An Overview of the Agent-Based Electronic Commerce System (ABECOS) Project

Ee-Peng Lim, Wee-Keong Ng
Center for Advanced Information Systems
School of Applied Science
Nanyang Technological University
Nanyang Avenue, Singapore 639798, SINGAPORE

Abstract

ABECOS (Agent-Based Electronic COmmerce System) is an ongoing project that investigates the problems involved in building a large-scale business-to-consumer e-commerce system on the web using agent technologies. In this paper, we describe the ABECOS's system architecture and some research problems studied by the ABECOS research group. Our approaches to apply database techniques to address some of these E-Commerce problems will also be presented.

1 Introduction

In the 1998 NSF Workshop on Research Priorities in Electronic Commerce[4], e-commerce has been defined as "sharing business information, maintaining business relationships, and conducting business transactions by means of telecommunication network." Under this broad definition, e-commerce has further been classified into *business-to-business* and *business-to-consumer*. As pointed out by Ozsu in [10], business-to-business e-commerce activities started as early as in the 1960's in the form of Electronic Data Interchange (EDI) although the telecommunication network adopted by the EDI systems were not Internet-based. In the case of Internet-based EDI [3], the main research problems include the interoperability between different EDI systems and the efficient and secure transmission of EDI messages on the Internet.

The business-to-consumer e-commerce, in contrast, has only started to pick up recently. Nevertheless, within a short span of time, it has created many new ways of online retailing on the Internet and has contributed to the successes of many new ".com" businesses. For example, Amazon.com has now become one of the world's largest booksellers, boasted with a customer base of 12 millions[12]. Yahoo.com, besides providing a very successful web search engine, has been hosting 6000 over online stores selling more than 3.75 million different products. Despite the huge successes of these ".com" businesses, many issues in the business-to-consumer e-commerce remain unresolved and require the attention of database researchers.

In the ongoing ABECOS (Agent Based Electronic COmmerce System) project[1] at the Center for Advanced Information Systems (CAIS), we are focusing on research issues related to building a large-scale e-commerce system that facilitates Internet shopping using different kinds of agents[13]. ABECOS aims to support a distributed

Copyright 2000 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.
Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

[1] See *http://www.cais.ntu.edu.sg:8000/∼abecos/*

marketplace where large number of different products offered by different online stores can be purchased by web users.

The ABECOS system architecture essentially consists of three types of software agents, i.e. **seller agents**, **buyer agents** and **directory agents**. Each buyer agent represents a web user and performs the appropriate tasks to purchase some product(s) specified by the user. In its operations, the buyer agent will emulate some intelligent behaviors of a typical human shopper in its attempt to find the "best" deal for the user. Here, the measure of a deal could be determined by the price of the product(s) or the some other criteria specified by the user. To buy a product from an online store, a buyer agent has to communicate with the seller agent representing the store. The seller agent has direct access to the product database of its online store and is capable of committing transactions on behalf of the store. The buyer agent may negotiate with the seller agents of multiple stores before a final purchase decision is reached. To allow buyer agents to locate seller agents of the appropriate online stores, the ABECOS system architecture has incorporated directory agents that maintain catalogs of product information in a summarized form and are capable of recommending the seller agents of appropriate online stores to the buyer agents based on product specifications provided by the users[9].

In the remaining sections of this paper, we will highlight some database research issues in the context of ABECOS project.

2 XML Search Capabilities

XML enables online stores to provide their product information in a self-describing and platform-independent manner to the web users. By marking web documents with well-formed user-defined tags, online stores can easily structure the document content so that remote e-commerce applications can extract the embedded data for their use[8]. When XML documents are given to a web browser together with their XSL style sheets(XML Style Sheet Language), their XML content will actually be presented in HTML format within the browser. Hence, XML is able to support both the browsing and interoperability needs of e-commerce applications.

In ABECOS, we adopt an XML approach to store both product catalog and online store catalog information at the directory agent's web site. Products are first classified into product categories and sub-categories forming one or more product hierarchies (also known as the *isa* hierarchies). The leaf nodes of the product hierarchies are XML pages of individual products while the internal nodes are XML pages containing directories of products or product categories. These XML pages are connected by XML links. The product hierarchies greatly facilitate web users in navigating the product catalog XML pages and formulating search queries on only products under some specific category. Beside the *isa* relationship between product categories and products, there also exists some other mundane or peculiar inter-relationships between products or product categories. For example, a personal computer consists of CPU, disk drives, and others as parts. A special mouse device can only be used with a special keyboard. All product information, product category information and the different types of relationships between them can be readily represented in XML. Similarly, we can maintain online store catalog information using XML. In the following, we describe our research efforts in developing a search engine for storing and indexing these XML information.

At present, our research in XML search capabilities focuses on providing **structured queries** on text data embedded in XML pages[5]. Here, structured queries are ones that involve **structure patterns** within the XML pages or across them. For example, a web user may want to look for an IBM personal computer consisting of an energy saving monitor and an ergonomic keyboard. Such requirement can be translated into a structured search on the "brand" element within "personal computer" pages that are linked to "monitor" pages containing the terms "energy" and "saving" in their "feature" elements, and "keyboard" pages containing the term "ergonomic" in their "feature" elements. This query can be represented as a query graph shown in Figure 1.

To efficiently evaluate the above type of queries on a collection of XML pages, we have developed a suite of indexes consisting of a **tag inverted index** and a **term inverted index**. The former allows us to locate the

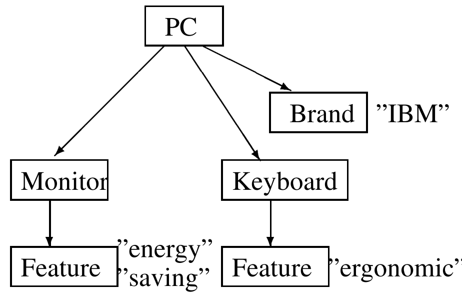

Figure 1: A Structured Query Example

start and end positions of tags within XML pages. The latter identifies the positions of terms within XML pages. Nevertheless, the indexes alone do not directly lead to efficient query evaluation if the structure patterns used in queries are not properly considered. Hence, coupled with these indexes, we have also developed an **incremental greedy** algorithm to decide the order of evaluating structure patterns in the queries. The algorithm consists of multiple iterations and in each iteration the query node expected to return the least result is evaluated. The algorithm terminates when it has evaluated all query nodes in the query graph. A prototype implementation of our proposed XML search engine has been made available on the web at *http://www.cais.ntu.edu.sg:8000/∼wires*.

3 Product Schema Integration

In the ABECOS e-commerce system, buyer agents and seller agents interact and transact on behalf of sellers and buyers while directory agents locate seller agents for buyer agents. Since seller agents distinguish themselves by the products they sell, directory agents use the buyer's specification of the desired product as a query to find relevant seller agents. In this case, directory agents no longer use a simple keyword search strategy adopted by traditional search engines; they use user-supplied product specifications as the search criteria. They compare the specifications with descriptions provided by different seller agents and return the locations of relevant seller agents. Once buyer agents obtain these addresses, they communicate directly with each relevant seller agent. They then query for the details of the products, negotiate prices and complete the deal.

As different sellers use different local product schemas, there is a need to resolve the semantic differences among local schema attributes. This is an ontology heterogeneity problem. Data integration is a solution to this problem. Product schema integration is a sub-field of data integration. In the Asilomar report on database research[1], in J. Gray's brief essay on database research [2] and in another paper on database research in the 21st century by A. Silberschatz *et al.*[11], heterogeneous data integration has been identified as one of the major research directions in the near future. It has also been pointed out that in e-commerce systems, heterogeneous information sources must be integrated and the integration is an extremely difficult problem. Although significant progress has been made in data integration, this problem is by no means solved and *name matching* across distributed sources is still one of the major problems remaining. Therefore, we need schema management facilities such as directory agents in ABECOS which can adapt to the dynamic nature of heterogeneous data. We may re-state the problem as a set of basic questions:

- What is the information needed for identifying the correspondence among attributes from heterogeneous local product schemas? In what form does such information exist?
- What is the algorithm to integrate product schemas? Is it capable of automated integration?
- How can the global schema be maintained?

We refer to this problem as the *Product Schema Integration* problem (PSI)[13, 14].

4 Electronic Brokering

In [7], *product brokering* and *merchant brokering* are two types of **electronic brokering (e-brokering)** to be carried out as part of the buying process. Product brokering refers to helping a web user to determine what product to buy. Merchant brokering involves helping the user to determine who to buy from. At this point, it is not always clear if product brokering always comes before merchant brokering. It is however quite certain that users will always have to start with some selection criteria on some product attributes, e.g. product type, brand, etc..

In our e-brokering research, we model the brokering of products and merchants as a *database selection problem*. In this database selection problem, we are given a set of user-specified selection criteria on some product attribute and we have to select from a large number of product databases from different online stores the ones that are likely to provide the products that satisfy the selection criteria. Once some online stores have been shortlisted by the user, the selection criteria can be forwarded to the shortlisted stores and be actually evaluated against the product databases of the stores. The products satisfying the selection criteria from the selected stores are finally combined and presented to the user for his final buying decision. Based on the above problem formulation, we seek to answer the following research questions in our e-brokering research:

- What should be the summarized knowledge one should acquire from the product databases such as database selection can be accurately performed? In this respect, the database selection technique must consider multiple attributes of different data types in the product databases.
- How do we measure the performance of database selection? The performance measure may possibly be determined by the end users.

5 Web Site Refreshing

As corporations advertise themselves and sell products through their websites, it is necessary to place the product and price information on web pages for consumer to browse. These information generally come from a corporation's product databases. For a large corporation with multiple websites spread across countries and continents, coordinating the refresh of information to websites from product databases is a complex task. There are several complications:

- A website contains information that come from different portions of the product databases. For instance, a single web page alone may contain different types and amount of information from different parts of the databases.
- Websites have different refresh frequencies. Pages in the same host machine may also have different refresh frequencies. Even different parts of the same page have different refresh frequencies.

In order to keep websites up-to-date, an appropriate set of queries must be made to product databases so that query results can be distributed to refresh the correct websites. Queries must be made regularly so that fresh information arrive within the refresh frequencies of websites.

A straightforward approach to refresh websites is to gather all the queries required by each website and execute each of them. However, this method may not work because query results may not arrive within the refresh frequencies of websites due to the lack of resources (processors). To alleviate this problem, we may use more computers and resources. However, this option is viable only if financial resources are available.

We observe a characteristic of information hosted in websites: As websites display information for consumers, care is taken not to overwhelm consumers with large amounts of complex information. Generally, information is distributed across multiple web pages and aesthetically presented. This has two implications:

- Fresh information from several web pages may be obtained by performing a single query to a corporation's product database.
- Likewise, a single query's result may be splitted up and distributed across multiple web pages, thereby satisfying their information requirements.

Given the information refresh requirements of a set of websites, we must find a suitable set of queries to pose to product databases so that *timely* refreshes are made to websites. Clearly, the query set should be minimal. We may re-state this problem as a set of basic questions: Given the information refresh requirements of a set of websites,

- How to determine the feasibility of scheduling a set of queries to satisfy them (i.e., query results arrive within refresh frequencies)?
- If the query set is non-unique, what is the optimal query set? What optimization techniques are there?

We refer to this kind of problem as the Websites Refresh Problem (WRP)[6]. This problem needs not necessarily occur for a large corporation with multiple geographically dispersed websites. It may also happen within a single, complex website containing multiple web pages requiring bits and pieces of information to be refreshed at different intervals from different portions of the corporation's product databases.

6 Concluding Remarks

In this paper, we gave an overall description of the ABECOS project and the database research problems involved. The demand for large-scale business-to-consumer e-commerce system has certainly brought about many new problems that require one or more solution techniques. While solving these research problems, we have started to develop a ABECOS system prototype to demonstrate the capabilities of buyer agents, seller agents and directory agents. To allow agent technology to be widely and quickly deployed by the online stores and web buyers, we are now looking into software toolkits that allow us to create and configure the different types of e-commerce agents.

References

[1] P. Bernstein, M. Brodie, S. Ceri, D. DeWitt, M. Franklin, H. Garcia-Molina, J. Gray, J. Held, J. Hellerstein, H. Jagadish, M. Lesk, D. Maier, J. Naughton, H. Pirahesh, M. Stonebraker, and J. Ullman. The Asilomar Report on Database Research. *ACM SIGMOD Record*, 27(3), December 1998.

[2] J. Gray. Database Systems: A Textbook Case of Research Paying Off. URL=*http://www.cs.washington.edu/homes/lazowska/cra/database.html*.

[3] T. Harding, R. Drummond, and C. Shih. *Internet Draft: Requirements for Inter-operable Internet EDI*. EDI-INT Working Group, URL=*http://www.ietf.org/internet-drafts/draft-ietf-ediint-req-08.txt*, September 1999.

[4] IC^2 Institute. NSF Workshop: Research Priorities in Electronic Commerce. Technical report, The University of Texas at Austin, September 1998.

[5] E-.P. Lim, C-.H. Tan, B-.W. Lim, and W-.K. Ng. Querying Structured Web Resources. In *ACM International Conference on Digital Libraries*, pages 297–298, Pittsburgh, June 1998. ACM Press.

[6] H.-F. Liu, W.-K. Ng, and E.-P. Lim. Some Preliminary Results on the Feasibility of Website Refresh Queries. In *10th International Conference on Database and Expert System Applications (DEXA'99)*, Florence, August 1999.

[7] P. Maes, R.H. Guttman, and A.G. Moukas. Agents That Buy and Sell. *Communications of the ACM*, 42(3):81–91, March 1999.

[8] B. Meltzer and R. Glushko. XML and Electronic Commerce. *SIGMOD Record*, 27(4):21–24, December 1998.

[9] K.-L. Ong, H.-H. Chew, W.-K. Ng, and E.-P. Lim. Calling All Agents, Where Are You? In *International Workshop on Information Technologies for Electronic Commerce (ITeC'99)*, Florence, August 1999.

[10] M.T. Ozsu. Data Management Issues in Electronic Commerce. In *ACM SIGMOD International Conference on Management of Data*, page 505, Philadelphia, June 1999.

[11] A. Silberschatz, M. Stonebraker, and J.D. Ullman. Database Research: Achievements and Opportunities Into the 21st Century. *ACM SIGMOD Record*, 25(1):52–63, 1996.

[12] T. Wolverton. CNET News.Com: Amazon tries on hosting hat, September 29 1999. URL=*http://news.cnet.com/news/0-1007-200-266350.html*.

[13] G.-H. Yan, W.-K. Ng, and E-.P. Lim. Toolkits for a Distributed, Agent-Based Web Commerce System. In *International IFIP Working Conference on Trends in Distributed Systems for Electronic Commerce (TrEC'98)*, Hamburg, Germany, June 1998. Also appeared at the *Web Development Journal* at *http://WebDevelopersJournal.com/articles/ecommerce/yan01.html*.

[14] G.-H. Yan, W.-K. Ng, and E.-P. Lim. Incremental Maintenance of Product Schema in Electronic Commerce—A Synonym-Based Approach. In *IEEE 2nd International Conference on Information, Communications and Signal Processing (ICICS'99)*, Singapore, December 1999.