network  $\mathcal{S}$ , they were asked to declare their own friends (max. five) who are

socially closer than others. Finally, we set the student's preference model by using questionnaires about the five courses. These datasets are available here.<sup>8</sup>

#### 4.1. Community organization

Each participant has noticed a set of students estimated in a same community, with respect to the heuristic function and blog activities of students. We computed the matching ratio with the communities organized by preference similarities from questionnaires (vector-space model). Table 1 shows the experimental results on community organization.

Eventually, we found out that heuristic function used in Eq. 20 outperforms the other two functions. Moreover, rather than any other activities, responding activities should be emphasized to measure the similarity among people on blogosphere.

#### 4.2. Recommendation generation

We have evaluated a set of BlogGrid's recommendation provided to the participants organized by three different heuristic functions and blogosphere activities. In order to measure the precision of recommendation, we computed the mean absolute error (MAE) given by

$$\text{MAE} = \frac{\text{Number of false recommendation}}{\text{Total number of recommendation}} \quad (22)$$

where false recommendation is discriminated by comparing with the pre-defined user preferences (e.g., a set of categories).

<sup>6</sup> eLearning@BlogGrid. <http://165.246.43.133:8080/blojsom/blog/default/>.

<sup>7</sup> Blojsom. <http://wiki.blojsom.com/wiki/display/blojsom/About+blojsom>.

<sup>8</sup> BlogGrid dataset. <http://www.intelligent.pe.kr/Blog+Wiki@Inha/BlogGrid>.

Table 1  
Experimental results on community organization (%)

Heuristic function	Linking $\mathcal{L}$	Posting $\mathcal{P}$	Navigation $\mathcal{N}$	Responding $\mathcal{R}$	Average
Eq. (18)	–	38.45	49.52	57.95	48.64
Eq. (19)	–	45.25	46.58	59.68	50.51
Eq. (20)	–	63.50	62.96	<b>64.93</b>	63.80
Average	–	49.07	53.02	60.85	

As presented in Table 2, the community organization by responding activity and heuristic function by Eq. (18) was the best formation among students for minimizing the recommendation error (about 21.34%).

Moreover, we have compared the performance of the information pushing service based on (i) the proposed contextual community organization with (ii) simple topological community organization during a month. As shown in Figs. 4 and 5, two measurements (i.e., recall and precision) were measured.

Contextual community organization-based service has shown better recall results (particulatly, in 7th day, 79% higher). It means that the proposed service can support the users, particularly in the early stage. With respect to precision, we found out that the information pushing system finally have shown the converged result over 80% precision level, because the service needed to collect user activities.

Table 2  
Experimental results on MAE of recommendation by BlogGrid (%)

Heuristic function	Linking $\mathcal{L}$	Posting $\mathcal{P}$	Navigation $\mathcal{N}$	Responding $\mathcal{R}$	Average
Eq. (18)	29.76	38.88	51.29	21.34	35.32
Eq. (19)	25.13	34.34	39.96	23.32	30.69
Eq. (20)	28.46	37.78	45.00	22.73	33.49
Average	27.78	37.00	45.42	22.46	

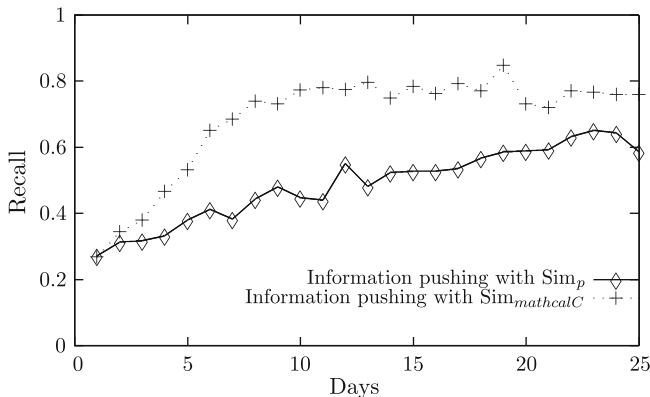


Fig. 4. Performance of information searching from three groups with recall.

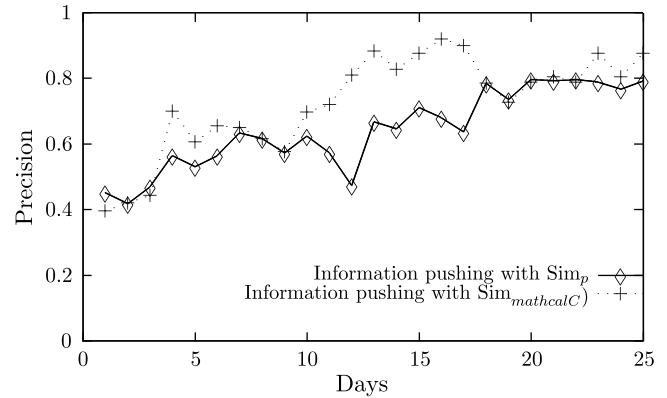


Fig. 5. Performance of information searching from three groups with precision.

### 5. Discussion and related work

We want to discuss some significant information uncovered in this study. Firstly, we found out that average weighting measurement scheme from co-occurrence patterns with the responding (e.g., comments and trackbacks) activities are the most significant patterns for information pushing on collaborative learning. We believe that students have taken the responding activities to a certain articles with “very high and reliable” interests. As a result, these activities is one of the most applicable implicit behaviors to measure the similarity between users on student social network (Reffay & Chanier, 2002; Saltz, Hiltz, & Turoff, 2004). They have shown 24.0% and 14.77% improvements, as compared to posting and navigating. Moreover, with respect to the heuristic functions,  $\frac{\sum_{i=1}^{|K|} W(k_i)}{|K|}$  in Eq. (20) has shown about 28.74% improvement.

Secondly, recommendation was able to be propagated on distributed personal information space. Responding activities and heuristic function  $\frac{|K|}{\max(|A|, |B|)}$  (Eq. (18)) have shown the best combination to organize communities for providing recommendations. In particular, communities organized by linking activity has shown relatively high recommendation performance.

#### 5.1. Comparing with client-server systems

Hereafter, we have to mention some improvements from traditional centralized systems for collaborative online learning. Comparing with client-server online learning systems such as HARMONY (Kojiri & Watanabe, 2001) and PLANETLAB (Chen, Hu, & Yu, 2006), major difference is dynamic socialization. Blogosphere can provide people a way to express their interests explicitly and implicitly, as configuring their own social networks. Henceforth, we consider a question “Does socialized information space improve the performance of collaborative learning among students?” The students in our system can be promptly switched into the most relevant communities in real-time. Expectedly, dynamic text sources such as RSS (Really Sim-



ple Syndication),<sup>9</sup> which is a family of web feed formats, specified in XML and used for Web syndication, can be exploited on our system. While it is used by (among other things) news websites, weblogs and podcasting, we can expect to embed this function to our system.

### 5.2. Category representation

Another discussion issue is the representation of blog category. We built five classes; SW (Semantic Web), DM (Data Mining), DB (Databases), WS (Web Service), and NN (Neural Network). The categories of classes are represented as simple lists and trees, respectively. As a result, hierarchical structure, in case of NN, has shown better performance. Bloggers were able to express their interests more in detail. Furthermore, for measuring the similarity between certain two patterns in vector-space, hierarchically represented patterns could reduce the uncertainty.

## 6. Conclusions and future work

Tradition paradigm of computer-aided instruction (CAI) has been shifted by diverse information systems. Our blogging-based online learning environment has shown to support students by providing efficient services with various forms of learning objects (e.g., multimedia content, instructional content, learning objectives, instructional software and software tools, and persons, organizations or events). In this paper, we believe that the strength of social ties between students is one of the most important criteria for collaborations on online learning systems. Therefore, we exploit the grid computing paradigm, which is capable of supporting an efficient framework of information sharing between heterogeneous sources, to blogosphere and students. Practically, we apply co-occurrence analysis methods for measuring the similarities between both activities. In addition, we empirically evaluated three different heuristic functions.

As our future plans, we have a plan to apply our system to a large set of students for scalability testing of our system. Especially, we are expecting that our system can support the beginning students to guide the up-to-data context on e-class. As another important future study, similar to Tscherteu (2005), we plan to implement information visualization tools to visualize the social interactions among students, and help students to improve the understandability. Finally, we will combine blackboard architecture with our grid-based system, like wiki-based learning systems (Bruns & Humphreys, 2005; Reinhold & Abawi, 2006).

### Acknowledgement

This work was supported by the Korean Science and Engineering Foundations (KOSEF) grant funded by the

Korean government (MEST). (No. R01-2008-000-10196-0).

### References

- Blood, Rebecca (2004). How blogging software reshapes the online community. *Communications of the ACM*, 47(12), 53–55.
- Bruns, Axel, & Humphreys, Sal (2005). Wikis in teaching and assessment: The M/Cyclopedia project. In *Proceedings of the 2005 international symposium on Wikis (WikiSym '05)* (pp. 25–32). New York, NY, USA: ACM Press.
- Chen, ZhangJian, Hu, TongSen, & Yu, Yan (2006). Analysis and research about an on-line collaborative learning teams based grids. In Zhigeng Pan, Ruth Aylett, Holger Diener, Xiaogang Jin, Stefan Göbel, & Li Li (Eds.), *Proceedings of the first international conference on technologies for e-learning and digital entertainment (edutainment)*, April 16–19. *Lecture notes in computer science* (vol. 3942, pp. 735–744). Springer.
- Chidananda Gowda, K., & Krishna, G. (1978). Agglomerative clustering using the concept of mutual nearest neighbourhood. *Pattern Recognition*, 10(2), 105–112.
- Downes, Stephen (2005). E-learning 2.0. *eLearn Magazine*, 2005(10), 1.
- Efimova, Lilia, & Fiedler, Sebastian (2004). Learning webs: Learning in weblog networks. In P. Kommers, P. Isaia, & M. B. Nunes (Eds.), *Proceedings of the IADIS international conference web based communities 2004* (pp. 490–494). IADIS Press.
- Freeman, L. C. (1979). Centrality in social networks: Conceptual clarification. *Social Networks*, 1, 215–239.
- Higgins, Christopher J., Reeves, Lenita, & Byrd, Eric (2004). Interactive online journaling: A campus-wide implementation of blogging software. In *Proceedings of the 32nd annual ACM SIGUCCS conference on user services (SIGUCCS '04)* (pp. 139–142). New York, NY, USA: ACM Press.
- Hiltz, Starr Roxanne, & Turoff, Murray (2002). What makes learning networks effective? *Communications of the ACM*, 45(4), 56–59.
- Jung, Jason J. (2005). Semantic preprocessing of web request streams for web usage mining. *Journal of Universal Computer Science*, 11(8), 1383–1396.
- Jung, Jason J. (2007). Exploiting semantic annotation to supporting user browsing on the web. *Knowledge-Based Systems*, 20(4), 373–381.
- Jung, Jason J., Ha, Inay, & Jo, GeunSik (2005). BlogGrid: Towards an efficient information pushing service on blogspace. In Hai Zhuge & Geoffrey Fox (Eds.), *Proceedings of fourth international conference on grid and cooperative computing, November 30–December 3, 2005. Lecture notes in computer science* (vol. 3795, pp. 178–183). Springer.
- Keith Edwards, W. (2005). Putting computing in context: An infrastructure to support extensible context-enhanced collaborative applications. *ACM Transactions on Computer-Human Interaction*, 12(4), 446–474.
- Kleinberg, Jon M. (1999). Authoritative sources in a hyperlinked environment. *Journal of ACM*, 46(5), 604–632.
- Kojiri, T., & Watanabe, T. (2001). Harmony: Web-based adaptive collaborative learning environment. In *Proceedings of international conference on computers in education (ICCE/SchoolNet)* (pp. 559–566).
- Nardi, Bonnie A., Schiano, Diane J., Gumbrecht, Michelle, & Swartz, Luke (2004). Why we blog. *Communications of the ACM*, 47(12), 41–46.
- Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69, 066133.
- Ras, Eric, Avram, Gabriela, Waterson, Patrick E., & Weibelzahl, Stephan (2005). Using weblogs for knowledge sharing and learning in information spaces. *Journal of Universal Computer Science*, 11(3), 394–409.
- Reffay, Christophe, & Chanier, Thierry (2002). Social network analysis used for modelling collaboration in distance learning groups. In Stefano A. Cerri, Guy Gouardères, & Fábio Paraguaçu (Eds.), *Proceedings of the sixth international conference on intelligent tutoring systems (ITS 2002)*, Biarritz, France and San Sebastian, Spain, June 2–

<sup>9</sup> RSS, [http://en.wikipedia.org/wiki/RSS\\_\(protocol\)](http://en.wikipedia.org/wiki/RSS_(protocol)).

- 7, 2002. *Lecture notes in computer science* (vol. 2363, pp. 31–40). Springer.
- Reinhold, Silvan, & Abawi, Daniel F. (2006). Concepts for extending wiki systems to supplement collaborative learning. In Zhigeng Pan, Ruth Aylett, Holger Diener, Xiaogang Jin, Stefan Göbel, & Li Li (Eds.), *Proceedings of the first international conference on technologies for e-learning and digital entertainment (edutainment), April 16–19, 2002. Lecture notes in computer science* (vol. 3942, pp. 755–767). Springer.
- Rosenbloom, Andrew (2004). The blogosphere. *Communications of the ACM*, 47(12), 31–33.
- Saltz, Jeffrey S., Hiltz, Starr Roxanne, & Turoff, Murray (2004). Student social graphs: Visualizing a student's online social network. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work (CSCW'04)* (pp. 596–599). New York, NY, USA: ACM Press.
- Schmidt, Andreas, & Winterhalter, Claudia (2004). User context aware delivery of e-learning material: Approach and architecture. *Journal of Universal Computer Science*, 10(1), 28–36.
- Tscherteu, Gernot (2005). The blogosphere map visualising microcontent dissemination inspired by maria montessori. In Theo Hug, Martin Lindner, & Peter A. Bruck (Eds.), *Proceedings of the microlearning* (pp. 109–120). Innsbruck University Press.
- Wasserman, Stanley, & Faust, Katherine (1994). *Social network analysis: Methods and applications*. Cambridge(UK): Cambridge University Press.
- Xu, Wen, Kreijns, Karel, & Hu, Jun (2006). Designing social navigation for a virtual community of practice. In Zhigeng Pan, Ruth Aylett, Holger Diener, Xiaogang Jin, Stefan Göbel, & Li Li (Eds.), *Proceedings of the first international conference on technologies for e-learning and digital entertainment (edutainment 2006), April 16–19. Lecture notes in computer science* (vol. 3942, pp. 27–38). Springer.