

Information Retrieval using Context Information on the Web 2.0 Environment

Sungrim Kim[†] and Joonhee Kwon^{††},

[†] Department of Internet Information, Seoil University,
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

Web 2.0 has become the face of the next generation web. Web 2.0 is more than anything else a new way for searchers to actually use the Web in a collaborative, interactive way. For web search, information retrieval technique is one of major researches. The recent advance in information retrieval is retrieval using context information. Considering the trends, a new information retrieval method is needed on context-aware Web 2.0 environment. We propose a new information retrieval method using contextual information on the Web 2.0 environment. We describe the procedure for the just-in time information retrieval.

Key words:

Information Retrieval, Context, Web 2.0, tag

1. Introduction

During the last years, the emergence of the Web 2.0 has revolutionized the way information is designed and accessed over the internet [1]. There are a number of Web-based services and applications that demonstrate the foundations of the Web 2.0 concept. These include blogs, wikis, multimedia sharing services, and content tagging services [2]. Especially, tagging as a phenomenon corresponds with a Web 2.0 mentality that users can create not only content but a richer, more adaptive and responsive way to navigate and search both existing and new media. The tagging promises better and more intuitive information access through tag-based browsing, information retrieval [3].

A user's context affects how they interact with an information retrieval system, what type of response they expect from a system and how they make decisions about the information objects they retrieve [4]. An information retrieval system is concerned with retrieving all and only those documents which are relevant to any information need of any user. Information retrieval is intrinsically context dependent; what is relevant to one user in one place at one time can no longer be relevant to another user,

in another place, or at another time. In principle, an information retrieval system should be context aware [5].

This situation calls for new approaches to information retrieval that incorporates the contextual information into the retrieval process, with the aim of delivering information to the users that is relevant within their current contexts.

In this paper, we propose an information retrieval method using context information on the Web 2.0 environment. In suggesting a new information retrieval method, we adopted the adapted PageRank [13] and context tags [14]. First, our proposed method is to model folksonomies with context tags. Then, it combines vector space based retrieval and adapted PageRank algorithms. Our proposed method provides the just-in time information retrieval.

The remainder of this paper is organized as follows. After introducing the related works in Section 2, the idea of information retrieval method using context information is illustrated in Section 3. Finally, in Section 4 we conclude the paper.

2. Related Works

A new breed of so-called 'Web 2.0' applications is currently emerging on the Web. These include user-centric publishing and knowledge management platforms like wikis, blogs, and social resource sharing tools [13].

An essential part of Web 2.0 is harnessing collective intelligence called folksonomy [9]. Folksonomy is collaborative tagging and method of collaboratively creating and managing tags to annotate and categorize content. Using traditional information retrieval, folksonomy contents can be searched textually. However, as the documents consist of short text snippets only (e.g., the web page title, and the tags themselves), ordinary ranking schemes such as TF/IDF are not feasible. FolkRank is adapted PageRank algorithm. In order to

employ a weight-spreading ranking scheme on folksonomies, FolkRank transforms the hyper-graph into an undirected graph. Then it applies a differential ranking approach that deals with the skewed structure of the network and the undirectedness of folksonomies [13].

Folksonomy-based information retrieval method uses contexts as tags [14]. 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.

Context-Aware Retrieval (CAR) systems are concerned with the acquisition of context, its understanding, and the application of behavior based on the recognized context. Thus the CAR model includes, among the classical information retrieval model elements, the user's context, that is both used in the query formulation process and associated with the documents that are candidates for retrieval [6, 7].

The collaborative annotation refers to the practice and methods of collaboratively creating and managing metadata to annotate and categorize content. Collaborative annotation is a feature of many Web 2.0 services, where it is known also as social tagging. With the diffusion of Web 2.0 services, social tagging has gained importance, thanks to the easy and informal approach that allows also non-expert users to classify and find information: differently from traditional hierarchical system, users can freely choose tags to categorize contents. Although most people use tagging to organize their own content collection, even resources tagged for personal use can benefit other users. For example, if many users find something funny, there is a reasonable likelihood that someone else would also find it to be so [6, 8].

Context-sensitive retrieval has been identified as a major challenge in information retrieval research. Several context-sensitive retrieval algorithms exist in the literature, most of them based on statistical language models to combine the preceding queries and clicked document summaries with the current query, for better ranking of documents. Towards the optimal retrieval system, the system should exploit as much additional contextual information as possible to improve the retrieval accuracy, whenever this is available. Ontology-based personalization is based on an ontology-driven representation of the domain of discourse, providing enriched descriptions of the semantics involved in retrieval actions and preferences, and enabling the definition of effective means to relate preferences and context [18]. However, they do not seriously consider information retrieval using context information on the Web 2.0 environment.

3. Information Retrieval using Context Information

Our method retrieves relevant information using the contexts on the Web 2.0 environment, and then provides the just-in-time information. In order to retrieve information on Web 2.0 context-aware environment, we consider both relevance to user's contexts and popularity. Moreover, mobile environment is considered.

The method is composed of two main tasks. The first task is to retrieve information using context and tags on Web 2.0. The second task is to provide information for each user's contexts in mobile computing environment. Figure 1 shows the flow of the main tasks. We describe the steps in the following section.

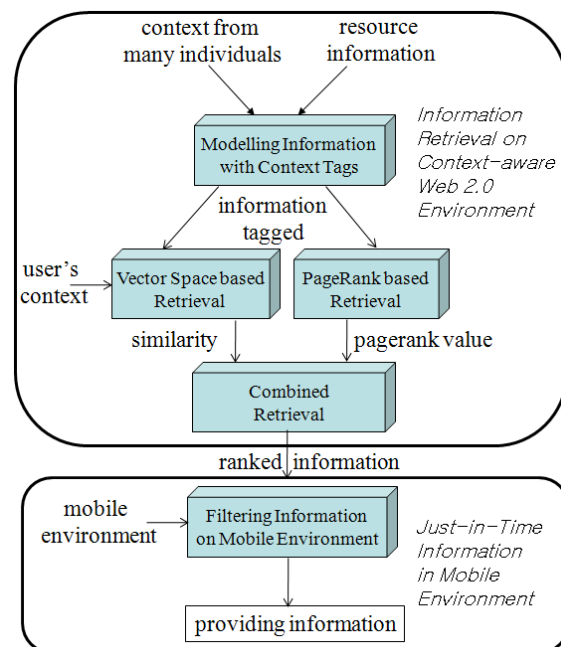


Fig. 1. Flow of information retrieval using context information

3.1 Information Retrieval on Context-aware Web 2.0 Environment

Traditional information retrieval is finding documents of a text that satisfies an information need from within large collections [12]. Context-aware environments provide a challenging and exciting new domain for information retrieval. Identification of relevant information can be achieved by integration of existing methods from information retrieval and context-aware technologies [11].

In this task, we propose a new method for information retrieval on context-aware Web 2.0 environment. Compared with the traditional information retrieval method, the method considers both context and

Web 2.0. For information retrieval on Web 2.0 environment, we adopt the adapted PageRank by Hotho [13]. Moreover, we adopt the context tags by Kim [14] for information retrieval on context-aware Web 2.0 environment.

Conventional search engines consider both relevance and popularity to retrieve information. Vector space model is the most popular method for relevance [5, 10]. For popularity, PageRank method is the most popular [8]. In our method, vector space model and PageRank is adopted for relevance and popularity of information.

The first task is to model folksonomies with context tags. The modeling algorithm is presented in [15]. The second task is to ranking information matching user's context.

In the second task, our method combines vector space model and PageRank algorithms. The task is composed of 3 steps. In the first step, we compute vector space similarity values on context-aware Web 2.0 environment. We adopt the existing our previous research [15]. A weight of information with context tag is calculated by adapting TF/IDF scheme. The context frequency is considered as term frequency in resource information. And then, similarity between current user's context and resource information is computed. User's current contexts are considered as a query.

In order to find some relevant information to a specific user's context C_q , similarity SC between a context C_q and information I_i is computed as following equation. w_{ji} is the weight for context C_j in information I_i , w_{jq} is the weight for context C_j in user's context C_q , and k is the number of context tags tagged in source information.

$$SC(C_q, I_i) = \sum_{j=1}^{j \leq k} w_{jq} * w_{ji} \quad (1)$$

In the second step, PageRank values on context-aware Web 2.0 are computed. Result from the first task is a graph with the vertices of the sets of tags, users and resource information. In this step, adapted PageRank algorithm based on the model are performed. Then, the resource information is assigned PageRank value globally. The value means that resource information which is tagged with many context tags becomes important.

In the third step, the result of the first step and the second step is composited. Formally, our method computes the composited ranking value on a context C_q and information I_i , $CR(C_q, I_i)$.

The composited ranking value for a information I_i is combination of the result of the second step, $PR(I_i)$ together with the context-dependent value from the first step. The result of first step and it of the second step are composited using single adjustable parameter α ($0 \leq \alpha \leq 1$).

$$CR(C_q, I_i) = \alpha * SC(C_q, I_i) + (1 - \alpha) * PR(I_i) \quad (2)$$

3.2 Just-in-Time Information in Mobile Environment

From the previous section, our method retrieves information considering user's context and popularity. Conventional search engine simply provides sorted information by rank value. However, in context-aware environment, information is mostly provided in mobile environment.

One of the main characteristics of mobile environment is limited resources such as limited memory, limited computational power, and small screen [16, 17]. Our method considers the limited resource of the mobile environment. It makes information to provide just-in-time on the mobile environment.

The first task is to sort information by ranking values. The second task is to provide information considering mobile environment.

In the first task, the ranking values on all information for a user and user's current context are sorted in descending order. And then, information is sorted according to sorted ranking.

In the second task, mobile environment such as small screen is considered. The amount of information is restricted to a limited amount, K , by small screen of mobile device. Our method filters only top- K information from the result of the first task.

4. Conclusion

Information retrieval is fast becoming the dominant form of information access, overtaking traditional database style searching [12]. However, previous researches are focused on traditional computing environment.

Web 2.0 and context-aware computing are the latest trends in information technology. Considering the trends, a new information retrieval method is needed on context-aware Web 2.0 environment.

We propose an information retrieval method using context Information on the Web 2.0 Environment. Our method retrieves information using the tagged contexts on the Web 2.0 environment.

In order to retrieve information on Web 2.0 context-aware environment, we consider both relevance to user's contexts and popularity. For relevance, we adopt traditional vector space method. For popularity, PageRank method is used.

The idea behind PageRank is that pages visited more often are more important. We already proposed the context tags in our previous research [14]. The context tags are useful information on context-aware Web 2.0 environment. In this paper, the idea of PageRank is

adopted for information retrieval using the context tags. That is, information tagged context tags more often is more important.

Moreover, the mobile environment is considered for context-aware computing environment. We consider a limited resource of the mobile device such as small screen. Our method filters only top-K information from retrieved information.

In future work, we will implement the algorithm using context from sensor such as RFID. We also will make some applications using this method.

Acknowledgments

The present research has been conducted by the Research Grant of Seoil University in 2008.

References

- [1] Mohanad Al-Jabari, Michael Mrissa, and Philippe Thiran, "Handling Users Local Contexts in Web 2.0: Use Cases and Challenges", Workshop on Adaptation and Personalization for Web 2.0, UMAP'09, June 22-26, 2009
- [2] Paul Anderson, "What is Web 2.0? Ideas, technologies and implications for education", JISC Technology and Standards Watch, Feb. 2007
- [3] Robert Graham, Brian Eoff, and James Caverlee, "Plurality: A Context-Aware Personalized Tagging System", WWW 2008, April 21-25, 2008, Beijing, China
- [4] Fabio Crestani and Ian Ruthven, "Introduction to special issue on contextual information retrieval systems", Information Retrieval, Vol.10, No.2, pp.111-113
- [5] Massimo Melucci, "A basis for information retrieval in context", ACM Transactions on Information Systems, Vo.26, No.3, ACM Press, June 2008
- [6] Stefano Mizzaro, Elena Nazzi, and Luca Vassena, "Collaborative Annotation for Context-Aware Retrieval", Proceedings of the WSDM '09 Workshop on Exploiting Semantic Annotations in Information Retrieval, pp.42-45, 2009
- [7] A. Schmidt, "Ubiquitous Computing-Computing in Context", PhD thesis, Lancaster University, 2003
- [8] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", <<http://infolab.stanford.edu/~backrub/pageranksub.ps>>
- [9] O'Reilly, T., "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>>.
- [10] Massimo Melucci, "Vectors, Planes and Context", PersDL 2007
- [11] Gareth J. F. Jones, and Peter J. Brown, "Context-aware retrieval for ubiquitous computing environments", Lecture Notes in Computer Science, Vol. 2954, 2004.
- [12] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze, "Introduction to Information Retrieval", Cambridge University Press. 2008.
- [13] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme, "Information Retrieval in Folksonomies: Search and Ranking", Proceedings of the 3rd European Semantic Web Conference. Budva, Montenegro, pp.411-426, 2006.
- [14] Sungrim Kim, and Joonhee Kwon, "Folksonomy-Based Information Retrieval in Context-aware Environment", International Journal of Computer Science and Network Security, Vol. 8, No. 11, pp. 252-257, Nov. 2008.
- [15] Joonhee Kwon, and Sungrim Kim, "Context aware Intelligent Service based on Folksonomies", Proceedings of the IADIS International Conference e-Society 2009, pp.181~184, Barcelona, Spain, Feb. 2009.
- [16] Evaggelia Pitoura, and George Samaras, "Data Mangement for Mobile Computing", Kluwer Academic Publishers, 1998.
- [17] Amit Thawani, Srividya Gopalan, V. Sridhar1, and Krithi Ramamritham, "Context-aware Timely Information Delivery in Mobile Environments", The Computer Journal, Vo. 50, No.4, pp.460-472, 2007
- [18] David Vallet, Miriam Fernández, Pablo Castells, Phivos Mylonas, and Yannis Avrithis, "Personalized Information Retrieval in Context", 21st National Conference on Artificial Intelligence - 3rd International Workshop on Modeling and Retrieval of Context, Boston, USA, 16-17 July 2006



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 University, 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 Computer Science Kyonggi University, Suwon-si, Kyonggi-do, Korea, Her research interests are ubiquitous computing, LBS, spatial database, GIS.