Ontology-Based Query Formation and Information Retrieval

Sheng-Uei Guan
National University of Singapore, Singapore

BACKGROUND

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 the Gartner, Inc. (Wong, 2005), 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 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.

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 algorithm, it is hoped that query effectiveness can be achieved, at the same time saving computational time.

Definition of Ontology

In artificial intelligence, ontology is defined as a design of a conceptualization to be reused across multiple applications (Bailin, 2004; Braga, Werner, & Mattosso, 2000; Fensel, 2000; Hendler, 2001; Riza & Oguz, 2002). 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 (Ambrosio, De Santos, De Lucena, & De Silva, 2004; Karp, 2000; McGuinness, 1998; Sugumaran & Storey, 2003).

Chandrashekaran and Josephson (1999) described ontologies as the conceptualization that underlies knowledge, without which there will not be a vocabulary for representing knowledge. Besides modeling knowledge, the ontology is also a structure that can be made publicly available, thereby enabling knowledge sharing.

According to Howarth (2004), in agent technology, ontology is especially useful, as agents are able to learn the behavioral patterns of constraints and resources of business networks. These intelligent agents are then able to store their knowledge into an ontological structure which, besides storage, enables the observation of the interaction of the resources. Ontology, therefore integrates intelligence into the nodes of the network, providing a context in which intelligent decisions are made.

Literature Review

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.

In genetic algorithm, much research (Boughanem et al., 1999; Guan & Zhu, 2004; Kraft, Petry, Buckles, & Sadasivan, 1994; Kouichi, Taketa, & Nunokawa, 1999; Yang & Korfhage, 1994) has been done on how it (GA) 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. GA actually extends the concepts of relevance feedback. The difference is that genetic algorithm uses more than one query and compares 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, whereas *recall* measures the percentage of the relevant documents retrieved (Kraft et al., 1994; Salton & McGill, 1983). These two terms tend to be inversely proportional so that one is traded for one 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 chapter 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 objective of this article is threefold: (a) to research the use of ontology to assist the users in shaping up their product enquiries; (b) to study the use of genetic algorithms and agents in query optimization, and (c) to develop information retrieval services for the m-commerce arena. This 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.

DESCRIPTION OF ONTOLOGY-BASED QUERY REFINEMENT APPROACH

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 and thus allow easy forming of a query. 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 are actually more efficient than using keywords.

Figure 1. Illustration of using ontology to form queries