Expert Systems with Applications 36 (2009) 10627–10633

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Knowledge distribution via shared context between blog-based knowledge management systems: A case study of collaborative tagging ☆

Jason J. Jung

Department of Computer Engineering, Yeungnam University, Dae-Dong, Gyeongsan 712-749, Republic of Korea

ARTICLE INFO

Keywords:
Personal knowledge management
Shared context
Knowledge distribution
Community of practice

ABSTRACT

Most of the existing personal knowledge management systems have been employing blogging services which is capable of providing various services (e.g., knowledge distribution and knowledge creation) to people. However, automated knowledge delivering service among the systems is difficult to take into account the corresponding *context* (or semantics) of the knowledge, so that the service can spread irrelevant information into knowledge management systems. In order to solve this problem, this study proposes a novel architecture, called *blog context overlay network*, to fulfill context matching between blog-based knowledge management systems. It is referred to as detecting 'shared' context between knowledge management systems. Thus, with respect to the contexts, we want to identify a community of practice (CoP) on a knowledge blogosphere where a set of blog-based knowledge management systems are incorporating with each other. As a result, newly generated knowledge can be proactively diffused to the blog-based knowledge management systems of which context is relevant to the knowledge, even before the bloggers' queries are explicitly asked.

1. Introduction

Personal knowledge management (PKM) can be broadly defined as an evolving set of knowledge, skills and abilities that allows an individual to survive, and prosper, through turbulent, complex and changing organisational and social environments. PKM is a concept that has grown out of a combination of fields as diverse as knowledge management (KM), personal information management, cognitive psychology, philosophy, management science, communications as well as others.

So far, very little research or significant conceptual development has been done with PKM. Main issues of the previous PKM studies have focused on helping individuals be more effective. The existing literature points to increasing individual effectiveness in work environments such as teams and organisations. While the traditional view of KM is primarily concerned with managing organizational knowledge, including the knowledge that individuals possess in their heads, through combinations of technology and management processes, the core focus of PKM is personal inquiry – the quest to find, connect, learn and explore relevant knowledge in distributed systems. It means that the people have to be able to automatically interact with each other for resolving the personal inquiry.

However, most of the KM systems have serious problems to support efficient interactions between people. To deal with the problems, these KM systems have been recently trying to employ the emerging Web 2.0 applications, e.g., blogs and wikis. Most importantly, the Web 2.0 application is a platform that maximize the usability of people. It means that people do not need to know how to write HTML code and edit multimedia contents on blogs. We need to note that such platforms are providing two important services in common; (*i*) publication, and (*ii*) sharing. In the context of blog-based KMS, particularly, we can address them, as follows:

Knowledge publication. Each person can publish their own experiences and know-hows obtained during conducting their jobs and tasks. Not only free texts but also relevant multimedia data can be added to describe the knowledge. More importantly, the published knowledge can be tagged with some keywords, the so-called *tags*. These tags can be regarded as a user-specified categorization system representing their situation and context (e.g., semantics and preferences). Sometimes, all of the tags are visualized as the so-called *tag cloud*, for better understanding.

Social communications. People can declare their social relations (e.g., friends, family, colleagues, and so on), which are referred to as blogrolls.

☆ This paper is significantly revised from an earlier version presented at the The 21st International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE) in June 2008.

E-mail address: j2jung@intelligent.pe.kr

0957-4174/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2009.02.052

These are playing an important role of linkage for information flowing on multiple KM systems. Not only this social network establishment but also social interactions can be made among people. Originally, the interactions are happened by comments. (In fact, more advanced comments are trackback.)

In addition, the blogs can be providing a variety of syndication services, e.g., RSS (RDF Site Summary[1]) along to the social linkages between them. It is represented as a family of web feed formats used to publish frequently updated content such as blog entries and news headlines. This is similar to the subscription of mailing lists. Hence, bloggers do not have to access to every blogs by periods, in order to find out new updates.

However, this automatic synchronization is simply based on the explicit social links. When communities are really different or when people do not want to describe explicitly their relationships, hidden social networks have to be inferred from secondary sources. These secondary sources are often the published contents (e.g., documents). The synchronization has to employ a certain estimation process to find out thematic relatedness between bloggers (or knowledge published on the blogs) for better performance.

To do so, in this paper, we want to utilize not only blogrolls but also tags and contents published in the blog-based KMSs. They are depicted solid arrows (blogrolls $A \rightarrow B, A \rightarrow C, C \rightarrow B, B \rightarrow C$) and dotted arrows (tags and contents $\langle$ `A:Korea`, `B:Korea`, `C:Korea`, `Equivalence` $\rangle$) in Fig. 1, respectively. Here, `Equivalence` is a way to represent a semantic relations.

We assume that the shortcomings caused by the traditional RSS-based services (i.e., spreading information to people without considering contexts) can be divided two sub-problems, as follows:

- Contextually irrelevant information may be reached to people. It is very similar to the *spam* mails. It means that in Fig. 1, resources in blog *A* are not guaranteed to be relevant to the blog *B* or *C*, even if there have equivalent tags *Korea* in common.
- It is difficult for people to receive contextually relevant information by finding out the contextually equivalent tags among blogs. In Fig. 1, two tags *Actions* in blog *B* and *Sports* in blog *C* may be matched with *Equivalence* relation.

In order to solve these problems, we propose a novel architecture, called *blog context overlay network* (BCON), to control information propagation through RSS channels on blog-based KMSs. Thereby, we want to focus on a set of tags representing the corresponding blogs' contexts, and present context matching algorithm for analyzing contextual relatedness (Jung & Euzenat, 2007; Jung & Koo, 2007) between both blogs. The most important aim of this algorithm is to uncover the hidden relationships between people whose contexts are most similar, even though they have been directly known with each other yet. Moreover, we want to organize a set of cohesive blog group (i.e., the contextually similar people), and refer to this blog group as a blog community of practice (BCoP). Thus, this paper hypothesizes that knowledge newly published in a blog can be propagated to any blogs of the same BCoP, and tests this hypothesis by conducting experimentations.

In the following Section 2, we will show mathematical formalization of a blog-based KMS, and describe how to fulfill context matching between two blogs. Section 4 explains query-free knowledge distribution on the proposed BCON platform. In Section 5, we will show the experimental results obtained from evaluation process, and then some issues related to blog context representation and matching will be discussed. Finally, in Section 6, we will draw a conclusion of this study and future work.

[1] http://web.resource.org/rss/1.0/.

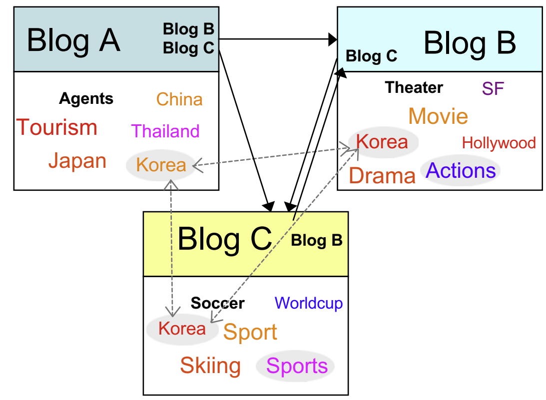

Fig. 1. Basic functionalities of blogs. It shows three main components of blogs, which are resources, blogrolls, and tags (tag clouds).The solid arrows are indicating blogrolls among blogs. More interestingly, the dotted ones are resulted from string matching between tags in different blogs.

2. Blog *context* overlay network

Most of these sources from knowledge blogosphere are based on explicit links that still need from knowledge creators (i.e., the blogger such as document authors and web page designers) to know each others. In this paper, we investigate the dual principles of:

- using the knowledge structure posed in individuals' blogs in order to infer some of their social relationships, and
- taking advantage of this inferred social structure in order to help people sharing their knowledge.

For these purposes, we introduce a structure made of three superposed networks that are assumed to be strongly linked with each other. In this paper, the proposed BCON for these blog-based knowledge management systems can be modeled as a three-layered architecture, which is composed of three layers; (*i*) blog layer, (*ii*) tag layer, and (*iii*) context layer. We call this stack of interlinked

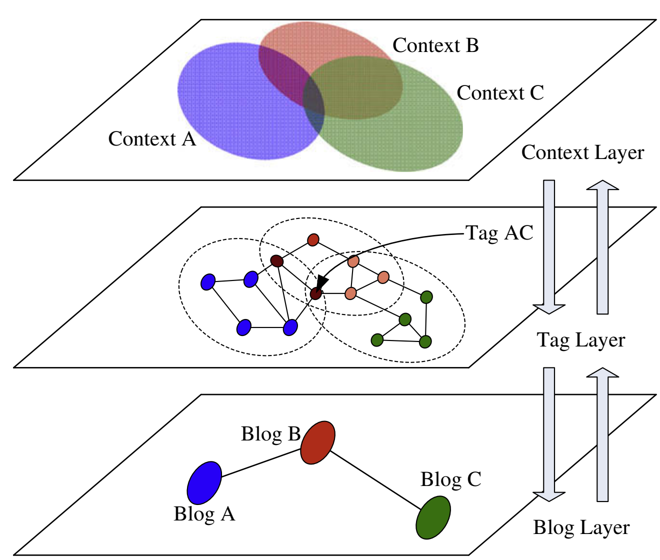

Fig. 2. A three-layered architecture for context matching on knowledge blogosphere.

networks a blog context overlay network (BCON). Fig. 2 is simply depicting the proposed BCON architecture for context matching between blog-based KMSs on the blogosphere.

Blog layer relating people on the basis of common interests;
Tag layer relating tags on the basis of explicit import relationships or implicit similarity;
Context layer relating contexts on the basis of explicit organizational relationships or implicit similarity.

In this architecture, context matching process for social relation discovery is conducted by the following two phases;

1. bottom–up phase from blog layer to context layer, and
2. top–down phase from context layer to blog layer.

Even though we can build various social networks by using blogrolls as well as co-occurrence patterns (e.g., resources and tags) (Staab et al., 2005), we want to find out contextual relationships among blog-based KMSs. In order to compare the contexts of two blogs, it is too difficult to directly analyze contents (e.g., posts) themselves stored from the blog-based KMSs. Thereby, in this paper, we assume that the context of a blog is implicitly represented as a set of corresponding tags while being transferred from the blogger (i.e., knowledge creator). It means that context matching between two blogs B_i and B_j can be replaced to context matching between two sets of corresponding tag.

However, main problem of this context comparison between the tags is that each tag expressing specific knowledge is *ambiguous* and *heterogeneous* with each other. As a simple example, some posts about 'Scientific Fiction' in blogs are tagged with '*sf*' in common. A tag *sf* in blog B_i, meaning 'Scientific Fiction', is contextually different from a tag *sf* in blog B_j, which is meaning 'San Francisco'. Hence, given a pair of the blog-based KMSs, we have to discover the best alignment condition between two tag sets, as finding out which two tag elements from both blogs are either contextually equivalent, opposite, or similar.

2.1. Contextualization of blogs

Above all, we have to assume that the knowledge blogosphere indicates a finite information space composed of a set of blogs which people can freely access to and make a link (e.g., blogroll) in their own blogs. The proposed BCON architecture can be employed in this blogosphere with a set of authorized blog-based KMSs.

Here, we want to explain how the context in the proposed BCON architecture can be managed and manipulated to enable better knowledge distribution services. A blog-based KMS is enhanced from a simple blog by contextualizing the tags. Consequently, the third layer can be built for representing the contexts. While a generic blog is simply composed of three elements (posts, tags, and blogrolls), a blog-based KMS B_k is formulated as

$$B_k = \{P_{B_k}, T_{B_k}, R^{Tag}_{B_k}, R^{Blogroll}_{B_k}, R^{Cmt}_{B_k}\} \tag{1}$$

where

- $\mathscr{P}_{B_k} = \{r_1, r_2, \ldots, r_{N_\mathscr{P}}\}$ is a set of posted knowledge (including simple textual documents as well as multimedia data, e.g., images, movies, and so on),
- $\mathscr{T}_{B_k} = \{t_1, t_2, \ldots, t_{N_\mathscr{T}}\}$ is a set of tags represented as user-generated keywords,
- $\mathscr{R}^{Blogroll}_{B_k} \subseteq B_k \times B_{k'}$ means a directed and unweighted social network which is explicitly represented as blogrolls among blogs within the knowledge blogosphere,

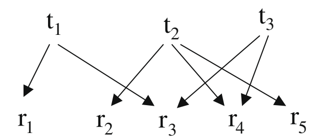

Fig. 3. A bipartite structure of tagged knowledge network.

- $\mathscr{R}^{\mathscr{T}}_{B_k} \subseteq \mathscr{P}_{B_k} \times \mathscr{T}_{B_k}$ is a set of linkages between knowledge $\mathscr{P}_{B_k}$ and tags $\mathscr{T}_{B_k}$ within a blog B_k, which is represented as a bipartite network, and
- $\mathscr{R}^{Cmt}_{B_k} \subseteq B_k \times B_{k'}$ indicates social interactions of which types can be *comment* and *trackback* between the blogs.

Here, the variables $N_\mathscr{P}$ and $N_\mathscr{T}$ are the numbers of posted knowledge and tags in the corresponding blog, respectively. The structure of tagging relationship $\mathscr{R}^\mathscr{T}$ is similar to a bipartite network, because the knowledge can be tagged by more than one tag, and the tags can also have relationships with more than single knowledge, as shown in Fig. 3.

In practice, within a blogosphere, $\mathscr{R}^{Blogroll}$ and $\mathscr{R}^{Cmt}$ are represented as asymmetric square matrices including the information about social awareness. Particularly, in case of the social interactions $\mathscr{R}^{Cmt}_{B_k}$, we want to record shortest path distance $spd(k, k')$ between two blogs, by computing matrix multiplication (Wasserman & Faust, 1994). If there are no path between B_k and $B_{k'}$, $spd(k, k') = M + 1$ where M is the diameter of $\mathscr{R}^{Blogroll}$ in the blogosphere. Of course, this formalization is rather simplified in this paper. But, important point is that these five components can cover main characteristics and features of other supplementary blog services. Within the blogosphere, mainly, the following three types of features are applied to contextualize the tags used in the blog-based KMS.

2.1.1. Background knowledge $\mathscr{O}$

It can indicate ontologies for lexical semantics and directories for topics. Meaning of each tag can be retrieved from the background knowledge (Moldovan & Novischi, 2004). A variety of senses defined in ontologies and thesauri (e.g., WordNet[2]) can be applied to compare the tag contexts (Chan & Ng, 2005; Curtis, Cabral, & Baxter, 2006). Certainly, if we have more background knowledge available, it can be applied to contextualize the tags. In this paper, we have emphasized to use WordNet and derive only three simple semantics for dealing with synonyms, antonym and hypernyms. Thus, we can make a tag t_i in blog B_k more contextually relaxed with semantics given from the ontologies,

$$\mathscr{O}(t_i) = \{Synonym(t_i), Antonym(t_i), Hypernym(t_i) | t_i \in \mathscr{T}_{B_k}\} \tag{2}$$

For example, as shown in Eq. (3), we can find out that a tag 'Car' is semantically related to several terms (e.g., 'Vehicle' and 'Automobile') in difference senses, by looking up WordNet library as the background knowledge,

$$\mathscr{O}(\text{Car}) = \{\text{Vehicle}, \text{Automobile}, \text{Machine}, \text{Motorcar}, \text{Railcar}\} \tag{3}$$

In practice, semantics of the background knowledge are easily accessible by using many APIs (Moldovan & Novischi, 2004). For instance, WordNet has already developed several functionalities for Java, PHP, JSP, Perl, and so on.

Once the tag has been contextualized to $\mathscr{O}(t_i)$, context matching between two tags t_i and t_j can be approximated to matching between $\mathscr{O}(t_i)$ and $\mathscr{O}(t_j)$. However, we are considering that this contextualization process is playing a simple role of preprocessing of

[2] WordNet. http://wordnet.princeton.edu/.

the tags, because it is impossible to recognize neither (*i*) personal contexts nor (*ii*) consensual contexts.

Thereby, in this paper, we want to decompose a given certain BCON into two-mode networks, i.e., (*i*) between tag layer and context layer ($L_{\mathscr{T}\mathscr{C}}$), and (*ii*) between tag layer and blog layer ($L_{\mathscr{T}\mathscr{B}}$). They are investigated to extract personal contexts and consensual contexts, respectively.

2.2. Tag cohesiveness $COH_{\mathscr{T}}$

To be aware of the personal context, we have to focus on the $L_{\mathscr{T}\mathscr{C}}$. A resource $p_i \in P_{B_k}$ is annotated with a set of tags $\mathscr{T}_{B_k}$, which can be easily acquired from *i*th column of $R^{Tag}_{B_k}$. We assume that this tag set $\mathscr{T}_{B_k}$ in a blog-based KMS have shown *contextual cohesiveness* among them. Thus, the hypothesis of tag cohesiveness can be established that the tags in a same blog (i.e., generated by a corresponding blogger) tend to be semantically consistent, and to be in a similar context. In other words, when the tags are generated to represent the knowledge by a certain human user, they might be represented in a contextually similar way. We refer to this cohesiveness as a *personal* context. As a simple example, if a person is interested in topic 'Movie', then his context of tag *sf* is probably related to 'Scientific Fiction'.

Similar to corresponding analysis (Borgatti & Everett, 1997; Roberts, 2000) and association rule mining, we have to statistically find out which tags are contextually more cohesive than others by the following equation

$$COH_{\mathscr{T}}(B_k) = \left\{ \langle t_1, \ldots, t_X \rangle | \min \sum_{x=1}^{X} \Delta_{\widetilde{t_x} \in \mathscr{O}(t_x), \widetilde{t_y} \in \mathscr{O}(t_y)}(\widetilde{t_x}, \widetilde{t_y}), \; t_x \in \mathscr{T}_{B_k}, t_y \in \mathscr{T}_{B_k} \right\} \tag{4}$$

where Δ indicates a string matching measurements, e.g., Edit, Levenshtein, and Substring distances. Here, variable X should be defined by users, and it means the number of principal tag components of the corresponding personal contexts. Each of the principal tag components can be understood as a center of the contexts in context layer shown in Fig. 2.

2.3. Social affinity $SA_{\mathscr{T}}$

Given two blog-based KMSs B_m and B_n, several sets of tags located within a certain social (geodesic) distance δ_{mn} along to the social relations (e.g., blogrolls) might be represented in a similar way. Instead of measuring the social affinity as $[0, 1]$, We refer to this social affinity pattern as *consensual* contexts of communities.

To be aware of the consensual context, we have to focus on the $L_{\mathscr{T}\mathscr{B}}$. A tag $t_i \in \mathscr{T}_{B_m}$ employed by blogs B_m can be easily acquired from *i*th column of $R^{Tag}_{B_m}$. Also, the social relation of the blogs can be retrieved from $\mathscr{R}^{Blogroll}_{B_m}$. For searching for the neighbors within a certain social distance δ, we can compute the distance by iterative matrix multiplication at δ times $\mathscr{R}^{Blogroll}_{(\delta)}$, which is represented as an asymmetric square matrix. Thus, the neighbors of B_m within distance δ, $NGB_{(\delta)}(B_m)$ are given by

$$NGB_{(\delta)}(B_m) = \left\{ B_n | \mathscr{R}^{Blogroll}_{(\delta)}(m, n) \geqslant \delta - 1 \right\} \tag{5}$$

where the distance δ can be adjusted by users.

Regarding the consensual context, we want to establish two hypotheses, and apply them to the neighbors selected by the social affinity (shown in Eq. (5)). First hypothesis of this social affinity is that as the distance between two blog-based KMSs is closer (i.e., their social affinity is stronger), the tag co-occurrence between two corresponding tag sets is higher. This co-occurrence pattern $C_{\mathscr{T}}(B_m, B_n)$ is computed by

$$C_{\mathscr{T}}(B_m, B_n) = \frac{|\{ \langle t_x, t_y \rangle | \Delta_{\widetilde{t_x} \in \mathscr{O}(t_x), \widetilde{t_y} \in \mathscr{O}(t_y)}(\widetilde{t_x}, \widetilde{t_y}) \leqslant \lambda_{C_{\mathscr{T}}}, t_x \in \mathscr{T}_{B_m}, t_y \in \mathscr{T}_{B_n} \}|}{\max(\mathscr{T}_{B_m}, \mathscr{T}_{B_n})} \tag{6}$$

where $\lambda_{C_{\mathscr{T}}}$ means a user-specific threshold value to control the number of co-occurrence patters. We want to find out the usage patterns of the number of semantically matched tags, as changing the social affinities between two blog-based KMSs.

Second hypothesis is to discover the relationship between consensual context and the social affinities. We regard a community as a group of neighbors extracted by function $NGB_{(\delta)}(B_m)$. It is simply an ego-centric approach (Jung, Juszczyszyn, & Nguyen, 2007).

Then, the consensual context is defined as the most common contexts integrating personal contexts (Eq. (4)) of which blogs are in a same community. Jung (2008) mentions that the context can be merged by intersecting the context components in common. Especially, in pervasive computing, a variety of context fusion methodologies (Sekkas, Anagnostopoulos, & Hadjiefthymiades, 2007) have been investigated, and they will be compared in next section.

The consensual context is given by three different set intersection operations

$$SA_{\mathscr{T}}(B_m) = \cap_1 \left\{ \langle t_1, \ldots, t_X \rangle_{B_n} | \min(COH_{\mathscr{T}}(B_n)), B_n \in NGB_{(\delta)}(B_m) \right\} \tag{7}$$

$$= \cap_2 \left\{ \mathscr{O}(\langle t_1, \ldots, t_X \rangle_{B_n}) | \min(COH_{\mathscr{T}}(B_n)), B_n \in NGB_{(\delta)}(B_m) \right\} \tag{8}$$

$$= \cap_3 \left\{ \frac{1}{\delta_{mn}} \times \mathscr{O}(\langle t_1, \ldots, t_X \rangle_{B_n}) | \min(COH_{\mathscr{T}}(B_n)), B_n \in NGB_{(\delta)}(B_m) \right\} \tag{9}$$

where $\frac{1}{\delta_{mn}}$ is expressing the inverse proportion to make an influence to consensual context.

Different from intersection $\cap_1$, the operation $\cap_2$ wants to employ WordNet ontology to conduct word sense disambiguation, and the operation $\cap_3$ wants to employ WordNet ontology and social distance together.

3. Matching between contextualized tags

Beside the numerous relationships that can be found by construction of the context layer, new relationships can be inferred between the entities. One particular relationship that will be interesting here is similarity. In order, to find relationship between contexts from different tags, identifying the entities denoting the same context is a very important feature. As a matter of fact, most of the matching algorithms use some similarity measure or distance in order to match entities.

Some distances, more in the spirit of network analysis, can be defined from the structure of the network. For instance, Euzenat and Valtchev (2004) defines all possible similarities of collection components. In case of tags in blog-based KMSs, such components are not only simple tags, but also personal context (i.e., *COH*) and consensual context component (i.e., *SA*). Given a pair of tags from two different blogs, the similarity measure $Sim_{\mathscr{C}}$ is assigned in $[0, 1]$. The similarity ($Sim_{\mathscr{C}}$) between t and t' is defined as

$$Sim_{\mathscr{C}}(t, t') = \sum_{E \in \mathscr{N}(\mathscr{T})} \pi^{\mathscr{T}}_E MSim_Y(E(t), E(t')) \tag{10}$$

where $\mathscr{N}(\mathscr{T}) \subseteq \{E^1, \ldots, E^n\}$ is the set of all relationships in which tags participate. The weights π^C_E are normalized (i.e., $\sum_{E \in \mathscr{N}(\mathscr{T})} \pi^{\mathscr{T}}_E = 1$).

To find out meaningful relationships between blog-based KMSs as top–down approach (from context layer to blog layer), we want to match between the contexts of the blogs. Most importantly, in this paper, the context of a blog is regarded as a collection of contextualized tags obtained by two major contextualization processes, i.e., *COH* and *SA*, which are described in previous section.

Let two blogs B_i and B_j matched with each other. We have to find out the best matching between three different cases, i.e., (*i*) a pair of simple tag sets, (*ii*) a pair of *COH*-contextualized tag sets, and (*iii*) a pair of *SA*-contextualized tag sets, as maximizing summation of contextual similarities between the given tag collections at each case. Thus, the proposed matching mechanism between two blog-based KMS is given by

$$Sim_{\mathscr{C}}(B_i, B_j) = \max_{\langle t,t'\rangle \in \xi(\mathscr{T}_{B_i}, \mathscr{T}_{B_j})} \sum_{t \in \mathscr{T}_{B_i}, t' \in \mathscr{T}_{B_j}} Sim_{\mathscr{C}}(t, t') \quad (11)$$

$$= \max_{\langle t,t'\rangle \in \xi(COH_{\mathscr{T}}(B_i), COH_{\mathscr{T}}(B_j))} \sum_{t \in COH_{\mathscr{T}}(B_i), t' \in COH_{\mathscr{T}}(B_j)} Sim_{\mathscr{C}}(t, t') \quad (12)$$

$$= \max_{\langle t,t'\rangle \in \xi(SA_{\mathscr{T}}(B_i), SA_{\mathscr{T}}(B_j))} \sum_{t \in T_{B_i}, t' \in T_{B_j}} Sim_{\mathscr{C}}(t, t') \quad (13)$$

where $Sim_{\mathscr{C}} \in [0, 1]$, and ξ provides a matching of the two set of tags. Methods like the Hungarian method allow to find directly the pairing which maximizes similarity. This algorithm is an iterative algorithm that compute this similarity (Euzenat & Valtchev, 2004).

Here, these similarity measurements Eqs. (12) and (13) are able to verify the hypotheses established in Sections 2.2 and 2.3, respectively.

As a result of context matching between blogs, we can obtain a set of tag correspondences

$$\widetilde{\mathscr{M}}(B_i, B_j) = \{\langle t, t', o\rangle | t \in R^{Tag}_{B_i}, t' \in R^{Tag}_{B_j}, o \in \mathscr{O}\} \quad (14)$$

where a correspondence is attached with a semantic relation *o* which is derived from a WordNet ontology $\mathscr{O}$. (For simplicity, $\mathscr{O} = \{\equiv, \sqsubseteq, \sqsupseteq\}$, meaning semantic equivalence and subsumption.) Eventually, we can enhance $\widetilde{\mathscr{M}}$ with social affinity (e.g., blogrolls) as

$$\mathscr{M}(B_i, B_j) \leftarrow \widetilde{\mathscr{M}}(B_i, B_j) \times \frac{1}{\log \delta_{ij}} \quad (15)$$

where δ_{ij} is the social geodesic distance.

4. Knowledge distribution over BCON

As next step of the BCON framework, we want to identify a community in which members are more cohesive than others by using contextual similarities among blog-based KMSs, i.e., the links between blogs are weighed by the contextual similarities. Particularly, we are considering dynamic identification (Tantipathananandh, Berger-Wolf, & Kempe, 2007), because the context might be able to change over time. Knowledge which is newly published on knowledge blogosphere is automatically distributed to the members in a same community.

4.1. Dynamic 'community of practice' identification

Subsequently, we can build an weighted social network of which links and weights are contextual relations and similarity between blogs, respectively. Formation of CoPs can be built by applying *Q*-modular function to the contextual social network. The modularity function $Q^{\diamond}$ is formulated by

$$Q^{\diamond}(\mathscr{S}) = \sum_{i=1}^{k} \frac{\sum_{B_a \in g_i, B_b \in g_i, ADJ_{(ab)}=1} Sim_{\mathscr{C}}(B_a, B_b)}{|g_i|} \quad (16)$$

where all possible pairs of bloggers should be considered.

4.2. Knowledge distribution on tag correspondences

In this work, two ways of knowledge distribution schemes have been proposed, as follows:

- systematic broadcasting of new resource to all members in a same community, and
- selective propagation along to tag correspondences.

Once a resource has newly published and tagged with *t* in B_a, the resource can be automatically pushed into $\langle t, t', o\rangle_{t \in B_a, t' \in B_b}$ in B_b. More importantly, the resource can be translated by tag replacement strategy.

Table 1
Manual evaluation of tag contextualization.

	$\mathscr{O}$	$\mathscr{O} + COH$	$\mathscr{O} + COH + SA$
U_1	0.54	**0.75**	0.58
U_2	0.72	0.68	**0.76**
U_3	0.74	**0.86**	0.64
U_4	0.52	**0.85**	0.82
U_5	0.62	**0.92**	0.9
U_6	0.6	**0.84**	0.63
U_7	0.38	**0.67**	0.58
U_8	0.63	**0.86**	0.53
U_9	0.54	**0.78**	0.48
U_{10}	0.58	**0.95**	0.52
U_{11}	0.63	0.74	**0.78**
U_{12}	0.45	**0.81**	0.57
U_{13}	0.48	**0.93**	0.5
U_{14}	0.58	0.82	**0.84**
U_{15}	0.63	**0.75**	0.65
U_{16}	0.65	**0.83**	0.63
U_{17}	0.43	**0.95**	0.57
U_{18}	0.66	**0.87**	0.78

5. Experimental results and discussion

In order to evaluate performance of the proposed blogosphere architecture, we have implemented collaborative tagging blogging system to share user-generated tags. More importantly, this system has applied context matching process to measure the contextual similarities between blogs. We have focused on two main evaluation issues to test the hypotheses.

- Evaluation of tag contextualization.
- Evaluation of knowledge distribution.

We have been exploiting WordPress blog open platform[3] and Blojsom platform.[4] For collecting the dataset, we have invited 18 graduated students (U_1 to U_{18}). They were asked them to define their offline social relations and to manage their contents on their own blogs. Their activities were monitored during ten days.

5.1. Evaluation of tag contextualization

Firstly, we have evaluated three different contextualization methods for matching tags in blog-based KMSs, as follows:

- using background knowledge (e.g., WordNet) ($\mathscr{O}$);
- using background knowledge and tag cohesiveness ($\mathscr{O} + COH$);
- using background knowledge, tag cohesiveness and social affinity ($\mathscr{O} + COH + SA$).

Table 1 shows the precision of contextual tags extracted by each method. This results were justified all students. Fifteen student of them (i.e., 83.3%) chose $\mathscr{O} + COH$ methods as best one for tag contextualization.

[3] WordPress. http://wordpress.org/.
[4] Blojsom. http://wiki.blojsom.com/wiki/display/blojsom3/About+blojsom.

Table 2
Contextual similarity between users.

	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_{12}	U_{13}	U_{14}	U_{15}	U_{16}	U_{17}	U_{18}
U_1	–																	
U_2	0.24	–																
U_3	0.32	**0.68**	–															
U_4	0.29	0.3	0.45	–														
U_5	0.42	0.25	0.24	**0.83**	–													
U_6	0.26	**0.74**	**0.78**	0.48	0.37	–												
U_7	0.18	**0.76**	**0.92**	0.26	0.45	**0.83**	–											
U_8	**0.86**	0.52	0.33	0.34	0.28	0.22	0.2	–										
U_9	0.41	0.42	0.21	**0.74**	**0.82**	0.35	0.52	0.44	–									
U_{10}	**0.82**	0.18	0.51	0.51	0.3	0.29	0.42	**0.85**	0.57	–								
U_{11}	0.29	0.23	0.52	**0.64**	**0.75**	0.49	0.18	0.56	**0.86**	0.1	–							
U_{12}	**0.91**	0.46	0.43	0.3	0.14	0.58	0.45	**0.72**	0.22	**0.72**	0.26	–						
U_{13}	0.27	0.37	0.18	**0.75**	**0.93**	0.37	0.37	0.3	**0.78**	0.05	**0.82**	0.4	–					
U_{14}	**0.68**	0.34	0.35	0.15	0.43	0.3	0.52	**0.68**	0.25	**0.85**	0.3	**0.83**	0.53	–				
U_{15}	0.35	**0.67**	**0.83**	0.39	0.5	**0.95**	**0.98**	0.28	0.36	0.52	0.25	0.43	0.3	0.44	–			
U_{16}	0.27	**0.84**	**0.73**	0.42	0.18	**0.84**	**0.88**	0.33	0.28	0.42	0.38	0.13	0.45	0.28	**0.82**	–		
U_{17}	0.53	**0.74**	**0.85**	0.27	0.42	**0.68**	**0.76**	0.43	0.23	0.18	0.38	0.45	0.37	0.21	**0.67**	0.93	–	
U_{18}	0.35	0.49	0.22	**0.95**	**0.74**	0.13	0.53	0.44	**0.69**	0.21	**0.83**	0.05	**0.74**	0.42	0.43	0.49	0.28	–

It proves that the proposed approaches (i.e., $COH_{\mathcal{T}}$ and $SA_{\mathcal{T}}$) outperform the simple word sense disambiguation based on background ontologies. Especially, $COH_{\mathcal{T}}$ has shown better performance than $SA_{\mathcal{T}}$. We found out that the explicit social relationships (e.g., blogrolls) can not indicate the contextual association between blog-based KMSs.

5.2. Evaluation of knowledge distribution

Table 2 shows contextual similarity of all possible pairs of blogs, in case of the best result of tag contextualization (i.e., $\mathcal{O} + COH$). Thus, we have chosen $\mathcal{O} + COH$ to compare a pair of contextualized tag collections.

As shown in Table 2, we measured the contextual similarity ($Sim_{\mathcal{C}}$) between all possible pairs of bloggers. Hence, by using community identification equation Eq. (16) ($k = 3$), we have organized three communities with highly cohesive blogs, as follows:

- Community_1: U_1, U_8, U_{10}, U_{12}, U_{14} (i.e., $Q^{\diamond}(\text{Community}_1) = 0.792$).
- Community_2: U_2, U_3, U_6, U_7, U_{15}, U_{16}, U_{17} (i.e., $Q^{\diamond}(\text{Community}_2) = 0.817$).
- Community_3: U_4, U_5, U_9, U_{11}, U_{13}, U_{18} (i.e., $Q^{\diamond}(\text{Community}_3) = 0.792$).

We found out that the proposed community identification method is working very well, because the modularity value is relatively high. In particularly, although Community_2 have more members, the cohesiveness is higher than other communities.

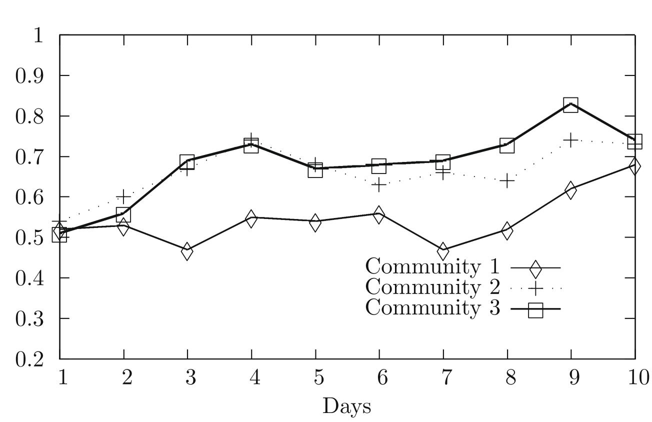

Fig. 4. Measuring user satisfaction with RSS feeds via BCON.

As second experimentation, we have conducted human evaluation for RSS-based information recommendation. During 10 days, we kept tracking of bloggers' rating patterns, as shown in Fig. 4. The precision (i.e., user satisfaction) of context-based RSS feeds in three communities has been increased over time in common. This is caused by the dynamic community identification process. The bloggers were able to be automatically involved into more relevant communities.

6. Concluding remarks and future work

As a conclusion, we have claimed a new measurement for contextual similarity between blogs.

Jung (2008) mentions that the context can be merged to foster semantically heterogeneous users who are not socially acquaint with each other.

When communities are really different or when people do not want to describe explicitly their relationships, additional social networks have to be inferred from secondary sources. These secondary sources are often the published contents (e.g., documents). For example, bibliometrics has for long made a speciality of inferring social networks and many other clues from databases of co-authoring and citations. The web is a rich source of documents that can be used as secondary sources for inferring social networks from analyzing the hyperlinked structure on the web (Kleinberg, 1999) (along the way used by google for inferring most authoritative web pages) to exploiting the more intimate personal infospheres made of web pages, weblogs, and so on.

Similar to Mika (2005), we may consider only the concepts applied to the specific knowledge. As mentioned in Tantipathananandh et al. (2007), dynamic CoP identification is NP-hard and APX-hard problem. We might have to evaluate our heuristics computing the similarity. Furthermore, this similarity can be extended to contextual centrality, which means the potential power of bridging among blogs on social network.

As future work of semantic centrality, we have three main plans to investigate the followings issues (*i*) *semantic subgroup discovery*, to organize the sophisticated user groups with enhancing Eq. (16), (*ii*) *query propagation*, to determine the ordering (or route) of potential peers to which the queries will be sent, and (*iii*) *semantic synchronization,* to maximize the efficiency interoperability by information diffusion. Furthermore, we have to consider to enhance the semantic centrality measurement $C^{\diamond}$ by combining with (i) authoritative and hub centrality measurement, proposed in Kleinberg (1999), and (ii) the modified shortest paths $spd(n,t) = \frac{1}{C^{\diamond}(n)+C^{\diamond}(t)}$. Finally, based on Jung (2005), we have plan to visualize the semantic dynamics on the social network.

Acknowledgement

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MEST) (No. R01-2008-000-10196-0).

References

Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks, 19*(3), 243–269.

Chan, Y. S., & Ng, H. T. (2005). Word sense disambiguation with distribution estimation. In L. P. Kaelbling & A. Saffiotti (Eds.), *Proceedings of the 19th international joint conference on artificial intelligence (IJCAI-05), Edinburgh, Scotland, UK, July 30–August 5* (pp. 1010–1015). Professional Book Center.

Curtis, J., Cabral, J., & Baxter, D. (2006). On the application of the Cyc ontology to word sense disambiguation. In G. Sutcliffe & R. Goebel (Eds.), *Proceedings of the 19th international Florida artificial intelligence research society conference, Melbourne Beach, Florida, USA, May 11–13* (pp. 652–657). AAAI Press.

Euzenat, J., & Valtchev, P. (2004). Similarity-based ontology alignment in OWL-Lite. In R. L. de Mántaras & L. Saitta (Eds.), *Proceedings of the 16th European conference on artificial intelligence (ECAI 2004), August 22–27* (pp. 333–337). IOS Press.

Jung, J. J. (2005). Collaborative web browsing based on semantic extraction of user interests with bookmarks. *Journal of Universal Computer Science, 11*(2), 213–228.

Jung, J. J. (2008). Ontology-based context synchronization for ad-hoc social collaborations. *Knowledge-Based Systems, 21*(7).

Jung, J. J., & Euzenat, J. (2007). Towards semantic social networks. In E. Franconi, M. Kifer, & W. May (Eds.), *Proceedings of the 4th European semantic web conference (ESWC 2007), Innsbruck, Austria. Lecture notes in computer science* (Vol. 4519, pp. 267–280). Springer.

Jung, J. J., & Koo, C.-M. (2007). Contextual matching-based blog overlay network for information sharing on blogoshphere. *Telecommunication Review, 17*(4), 651–659 (in Korean).

Jung, J. J., Juszczyszyn, K., & Nguyen, N. T. (2007). Centrality measurement on semantically multiplex social networks: Divide-and-Conquer Approach. *International Journal of Intelligent Information and Database Systems, 1*(3/4), 277–292.

Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM, 46*(5), 604–632.

Mika, P. (2005). Ontologies are us: A unified model of social networks and semantics. In Y. Gil, E. Motta, V. Richard Benjamins, & M. A. Musen (Eds.), *Proceedings of the 4th international semantic web conference (ISWC 2005), November 6–10. Lecture notes in computer science* (Vol. 3729, pp. 522–536). Springer.

Moldovan, D., & Novischi, A. (2004). Word sense disambiguation of wordnet glosses. *Computer Speech and Language, 18*(3), 301–317.

Roberts, J. M. (2000). Correspondence analysis of two-mode network data. *Social Networks, 22*(1), 65–72.

Sekkas, O., Anagnostopoulos, C. B., & Hadjiefthymiades, S. (2007). Context fusion through imprecise reasoning. In *Proceedings of the IEEE international conference on pervasive services 2007 (ICPS 07)*.

Staab, S., Domingos, P., Mika, P., Golbeck, J., Ding, L., Finin, T., et al. (2005). Social networks applied. *IEEE Intelligent Systems, 20*(1), 80–93.

Tantipathananandh, C., Berger-Wolf, T., & Kempe, D. (2007). A framework for community identification in dynamic social networks. In *Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'07)* (pp. 717–726).

Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge University Press.