Activity-Based Query Refinement for Context-Aware Information Retrieval

Shun Hattori, Taro Tezuka, and Katsumi Tanaka

Department of Social Informatics, Graduate School of Informatics, Kyoto University
Yoshida-Honmachi, Sakyo-ku, Kyoto 606-8501, Japan
Tel.: +81-75-753-5385; Fax: +81-75-753-4957
{hattori, tezuka, tanaka}@dl.kuis.kyoto-u.ac.jp

Abstract. Mobile Web search will gain more importance. This paper proposes a novel method for query refinement based on real-world contexts of a mobile user, such as his/her current geographic location and the typical activities at the location which are extracted by Blog mining. Our method enhances location-awareness and even further context-awareness to the existing location-free keyword-based Web search engines.

1 Introduction

With the exponentially growing amount of information available on the Web and the advances in wireless and mobile computing environments, we have been able to access information anywhere at any time in our daily life. Mobile Web search engines will increase their significance in the future. It is remarkably crucial to refine better the retrieval results in mobile computing environments, because mobile devices have poorer I/O user interfaces and we have little time for browsing information slowly while moving or doing some activities in the real world. However, mobile users' original queries are often so short and ambiguous that mobile Web search engines cannot guess their information demands accurately and thus would present too many and too noisy retrieval results.

There exist various approaches proposed for the so-called "mismatched query problem" in IR (Information Retrieval): query modification such as query expansion [1], query relaxation or query substitutions [2], automatic classification and visualization of the retrieval results by clustering, and written or spoken natural language querying [3]. Our work also relates very much to the research fields of Mobile IR [4, 5] and Context-Aware IR [6, 7, 8] for mobile and ubiquitous/pervasive computing environments.

In this paper, we propose a novel method for query refinement based on real-world contexts of a mobile user, such as his/her current geographic location and the typical activities at the location which are extracted by Weblog mining techniques. For instance, when a mobile user in a bookstore issues ["da vinci"] as an original query, our system would infer from his/her original query and current place-name as a real-world context that he/she is requesting information about not a movie but a book of "da vinci code", and offer or retrieve automatically by ["da vinci code" AND "book" AND "buy"] as one of its alternatives.

S. Sugimoto et al. (Eds.): ICADL 2006, LNCS 4312, pp. 474–477, 2006.

2 Query Refinement Based on Real-World Contexts

Our proposed method for refining a mobile user's original query based on real-world contexts, that consist of (1) the place-names which are obtained from his/her current geographic location and (2) the typical activities at the location which are extracted by Weblog mining techniques, executes the following flows.

Step 0. Query Inputted by Mobile User:
A mobile user at certain location issues his/her original query q.

Step 1. Sensing Real-World Contexts:
Obtain the user's current geographical coordinates (latitude and longitude or altitude) by receiving GPS or the user's current place-names by reading real-world embedded RFID tags linking location information.

Step 2. Translating Contexts into Contextual Words:
Convert the above real-world contexts into "contextual words" that are place-names by using a GIS with data from residential maps and the typical activities (object-names and action-names) at the location of each place-name by mining blog data [9].

Step 3. Assigning Weight to Contextual Word:
Assign to each contextual word c_i such as names of place, object and action, the weight $W(c_i; q)$ that evaluates its usefulness for expanding the query q. We will describe "context weighting" in detail in the last half of this section.

Step 4. Enforcing Query Refinement:
Generate alternative queries that consist of the original query and the contextual word with higher weight and let the user select from among them, or presents the results retrieved by the alternative query with highest weight.

The advance of sensor technologies such as GPS or RFID allows mobile users to obtain their numerous real-world contexts. All of those information are, however, not always useful for query refinement to match their information demands. Therefore, it is very necessary to assign some sort of weight based on their original query to each real-world context, in order to classify whether or not it is useful for query refinement. We define the importance $W(c; q)$ of a contextual word c with regard to a user's original query q as follows. It is defined based on the proportion of the local probability $Pr(c|q) = \frac{DF(c \wedge q)}{DF(q)}$ of the contextual word c in the retrieval documents by the user's original query q to the global probability $Pr(c) = \frac{DF(c)}{N}$ of the contextual word c in the whole documents of the target corpus,

$$W(c; q) = \frac{Pr(c|q)}{Pr(c)} = \frac{DF(c \wedge q)}{DF(c)} \cdot \frac{N}{DF(q)} \simeq \frac{DF(c \wedge q)}{DF(c)} \cdot \alpha, \qquad (1)$$

where $DF(q)$ stands for the number of searched documents by submitting a query q to a search engine, N stands for the total number of documents in the corpus of the search engine, and α stands for the certain constant value when given the query. In this paper, we calculate each importance $W(c_i; q)$ where $\alpha = 1$ by using Google Blog Search [10] as the corpus.

3 Experimental Results

We have carried out some experiments, in order to justify that our method can improve the retrieval results by refining an original query by its related contextual words such as names of place, object and action. Table 1 shows each weight $W(c;q)$ of contextual word to "da vinci code" as a mobile user's original query. Because "da vinci code" has some subtopics such as "book" and "movie", we expected that the top 4 names of each subtable for Places, Objects and Actions have higher weight. Table 1 shows almost the same results as our expectation with respect to Places and Objects, but has some undesirable results with respect to Actions. Table 2 shows that our method can improve the approximate precision in the top 20 web pages googled [11] by adding each contextual word such as name of place, object or action to "da vinci code" as an original query. For mobile users, precision is generally more important than recall.

Table 1. Weight of Place/Object/Action-Name c to "da vinci code" as User Query q

Place	W(c;q)	Object	W(c;q)	Action	W(c;q)
"bookstore"	(3) 0.00620	"book"	(4) 0.00940	"buy"	(7) 0.00116
"library"	(6) 0.00246	"novel"	(1) 0.03808	"order"	(3) 0.00156
"movie theater"	(1) 0.01059	"movie"	(3) 0.01163	"read"	(1) 0.00347
"theater"	(5) 0.00462	"film"	(2) 0.01429	"see"	(2) 0.00175
"church"	(2) 0.00891	"dvd"	(6) 0.00234	"climb"	(10) 0.00105
"hotel"	(10) 0.00094	"comic"	(5) 0.00425	"drive"	(8) 0.00109
"museum"	(4) 0.00516	"magazine"	(7) 0.00207	"learn"	(9) 0.00108
"restaurant"	(8) 0.00145	"game"	(8) 0.00156	"sell"	(6) 0.00126
"station"	(7) 0.00157	"drug"	(10) 0.00090	"swim"	(5) 0.00136
"university"	(9) 0.00142	"food"	(9) 0.00147	"walk"	(4) 0.00151

Table 2. Approximate Precision of the Top 20 Web Pages by Refining Query

Original/Refined Query	book (dvc)	movie (dvc)	both	others
"da vinci code"	0.50 (0.50)	0.30 (0.30)	0.05	0.15
"da vinci code" & "bookstore"	0.75 (0.70)	0.15 (0.15)	0.00	0.10
"da vinci code" & "book"	0.60 (0.55)	0.05 (0.05)	0.05	0.30
"da vinci code" & "novel"	**0.85 (0.80)**	0.00 (0.00)	0.05	0.10
"da vinci code" & "buy"	0.55 (0.50)	0.15 (0.15)	0.00	0.30
"da vinci code" & "read"	0.50 (0.45)	0.10 (0.10)	0.00	0.40
"da vinci code" & "theater"	0.00 (0.00)	0.65 (0.65)	0.10	0.25
"da vinci code" & "movie"	0.05 (0.05)	**0.80 (0.80)**	0.05	0.10
"da vinci code" & "film"	0.05 (0.05)	0.60 (0.60)	0.00	0.35
"da vinci code" & "see"	0.10 (0.10)	0.45 (0.45)	0.00	0.45

The number followed by '(' is the precision of web pages about the books or movies of "Leonardo da Vinci", and the number placed between '(' and ')' is the precision of web pages about the book or movies of "The Da Vinci Code".

4 Conclusion and Future Work

In this paper, we proposed a method for query refinement based on real-world contexts of a mobile user, such as his/her current geographic location and the typical activities at the location which are extracted by Weblog mining techniques, aiming to enhance location-awareness and even further context-awareness to the existing location-free keyword-based Web search engines. In the near future, we plan to develop and evaluate a prototype system based on our method, and then we would like to challenge to utilize not only current real-world contexts but also history of continuous past ones and/or prospective ones.

Acknowledgments. This work was supported in part by the Japanese MEXT the 21st Century COE (Center of Excellence) Program "Informatics Research Center for Development of Knowledge Society Infrastructure" (Leader: Katsumi Tanaka, 2002–2006), and the Japanese MEXT Grant-in-Aid for Scientific Research on Priority Areas: "Cyber Infrastructure for the Information-explosion Era", Planning Research: "Contents Fusion and Seamless Search for Information Explosion" (Project Leader: Katsumi Tanaka, A01-00-02, Grant#: 18049041).

References

1. Xu, J., Croft, W.B.: Query Expansion Using Local and Global Document Analysis. In: Proc. of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. (1996) 4–11
2. Jones, R., Rey, B., Madani, O., Greiner, W.: Generating Query Substitutions. In: Proc. of the 15th International Conference on WWW. (2006) 387–396
3. Chang, E., Seide, F., Meng, H.M., Chen, Z., Shi, Y., Li, Y.C.: A System for Spoken Query Information Retrieval on Mobile Devices. IEEE Transactions on Speech and Audio Processing **10** (2002) 531–541
4. Kawai, H., Akamine, S., Kida, K., Matsuda, K., Fukushima, T.: Development and Evaluation of the WithAir Mobile Search Engine. In: Proc. of the 11th International Conference on World Wide Web (WWW'02), Poster-ID:102. (2002)
5. Jose, J.M., Downes, S.: Evaluation of Mobile Information Retrieval Strategies. In: Proc. of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries. (2005) 411
6. Jones, G.J.F., Brown, P.J.: Context Aware Retrieval for Ubiquitous Computing Environments. In: Proc. of the Mobile HCI 2003 International Workshop on Mobile and Ubiquitous Information Access, LNCS Vol.2954. (2004) 227–243
7. Yau, S.S., Liu, H., Huang, D., Yao, Y.: Situation-Aware Personalized Information Retrieval for Mobile Internet. In: Proc. of the 27th Annual International Computer Software and Applications Conference (COMPSAC'03). (2003) 638–645
8. Rhodes, B.: Using Physical Context for Just-in-Time Information Retrieval. IEEE Transactions on Computers **52** (2003) 1011–1014
9. Kurashima, T., Tezuka, T., Tanaka, K.: Blog Map of Experiences: Extracting and Geographically Mapping Visitor Experiences from Urban Blogs. In: Proc. of the 6th International Conference on WISE, LNCS Vol.3806. (2005) 496–503
10. Google Blog Search: `http://blogsearch.google.com/`.
11. Google: `http://www.google.com/`.