Keyword-based Semantic Retrieval System using Location Information in a Mobile Environment

Tae-Hoon Lee, Jung-Hyun Kim, Hyeong-Joon Kwon and Kwang-Seok Hong
Sungkyunkwan University/School of Information and Communication Engineering, Suwon, South Korea
Email: {lthysk, kjh0328, katsyuki}@skku.edu, kshong@skku.ac.kr

Abstract—Many services for mobile devices and wireless network have been studied in a Ubiquitous Environment. Mobile information retrieval is crucial to share the vast information and multimedia content. Mobile information retrieval has three challenges; 1) Mobile devices have a small display area. Thus, users are more interested in precision than in recall. 2) Mobile device portability means users frequently change their interest with location. 3) Though wireless development is progressing; its response rate and mobile device processing ability are still lower than for a wired network and PC. To meet these challenges, we propose keyword-based semantic mobile search system using ontology and location information. Through the experiment, we prove that providing the user’s selection of the semantic result improved precision. Current location information can be used to provide location-based multimedia content using GPS module and map with mash-up service. Finally, the response rate of the service was demonstrated to be faster through a java-based Lucene engine for keyword-based retrieval module and dividing the result of retrieval.

Index Terms—Mobile Information Retrieval, Semantic Information Retrieval, Location-based Service

I. INTRODUCTION

Demand for mobile devices continually increases in the ubiquitous generation. The development of ICT (Information Communication Technology) improves ubiquitous computing environments (i.e. wireless network and mobile device computing). Accordingly, many services for mobile devices using wireless network have been studied [1–3]. People will gather information from the physical world and upload it to the service server. This is unlimited in time and space using wireless networked mobile devices. Therefore, we can readily expect information overload in mobile environment.

Mobile information retrieval is important to share the vast information and multimedia content. Mobile information retrieval differs from existing non-mobile information retrieval. The features of mobile device and ubiquitous computing environment confer many challenges. Three challenges are:

- The mobile device has a small display area. It can be inadequate displaying a full list of relevant items. Therefore, in the mobile search environment, the users are more interested in precision than in recall [4].
- The locality where the search for information is focused may continually change due to the portability of mobile devices. Thus, the user’s interest is frequently changed as the user moves to a new location [5].
- Fast response is important for user satisfaction. Although there is rapid progress in research and development of wireless networking and communication technologies [6], the speed and mobile device processing ability remain lower than for a wired network and PC.

To meet these challenges, we propose the following:

- To improve precision, the result of keyword-based retrieval and semantic-based retrieval using ontology from the two modules operate simultaneously. The user manually selects the correct result, so he/she can check the desired information.
- The coordinates of the current location received using the GPS module are mapped to the address. This enables the retrieval of multimedia content including address information. They can check the registered content near the current location using the map with a mash-up service.
- We imported the java-based Lucene engine for keyword-based retrieval module to improve the service response rate. We divide the result of retrieval (XML document) using the WiBro network environment for fast network access speed.

The paper is structured as follows: Section 2 outlines related work. In Section 3, we describe the entire system, keyword-based semantic retrieval module and location-based retrieval module and mash-up service. In Section 4, we demonstrate implementation and experimental results. The last section contains concluding remarks and future directions.

II. RELATED WORKS

A. Mobile Information Retrieval

P. Coppola [4] divides information retrieval into e-relevance and w-relevance. He used the term “relevance” to denote information relevant to the user. E-relevance denotes classical relevance, i.e. the non-mobile information retrieval case. Conversely, w-relevance concerns mobile information retrieval. He proposed four dimensions – information resources, representation of the
user problem, time, and components - that can be used in the implementation and evaluation of information retrieval systems. The two relevances differ, because e-relevance is in the “information world”, whereas w-relevance is in the “real/physical world”.

B. Semantic Search

The search is divided generally in two; navigational search and research search [7]. They differ with the type of the query; navigational search requires the exact word or sentence the user wants to find. For research search, the user inputs the query that explains or describes information related to that the user wishes to collect or research. Then, research search differs from navigational search, because the user does not know about the documentation before searching. Thus, the user tries to find the related document by getting information about the place of the document, as the result of navigational search. Semantic-based search belongs to research search; it is formulated to improve existing keyword-based web search. The query user inputs indicate one (or two) real world concepts. This helps understand the user’s expected categorization of the search result. The result of semantic search is independent of the result of the keyword-based search, because it is based on semantic web. Therefore, it can increase the quality of the result, as well as the quantity.

III. KEYWORD-BASED SEMANTIC RETRIEVAL SYSTEM USING LOCATION INFORMATION IN A MOBILE ENVIRONMENT

A. System Architecture

Fig. 1 shows the system architecture of proposed system. The system is largely divided into content creation and content sharing. At the creation level, the user creates and transmits multimedia contents based on their own five senses using a mobile device. The user employs the mobile device and external devices to discover the five senses. The user transmits the created multimedia content to the web server through the wireless TCP/IP socket after commenting on the content, and receives the coordinates of the current location through the GPS. At the sharing level, the user can share the five sense-based multimedia content created and ontology information using the keyword-based semantic retrieval module. First, the user creates and sends the query to find the information or content. The server receives the query and resends it to the keyword-based module and semantic-based module respectively. Each module executes the searches, after analyzing the user query. The system integrates the search result and creates the structured XML document. The user checks the list resulting from the search result by accessing this XML document. The user selects the content and displays it if s/he finds what they wanted.

B. Keyword-based Semantic Retrieval Module

We designed and implemented the keyword-based semantic retrieval module using the five senses multimedia ontology. Fig. 2 depicts the integrated block-diagram of the keyword-based retrieval interface and semantic-based retrieval interface. The keyword-based query the user inputted retrieves using the two interfaces; keyword-based retrieval interface and semantic-based retrieval interface.

Keywords are used generally in retrieving items from a vast information source, for instance a catalog or a search engine. An index term or descriptor in information
retrieval is a term that captures the essence of the topic of a document, and can consist of a word, phrase, or alphanumerical term. They are created by analyzing the document either manually with subject indexing or automatically with automatic indexing. Methods that are more sophisticated use keyword extraction. Keywords are stored in a search index. In this paper, we adopt the Apache Lucene library [8], a high-performance, full-featured text search engine library written entirely in Java. Lucene has a very flexible and powerful search capability that supports a wide array of possible searches including AND OR and NOT, fuzzy logic searches, proximity searches, wildcard searches, and range searches to locate indexed items. The technology is suitable for nearly any application that requires full-text search, especially cross-platform. The first thing it does is create an IndexSearcher object pointing to the directory where the contents have been indexed by title, location, user’s annotation and tags, and then creates a StandardAnalyzer object. The StandardAnalyzer is passed to the constructor of a QueryParser along with the name of the default field to use for the search. This will be the default field, if the user does not specify a field in the search criteria. It then parses the actual search criterion specified, giving us a Query object. We can run the Query against the IndexSearcher object. This returns a Hits object, a collection of all the content that met the specified criteria.

The semantic-based retrieval module is depicted in Fig. 3. It uses five sense multimedia ontology. It reflects real world information connected to location information by a semantic relation. For example, restaurant, hotel, and school information in Washington DC. In this case, each instance (restaurant, hotel, or school) has a semantic relationship with the address (Washington DC). Each class has a semantic relationship to each other as well as the address. We use the semantic relationship between classes in the ontology for semantic-based retrieval. The implemented semantic-based retrieval interface includes three major steps referred to in [9]; 1) Term mapping maps between ontology resource and the input query language; 2) Query graph construction uses relations between the mapped ontology resources in step 1; and 3) SPARQL(SPARQL Protocol and RDF Query Language) query conversion using the query graph created in step 2.

After the three steps, the constructed SPARQL query is used to search the knowledge in the five senses ontology, and form the XML document from the search result. Term mapping is the process of connection between the input query and the ontology resource. The query is split by the space, and each token is mapped when the token is the same as the name or label of the class or instance in the ontology. The mapped ontology resource constructs all possible query graphs using the ‘Minimum Spanning Tree’ algorithm, then a spanning tree with weight less than or equal to the weight of every other spanning tree. The relation between the constructed query graphs is checked to expand the query graph using the ontology’s schema. Schema define the relation between classes, then if the query user input is a instance, the class that belongs to the instance and the higher classes are used to check the relations. After checking the relation, both graphs in the relation link to the relation’s direction for expansion. For example, Fig. 4 shows an example of the query graphs generation and expansion. and indicate instance variant x and y belong to class 1 and class 2 respectively. The system automatically finds the that indicates the relation between variant x and y. is the result of combining the three conditions.

The system translates final query graph to SPARQL language for semantic-based retrieval. SPARQL is a query language based on graph pattern, so the translation from query graph to SPARQL is basic. Table 1 shows an example of SPARQL translated by the graph in Fig. 4.

<table>
<thead>
<tr>
<th>Num</th>
<th>Command</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>PREFIX rdf: <a href="http://www.w3.org/1999/02/22-rdf-syntax-ns#">http://www.w3.org/1999/02/22-rdf-syntax-ns#</a></td>
</tr>
<tr>
<td>2</td>
<td>PREFIX : <a href="http://www.owl-ntologies.com/Ontology1215649820.owl#">http://www.owl-ntologies.com/Ontology1215649820.owl#</a></td>
</tr>
<tr>
<td>3</td>
<td>SELECT ?x</td>
</tr>
<tr>
<td>4</td>
<td>WHERE (?x rdf:type :Class 1&gt;</td>
</tr>
<tr>
<td>5</td>
<td>?y rdf:type :Class 2.</td>
</tr>
</tbody>
</table>

Figure 3. Semantic-based retrieval module

Figure 4. Example query graph generation and expansion

TABLE I. AN EXAMPLE OF SPARQL
C. Location-based Retrieval Module and Mash-up Service

Two or more geographic maps are produced with the same geographic parameters and output size. The results can be accurately overlaid to produce a composite map. This study designed and implemented the mobile station with a GPS unit-based mixed-Web map interface using a location-based mobile mash-up with Google and Yahoo maps. This application is an essential component technology that enables the design and realization of the LBS-based intelligent agent including a user-centric automatic five senses multimedia recommender interface, mobile social networking-based user detection interface, and LBS-based five senses multimedia content retrieval and representation module.

Fig. 5 shows the block-diagram of the location-based retrieval module. The user obtains current coordinates using the GPS module. The location mapping module maps the coordinates to the real address we called. The obtained address is used to query. Therefore, the user can select whether or not to use the address as the query. The proposed mobile station-based mixed-map interface is designed using 2- and 3-dimensional geographical position information provided by Google and Yahoo maps. It controls the map data independently in the location server side.

IV. EXPERIMENTAL RESULTS

A. Experimental Conditions

The proposed system is implemented in Microsoft Visual Studio 2008-supported c# using the Microsoft.Net Compact Framework 3.5 and Windows Mobile 5.0 Pocket PC SDK.

Experimental environments consisted of a blue-tooth module for data communication between the mobile station and GPS unit, and WiBro-based mobile stations for wireless mobile networking. Three WiBro mobile stations with GPS units are used in the experiments: MSM 6500(EVDO) and 520MHz CPU-based SAMSUNG SPH-M8200. This is a PDA type WiBro station that runs Window Mobile 5.0 (Pocket PC 2005) OS and supports WiBro and CDMA EV-DO connectivity. The bar type device is 16.6mm thick and has enhanced usability with a full touch screen (2.8’’). The user can create five senses based multimedia content using the mobile device’s embedded camera and recorder and additional external devices with the blue-tooth module. We adopted the KT (Korea Telecom) WiBro service. It is able to use a broadband Internet service whilst moving at a speed of up to 120 km/h and has a faster rate than HSDPA (High Speed Downlink Packet Access).

B. Experimental Results

1) Example of Sematic Search using Location:

Fig. 6 shows an example of mobile semantic search. Fig. 6 (a) shows the initial screen of the proposed system. The user can insert the query to find the desired information. In the example, the query “chunchun-dong” is entered automatically by the obtained location using GPS module. And user inputs the query ‘school’ additionally using a stylus pen. ‘chunchun-dong’ means Korean style address. ‘dong’ means a village and ‘chunchun’ is the name of a place. Thus, it indicates the user wishes to know about the ‘school’ in the ‘chunchun-dong’. The result of the inputted query is shown in Fig. 6 (b). ‘SungKyunKwan University’ is the only result retrieved. It is semantic-based retrieval result. The keyword search result is empty, because the current DB content does not include ‘chunchun-dong’ or ‘university’. If the user wants to know about ‘SungKyunKwan University’, s/he just clicks that. Thus, Fig. 6 (c) shows the result of the query ‘SungKyunKwan University’. If the user clicks ‘show map’, s/he can check the content near the current location. This is depicted in Fig. 6 (d).

2) Information Retrieval Test: We used 50 keyword

Figure 5. Location-based retrieval module block-diagram

Figure 6. An example of simulation; (a) the level of entering query (b) the result of semantic search (c) the result of keyword search using the result of semantic search (d) mash-up service with google
queries for the three subjects; sports, restaurant, university. And we calculate the precision and recall of the three methods of retrieval; keyword-based retrieval, semantic based retrieval, user-driven retrieval. User-driven retrieval is integrated retrieval method (i.e. user can choose one between the results of keyword- and semantic-based information retrieval). Table 2 shows the results of information retrieval test. From the experimental result, the precision and recall of semantic-based retrieval is less than keyword-based retrieval. It may be caused by the keyword’s ambiguity. The user-driven retrieval’s precision (88.1%) shows best result though the recall is low (62.9%).

TABLE II. PRECISION AND RECALL IN IR TEST

<table>
<thead>
<tr>
<th>Query</th>
<th>Keyword</th>
<th>Semantic</th>
<th>User-Driven</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sports</td>
<td>P (%) 71.4</td>
<td>35.8</td>
<td>82.5</td>
</tr>
<tr>
<td></td>
<td>R (%) 68.9</td>
<td>56.8</td>
<td>64.8</td>
</tr>
<tr>
<td>Restaurant</td>
<td>P (%) 82.5</td>
<td>45.8</td>
<td>91.6</td>
</tr>
<tr>
<td></td>
<td>R (%) 72.8</td>
<td>72.6</td>
<td>67.4</td>
</tr>
<tr>
<td>University</td>
<td>P (%) 84.7</td>
<td>30.2</td>
<td>90.2</td>
</tr>
<tr>
<td></td>
<td>R (%) 67.8</td>
<td>70.6</td>
<td>56.5</td>
</tr>
<tr>
<td>Average</td>
<td>P (%) 79.5</td>
<td>37.3</td>
<td>88.1</td>
</tr>
<tr>
<td></td>
<td>R (%) 69.8</td>
<td>66.7</td>
<td>62.9</td>
</tr>
</tbody>
</table>

3) Speed of Information Retrieval: We calculate the speed of information retrieval by the mobile system’s stopwatch function in the WiBro environment. It is from the request time of information retrieval (i.e. click the send button after enter a query) to the receiving time of results. We use the user-driven information retrieval method with 114,857 contents. And the process of keyword-based information retrieval is divided in two; non-using and using Lucene engine. Non-using Lucene engine substitutes My-SQL query. We ignored the precision of results. The Average speed in Table 3 proves that the inverted indexing structure-based Lucene engine is faster than full text search. And we divide the generated XML file by information retrieval’s result into several pieces with 10 contents. It is not much difference but speed of the divided list is faster than full list.

TABLE III. AVERAGE SPEED OF INFORMATION RETRIEVAL

<table>
<thead>
<tr>
<th></th>
<th>Lucene Engine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Non-Usage</td>
</tr>
<tr>
<td>XML File</td>
<td></td>
</tr>
<tr>
<td>Full List</td>
<td>7.88 sec</td>
</tr>
<tr>
<td>Divided List</td>
<td>6.96 sec</td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

We propose a keyword-based semantic retrieval system using location information in a mobile environment. Through the experiment, we prove that providing the user’s selection of the semantic result improved precision. Current location information can be used to provide location-based multimedia content using the GPS module and map with a mash-up service. Finally, the response rate of the service was faster using the java-based Lucene engine for the keyword-based retrieval module and dividing the result of retrieval. Future work will expand the research to apply other context information including location.

ACKNOWLEDGMENT

This paper was supported by Samsung Research Fund, Sungkyunkwan University, 2008 and the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MEST) (No. 20090058909).

REFERENCES