Folksonomy-Based Information Retrieval in Context-aware Environment

Sungrim Kim† and Joonhee Kwon††,

† Department of Internet Information, Seoil College,
49-3, Myonmok-dong, Jungrang-Ku, Seoul, Korea
†† Department of Computer Science, Kyonggi University,
San 94-6, Yiui-dong, Yeongtong-ku, Suwon-si, Kyonggi-do, Korea

Summary
The Web 2.0 approach has revolutionized the way we use the information system. An essential part of Web 2.0 is folksonomy. Most folksonomy-based systems do not consider new breed of context-aware retrieval. It requires the new convergence of folksonomy-based information retrieval and context-aware systems.We propose a folksonomy-based information retrieval in context-aware environment. We describe the procedure and some experiments are performed. The experiments show our method outperform other approaches. Our proposed method is an effective information retrieval method in folksonomy-based context information retrieval.

Key words:
Folksonomy, Information Retrieval, Context, Web 2.0

1. Introduction

Web 2.0 systems, such as Flickr or Delicious have acquired large numbers of users within less than two years. The reason for their immediate success is the fact that no specific skills are needed for participating, and that these systems yield immediate benefit for each individual use without too much overhead [1].

An essential part of Web 2.0 is harnessing collective intelligence called folksonomy [2]. The frequent use of systems such as Flickr or Delicious shows clearly that folksonomy-based approaches are able to overcome the knowledge acquisition bottleneck [1]. The term folksonomy is generally attributed to Thomas Vander Wal. [3]. It is a portmanteau of the words folk and taxonomy that specifically refers to subject indexing systems created within internet communities.

Most approaches to searching folksonomy-based systems are to employ traditional information retrieval. However, most approaches suffer from mandatory tagging. Moreover, they don't consider new breed of context-aware environment. It requires the new convergence of folksonomy-based information retrieval and context-aware systems.

Context-aware systems require proactive information retrieval, where information is presented to the user automatically [4]. Compared with it, previous folksonomy-based information retrieval systems are interactive, where the user directly issues a request to retrieve relevant information.

In this paper, we consider folksonomy-based information retrieval in context-aware environment. Our method uses contexts as tags. Automatic tagging generated from context information provides effective folksonomy-based information retrieval in context-aware environment. Information is proactively suggested by taking into account the user's current context.

This paper is organized as follows. We start giving an overview of the related works, and then describe a folksonomy-based information retrieval method in context-aware environment. Finally, we present some experiments, showing the benefits of our system and discuss the results and conclusions of our work.

2. Related Works

In Web 2.0 services, prosumers, that is producers and consumers, collaborate not only for the purpose of creating content, but to index these pieces of information as well. Folksonomy permits actors to describe documents with subject headings, tags [5]. The term folksonomy is a portmanteau of the words folk and taxonomy [3].

We are witnessing an increasing number of tagging services on the web, such as Flickr [6] and Delicious [7]. The systems can be distinguished according to what kind of resources is supported. Flickr, for instance, allows the sharing of photos and Delicious allows the sharing of bookmarks. In their core, these systems are all very similar. Once a user is logged in, he can add a resource to the system, and assign arbitrary tags to it. The collection of all his assignments is his personomy, the collection of all personomies constitutes the folksonomy. The user can

Manuscript received November 5, 2008
Manuscript revised November 20, 2008

explore his personomy, as well as the personomies of the other users. When clicking on a tag, one sees information with the tag [1]. As shown in the previous systems, folksonomies have a shortcoming that users assign tags manually. It makes uncontrolled vocabulary and inconvenience.

Recently, the dawn of the age of ubiquitous computing is upon us. Context-aware computing is a core technology of ubiquitous computing. Context is any information that can be used to characterize the situation of an entity. An entity is any person, place or object that is considered relevant to the interaction between a user and an application, including the user and application themselves [8]. A recent analysis of promising future context-aware applications reports that one of a key component is context-aware information retrieval in context-aware environment [4].

There are a few folksonomy-based systems, such as Tagzania, Sociallight, DSG project in context-aware environment. Tagzania [9] is a geofolksonomy and a mash-up which combines social tagging in a Delicious style with geographical information coming from GoogleMaps. With the help of its web front-end, users may associate tags to planet locations and then browse through them. The Socialight [10] enables sharing location-based notes, termed StickyShadow, pictures and sounds from the web and a phone. A StickyShadow is made up of media, such as text and a picture, and information about who can see it and when and where that note is available. The DSG [11] aims to tag the planet not only with location but more general context based annotations, which are triggered only when a user's mobile device matches some contextual attributes. However, they do not seriously consider folksonomy-based information retrieval in context-aware environment.

3. Folksonomy-Based Information Retrieval in Context-aware Environment

In this section, we describe our mechanism for folksonomy-based information retrieval in context-aware environment. The first task is to tag the contexts by user generated contexts. The second task is to retrieve information for each user's context. Figure 1 shows the flow of main tasks.

The first task is to tag the contexts in resource information automatically. The task computes the number of occurrences of a context divided by the total number of context occurrences for each resource information. If the value is above a certain threshold, the context is tagged with the computed value in resource information.

The second task is to retrieve information for each user's context. Firstly, resource information with the user's current context tags from source information is extracted. Secondly, the type of information to be provided to the user is determined. In this paper, we provide a tag instead of simply providing the information for each context. It makes more effective folksonomy-based information retrieval. When the number of information to be retrieved is greater than the limited number of information to be provided in the user, we provide contexts as tags. Otherwise, information itself is provided.

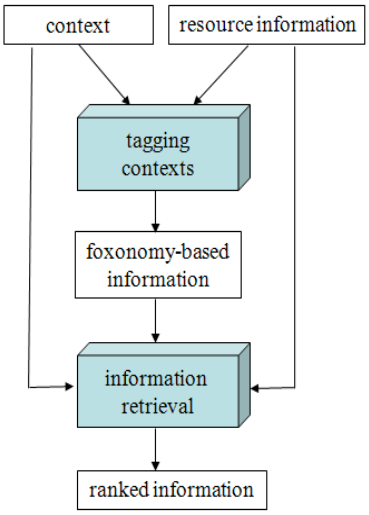


Fig. 1. Flow of folksonomy-based information retrieval

We use the characteristics of context with hierarchy. For example in location context, a location contains other location. Location context has a contain hierarchy. In general, amount of information in a parent location is more than it in a child location. Moreover, in a parent location, user wants to know a child location with specific information, not excessive specific information. Using the characteristics, we retrieve tags instead of information itself.

Finally, the information in the second step is refined in relation to the relevance and attention. For relevant information service for user's context, we adopt the traditional information retrieval, vector space information retrieval scheme [12]. Moreover, we adopt the parameter of attention weight, AW, as found in the previous research [13]. Attention is defined as focused mental engagement on a particular message or piece of information [14]. In the second task, the amount of information needed for a context with a high weight of AW is larger than that needed for a context with a low weight.

A weight of information with context tag is calculated by adapting most often used TF/IDF scheme. We compute the weight for context in resource information using term frequency and inverse document

frequency. We adopt context frequency as term frequency in resource information. And then, similarity between current user's context and resource information is computed. Compared with the traditional method, our method considers the query as user's current context.

In context-aware environment, mobile devices often are used. Considering it, when mobile device's resource is limited, tags is retrieved. The tags are ranked by summation of similarities contained in the current user's context. Otherwise, resource information itself is ranked by similarities.

4. Experiments

4.1 Experimental Setting and Data Sets

We implemented both the proposed method, called Approach1 and the existing method, called Approach2. All programs were implemented in Java. The existing method was implemented without considering context tag, retrieval of context tag, and attention weight.

In the experiments, the number of resource information is 750. We consider the location context with hierarchy. The location context with hierarchy is consisted of top-level zone, middle-level zone, and bottom-level zone. The number of instances of location context is 5 in the top level zone, 19 in the middle-level zone, and 95 in the bottom-level zone. We also varied the tagging rate by 10% from 20% to 50%. The information with tag was randomly selected.

We varied the participation difference between the number of information with context tag and the number of information without it by 10%; that is, from 10% to 100%. In addition, we matched the number of responsive information with number of tagged infomation. Of the responsive information, 80% were tagged information and 20% were non-tagged information.

The attention weight was set to 0.5 in the top level zone. If contexts are tagged in a bottom-level zone, the attention weight of the selling zone as well as that of the middle level zone that contained the selling zone was set to 1. Otherwise, the attention weights were set to 0.5.

The maximum number of information to be retrieved in the user's device was set to 8. For the performance evaluation, we conducted the experiment 50 times.

4.2 Experimental Results

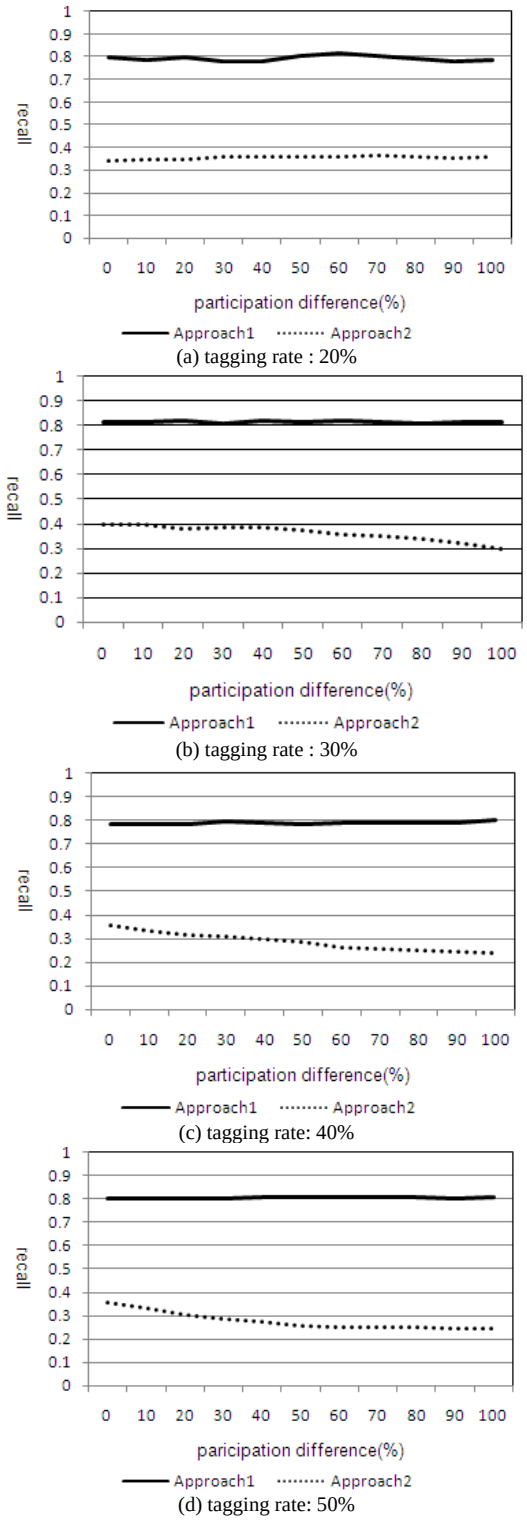


Fig. 2. Average recall in all zone

To evaluate the accuracy of retrieved information, we used the recall of retrieved information. The recall is the proportion of the available relevant information that has been retrieved. The recall is usually defined as shown below.

$$\text{Recall} = \frac{|R \cap P|}{|P|}$$

In the equation, R is the information retrieved to a user, P is the set of information responded by the user, and |S| is the cardinality of a set S. We computed the average recall about top 8 retrieved information.

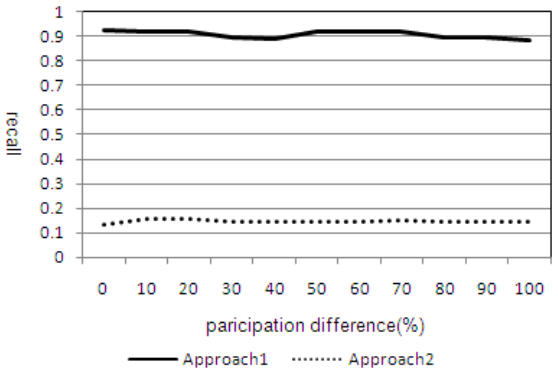
Figure 2 shows the average recall in all zones. Figure 3 and Figure 4 show the average recall for each zone. Several of the same observations are found in these results. First, the average values of the recall in Approach1 are consistently better than those of Approach2. Second, as the tagging rate and participation difference increase, the difference in values of recall between Approach1 and Approach2 increases. This phenomenon is due to the fact that when the number of tagged information is greater, more relevant information is retrieved.

Figure 3 and Figure 4 show the average recall for each zone. Note that the difference between two approaches in Figure 3 is greater than it in Figure 4.

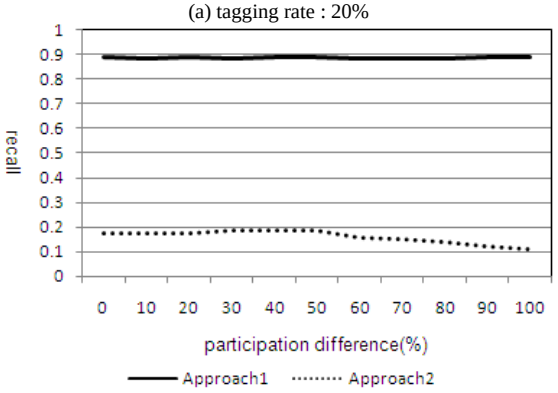
Figure 3 shows the average recall in top and middle-level zone. In those zones, the number of retrieved information increases more than it in bottom-level zones. Recall value of Approach1 is near 1.

Compared it, recall value of Approach 2 is about 0.1. Approach1 retrieves context tags instead of information itself. The context tag enables it to summarize information. It makes to contain more relevant information in top 8 retreived information. Moreover, Approach1 retrieves levelized context tags considering the amount of information provided in top 8 ranking. It makes more difference between two approaches.

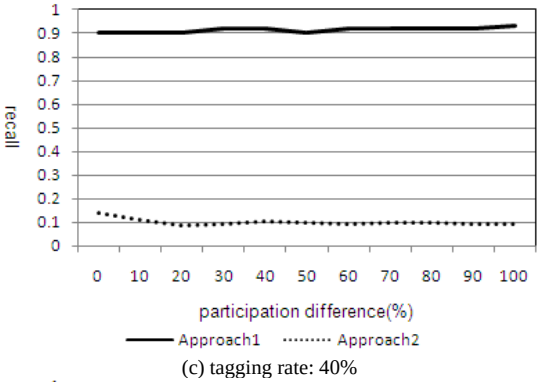
Figure 4 shows the average recall in bottom-level zones. In bottom-level zones, all approaches retrieve information itself. Compared with Approach2, our approach considers user's context and attention. Considering them, the recall value of Approach1 is consistently near 0.7. However, Approach2 shows the more variant results according to the tagging rate and participation difference.



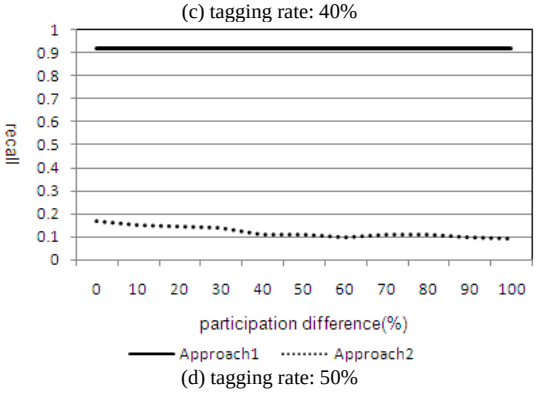
(a) tagging rate : 20%



(b) tagging rate : 30%

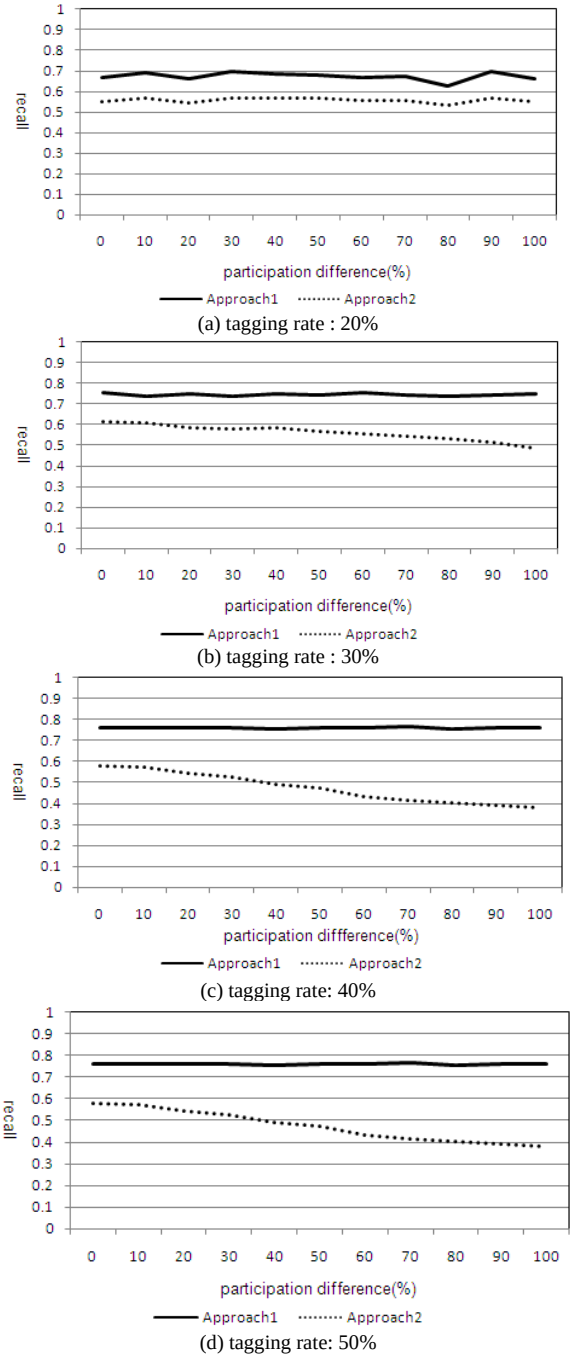


(c) tagging rate: 40%



(d) tagging rate: 50%

Fig. 3. Average recall in top and middle-level zone



(a) tagging rate : 20%

(b) tagging rate : 30%

(c) tagging rate: 40%

(d) tagging rate: 50%

Fig. 4. Average recall in bottom-level zone

5. Conclusion

Web 2.0 and context-aware computing are latest noticeable technology. An essential part of Web 2.0 is harnessing collective intelligence called folksonomy. Promising future context-aware computing technology is context-aware information retrieval.

Folksonomy-based information retrieval is very interesting research area in context-aware environment. However, previous researches do not consider seriously context-aware environment.

We propose a folksonomy-based information retrieval in context-aware environment. Our work shows that the method is suitable for effective information retrieval method in folksonomy-based context information retrieval.

We present new information retrieval method. The method use folksonomy based on contexts and traditional information retrieval method. Our experimental results also confirm that our method has greater accuracy in terms of recall.

Acknowledgments

The present research has been conducted by the Research Grant of Seoil College in 2007.

References

[1] Hotho, A. et al, 2006. Information Retrieval in Folksonomies: Search and Ranking. Proceedings of the 3rd European Semantic Web Conference. Budva, Montenegro, pp. 411-426.

[2] O'Reilly, T., 2005. What Is Web 2.0, O'Reilly Media, Inc., USA, viewed 1 October 2008, <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>.

[3] Vanderwal, T. 2005. Off the Top: Folksonomy Entries November 5 2005, viewed 1 October 2008, <.http://www.vanderwal.net>.

[4] Brown, P. J. and Jones, G. J. F., 2001, Context-aware Retrieval: Exploring a New Environment for Information Retrieval and Information Filtering, Personal and Ubiquitous Computing Vol 5, Issue 4, pp.253 – 263.

[5] Isabella P. and Wolfgang G. S. 2007. Folksonomy and information retrieval, Proceedings of the 70th ASIS&T Annual Meeting. Vol.44, pp.1510-1542.

[6] Flickr, <http://www.flickr.com>.

[7] Del.icio.us, <http://del.icio.us>.

[8] Abowd, G.D. et al, 1999. Towards a Better Understanding of Context and Context-Awareness. In Lecture Notes In Computer Science, Vol. 1707, pp. 304-307.

[9] Tagzania 2005, CodeSyntax, Spain, viewed 1 October 2008, <http://www.tagzania.com/>.

[10] SOCIALIGHT 2005, Kamida, USA, viewed 1 October 2008, <http://socialight.com/>.

[11] Lopez-de-Ipia, D. et al, 2007. A context-aware mobile mash-up platform for ubiquitous web. Proceedings of 3rd IET International Conference on Intelligent Environments, Ulm, Germany, pp. 116-123.

[12] Salton, G. and Buckley, C., 1988. Term-weighting approaches in automatic text retrieval. In Information Processing and Management: an International Journal, Vol. 24, No. 5, pp.513-523.

[13] Kwon, J. and Kim, S., 2006. Multicontext-Aware Recommendation for Ubiquitous Commerce. In Lecture Notes In Computer Science, Vol. 4317, pp. 291-304.
[14] Davenport T. H. 2001. May We Have Your Attention, Please?, Ubiquity 2(17).

Sungrim Kim received the B.S., M.S., and Ph.D. degrees in Computer Science from Sookmyung Women's University, Seoul, Korea in 1994, 1997 and 2002, respectively. She is currently an associate professor, Dept. of Internet Info., Seoil College, Seoul, Korea, Her research interests are ubiquitous computing, web database, multimedia medical database.

Joonhee Kwon received the B.S., M.S., and Ph.D. degrees in Computer Science from Sookmyung Women's University, Seoul, Korea in 1992, 1994 and 2002, respectively. She is currently an associate professor, Dept. of Conputer Science Kyonggi University, Suwon-si, Kyonggi-do, Korea, Her research interests are ubiquitous computing, LBS, spatial database, GIS.