Ontology-Based Query Formation and Information Retrieval

Sheng-Uei Guan
National University of Singapore, Singapore

INTRODUCTION

With the introduction of new technologies such as WAP, HSCSD, GPRS, UMTS, and Bluetooth, it is believed that the e-commerce arena will sooner or later merge its applications with handheld devices to create more opportunities for the birth of mobile commerce (m-commerce). However, m-commerce is largely unrealized to date because there still does not exist a single killer application that can attract wireless users to use wireless services. According to a recent survey by Gartner, Inc., besides the importance of coverage of wireless network and pricing issues, the wireless Internet and data services is the next crucial factor that attracts users to use wireless service. As such, there is a need to improve the data services over the wireless network. One of these services is the information retrieval service.

Most electronic product information retrieval systems are still not efficient enough to cater to the increasing needs of customers. This is especially serious in the m-commerce arena, where the bandwidth of mobile devices is low and large data would not be possible. Thus, the discovery of new information retrieval techniques is inevitable.

The main objective of this article is three-fold: 1) to research the use of ontology to assist users in shaping their product enquiries; 2) to study the use of genetic algorithms and agents in query optimization; and 3) to develop information retrieval services for the m-commerce arena. The article proposes a methodology for efficient query formation for product databases and for effective information retrieval systems, which includes the evaluation of retrieved documents to enhance the quality of results that are obtained from product searches.

This article discusses the usage of ontology to create an efficient environment for m-commerce users to form queries. The establishment of a method that combines keyword searches with using ontology to perform query formation tasks further allows a more flexible m-commerce environment for users. Also, with the use of genetic algorithms, it is hoped that query effectiveness can be achieved, at the same time saving computational time.

BACKGROUND

Definition of Ontology

In artificial intelligence, ontology is defined as a design of a conceptualization to be reused across multiple applications (Fensel, 2000; Braga, Werner & Mattosso, 2000; Hendler, 2001). A conceptualization is a set of concepts, relations, objects, and constraints that define a semantic model of some domain of interest. In other words, ontology is like the structure that describes or encodes the conceptualization in any relational aspect (McGuinness, 1998; Karp, 2000).

Literature Survey

In this section, a survey of present query formation methods and information retrieval methods will be discussed.

Unlike in e-commerce, query information using keywords alone in m-commerce is unrealistic, as mobile devices are too small and keypads are not suitable for typing. Moreover, it may be difficult for the user when vocabulary of subject is unfamiliar. Thus, relevance feedback is still the main technique for query modification.

Relevance feedback technique has been investigated for more than 20 years in various information retrieval models, such as the probabilistic model and vector space model (Boughanem, Chrisment & Tamine, 1999; Salton, 1989). It is based on randomly changing the set of query terms, as well as the weights associated with these terms, according to the document retrieved and judged during the initial search.

A lot of research (Boughanem et al., 1999; Yang & Korfhage, 1994; Kraft, Petry, Buckles & Sadasivan, 1994; Kouichi, Taketa & Nunokawa, 1999) has been done on how genetic algorithms (GAs) can be used in information retrieval. One popular approach is query restructuring, which is used to improve the efficiency and effectiveness of the queries formed. GAs actually extend the concepts of relevance feedback. The difference is that genetic algorithms use more than one query and compare the

fitness among these queries. The fittest query will survive in the end. Thus, this article focuses on extending the concepts of using genetic algorithms in query restructuring.

Fitness Functions

There are a number of measures of query fitness used in previous works, namely precision and recall retrieved (Kraft et al., 1994; Salton & McGill, 1983), average search length (Losee, 1991), and average maximum parse length (Losee, 1991).

Precision is the percentage of documents retrieved that are relevant, while *recall* measures the percentage of the relevant documents retrieved (Kraft et al., 1994; Salton & McGill, 1983). These two tend to be inversely proportional, so that one is traded for another in most situations. *Average search length* is the average number of documents or text fragments examined in moving down a ranked list of documents until arriving at the average position of a relevant document (Losee, 1988, 1996). Evaluating the performance of a filtering or retrieval process with average search length provides a single number measure of performance. *Average maximum parse length* is the average (over a set of sentences) of the largest number of terms in a parse for each sentence. There are also measures that combine both average search length and average maximum parse length.

Typically, present methods had only dealt with the relevance of the document retrieved. This is reasonable but inefficient, because it is rather difficult to indicate the relevance of a document when the number of documents could be very large. This article measures the relevance of queries instead of documents retrieved. Based on this, efficiency will be improved significantly as the number of queries will be much smaller than the number of documents retrieved, which is ideal for mobile devices.

The Proposed Approaches

Both keyword- and ontology-based approaches have their advantages and disadvantages. Ontology provides the structure, context, and visual aid, while keyword provides a direct search mechanism. Both approaches are relevant for mobile commerce because they save time in browsing and searching, which is very much required by mobile users who are always on the move. Thus, by combining keyword queries with ontology, it is possible to achieve a better and more effective query formation. Before ontology terms are accessed to form the queries, there will be a keyword search to find the required ontology term. For example, "ps2" can be hidden in the node "mouse" when presented in the ontology. The user will not be able to know where "ps2" can be found intuitively without eyeballing the ontology. With the help of keyword search, the term "ps2" can be found easily.

In forming queries, there can be a high chance that the vocabulary used by the user to describe a query does not exactly match the vocabulary used by a query system (Preece, 1999). This will result in getting insufficient information. Therefore, restructuring dealing with domain ontology relationships might be useful. These relationships involve semantic links such as hyponyms and synonyms (Braga et al., 2000). Here, using synonyms is an adequate option to restructure queries because it correctly broadens the scope of search even to the extent of different languages.

When too little information is retrieved, the use of synonym or hyponym might be necessary in order to relax the constraints of the query. However, this approach has a major disadvantage. By relaxing the constraints of a query using synonym or hyponym to increase the number of documents retrieved, one could actually deface the meaning of the original query such that it could drift away from the user's intention. This concern can be alleviated by having user feedback along the process. Also, we have considered relaxing constraints step-by-step. This option can better eliminate the chances of constructing far-fetched queries from the use of genetic algorithms.

MAIN THRUST OF THE ARTICLE

Prototype Design and Implementation

Query Formation Using Ontology

Query formation will be done with the aid of tree ontology. Following the tree path will help form the requirements of a query. This allows forming a query easily. An illustration of the query formation process is shown in Figure. 1. As can be seen from this illustration, using ontology helps the user to save several steps by forming a query using the ontology path that is selected. Thus, it can be claimed that forming queries using ontology is actually more efficient than using keywords.

Combining Keywords and Ontology

The design of parallel combination is rather straightforward. Ontology does not cover everything. Thus, besides having ontology for the user to click on when forming a query, there should be some fields present for the user to fill in. When these fields are being filled in, they can replace the use of ontology either partially or completely. For a serial combination, keywords are used to look for