8-16-1996

An Intelligent Agent-based Architecture for Interoperability among Heterogeneous Medical Databases

V. Ramesh
Department of Information Systems, University of Maryland, venkatar@umbc.edu

Kip Canfield
Department of Information Systems, University of Maryland, canfield@umbc.edu

Steve Quirologico
Department of Information Systems, University of Maryland, squiro1@umbc.edu

Marcelo Silva
Department of Information Systems, University of Maryland, silva@umbc.edu

An Intelligent Agent-based Architecture for Interoperability among Heterogeneous Medical Databases

V. Ramesh, Kip Canfield, Steve Quirologico and Marcelo Silva
Department of Information Systems
University of Maryland, UMBC
Baltimore, MD 21228
{venkatar, canfield, squiro1, silva}@umbc.edu

Introduction

Modern medical information systems utilize a number of diverse databases to accomplish their day-to-day data management functions. Typically, these databases are heterogeneous in that they store different types of data, represent data differently, use different software to manage the data and run on different computer hardware. Effective decision making often requires access to data from multiple such databases. Research on heterogeneous database management has focused on the development of mechanisms that can access data from multiple databases, while preserving the local autonomy of the databases, i.e., without making changes to the existing databases.

To achieve this objective, it is necessary to develop integrated representation(s) that convey the semantics of the underlying databases. Such representations are expected to hide the existence of multiple databases as well as the inherent heterogeneity in them from the user. Users can use these representation(s) to formulate queries on the underlying databases. The process of deriving these integrated representation(s) as well as answering queries requires cooperation among the underlying autonomous heterogeneous databases.

Traditional approaches to heterogeneous database integration have attempted to resolve structural differences among the underlying databases. Two popular such approaches are the global schema and the federated schema approach (Sheth and Larson, 1990). In the global schema approach, schemas corresponding to each local database are combined into a single integrated schema. In the federated approach, each local database provides an export schema, i.e., a portion of its schema that it is willing to share with other databases. Each database then uses these export schemas to define an import schema, a partial global schema representing information from remote databases accessible locally.

Researchers have questioned the scalability of either approach when a large number of databases need to be integrated. Kashyap and Sheth (1994) describe an alternative approach to heterogeneous database interoperability that is based on understanding the semantics of the underlying databases. The authors use contexts to represent the semantics of the underlying databases. These contexts, in turn, are defined using domain ontologies. Thus, the problem of understanding and resolving the difference in structure and contents of the underlying databases is reduced to understanding and resolving differences in domain ontologies relevant to a query . It should be noted that using their approach, the problem of integration is not completely solved. There is still the need for the generation of a federation ontology, i.e., an ontology that is generated by integrating individual domain ontologies. The authors propose the use of CYC or reuse of other existing classifications, such as, ISBN etc.

Implementing an approach such as the one described in Kashyap and Sheth (1994) requires the development of mechanisms that will allow cooperation among autonomous databases. Recently, there has been an increased amount of interest in the use of agent-based architectures for achieveing cooperation among autonomous (information) systems. This interest has been fueled, in part, by the development of agent communications languages, such as, the Knowledge Query and Manipulation Language (KQML) (Labrou and Finin, 1994) and the Knowledge Interchange Format (KIF) (Geneserth and Ketchpel, 1994).

In this paper, we describe an architecture for achieving interoperability among autonomous heterogeneous databases using intelligent agents. This work extends the work presented in Ram and Ramesh (1995) which demonstrated the use of blackboard architectures for coordinating communication among multiple agents (human and computational) during the schema integration process.

An Intelligent Agent-based Architecture for Heterogeneous Medical Databases

Fig. 1 shows the proposed architecture for providing interoperability among heterogeneous medical databases using intelligent agents. Each agent in the diagram "serves" a single database and performs the following functions:

1. It serves as an intermediary between the users and the underlying databases. Hence, each agent contains the know-how required to translate a user query into a query on the local database that it serves.

2. Its knowledge base contains meta information needed to facilitate heterogeneous database interoperability. Examples of knowledge that an agent would store include, terminological and domain specific knowledge (data dictionary), the database schema of the database served by the agent, and semantic integrity constraints that are specified on the local database (Ramesh and Ram, 1994).

In essence, we envision that each agent will store a representation of the semantics of the database it serves. The semantics of a database will be conveyed by generating a meta-data representation using concepts from a shared ontology. We intend to use the Unified Medical Language System (UMLS) (NLM, 1996) for this purpose. The UMLS is a set of knowledge sources designed to promote retrieval and integration from multiple biomedical information systems. It consists of four knowledge sources: a Metathesaurus, a Semantic Network, an Information Sources Map (ISM) and a SPECIALIST lexicon. The metathesaurus presents a description of various biomedical concepts. The semantic network relates concepts described in the metathesaurus by placing them into general categories or semantic types. The ISM describes sources where more information on concepts described in the metathesaurus can be found. Since, the purpose of an ontology is to specify a representational vocabulary for a shared domain of discourse (Gruber, 1993). the UMLS metathesaurus can be thought of as defining a ìglobal ontologyî for the biomedical domain.

Each agent (serving a database) will register with an agent broker a representation of its meta-data that conveys the domain of problems that can be handled by the database it serves. Currently, we envision this representation to be in the format specified in the ISM.

The sequence of events that transpire in answering a user query is as follows:

1) A user issues a query at one of the client workstations. This query gets sent to the agent broker.

2) The agent broker in consultation with the UMLS agent looks at the set of registered meta-data representations and determines the agents that can provide answers to the user's query and sends the query to the agents. For example, the UMLS metathesaurus and semantic network would be utilized to map the terms used in the user's query to more general concepts (that individual databases may have used in their meta-data representation).

3) Each agent (that receives the query) translates the high-level query into a database specific query (using its knowledge of the underlying databaseís structure).

4) The results of the query are sent back to the agent broker who assembles the results from the various agents and sends it back to the client.

Note that in the above architecture the need for integration of any kind has been eliminated. Similar to the approach presented in Kashyap and Sheth (1994), we have reduced the problem of understanding the structure and contents of the underlying databases, to the problem of being able to generate an appropriate meta-data representation of the domain of problems that a database can solve. We have currently used the architecture described above to implement a prototype of a computerized patient-record system that needs to interact with other medical repositories (Canfield et. al., 1996). Currently, we use the agent communication language KQML for sending messages between the clients and the various servers.

Conclusions

This paper brings together two areas of immense interest to the research and business communities: heterogeneous databases and intelligent agents. The architecture described in the paper presents an elegant platform for providing interoperability among heterogeneous databases (while maintaining the autonomy of the underlying databases). However, several questions still need to be answered. What kind of meta-data representation is needed at each agent? How does one arrive at a subset of the meta-data that we register with the broker? What classes of queries can be supported by such an architecture ? We hope that as our project develops further we will be able to provide better answers to these questions.

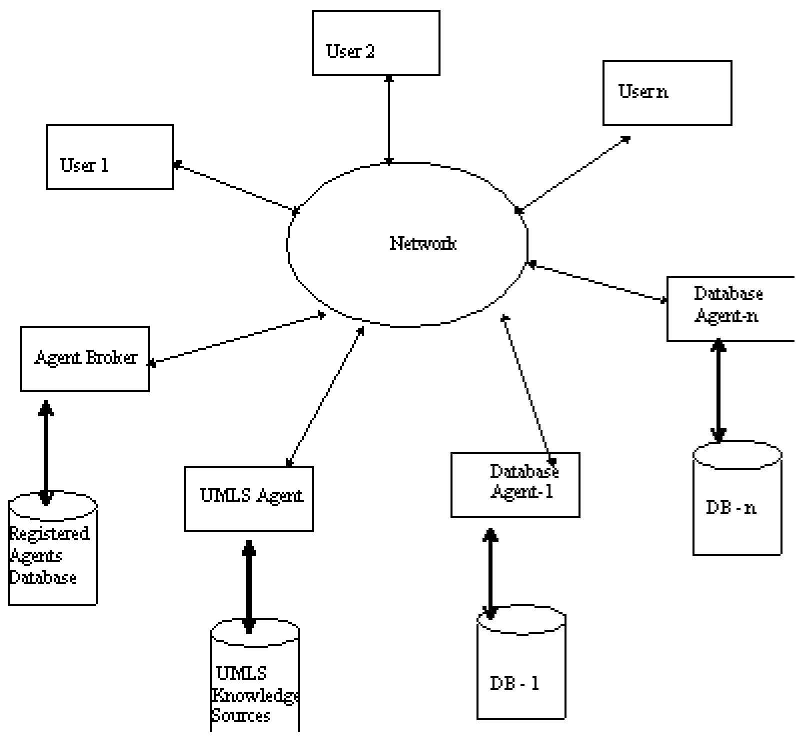

Fig. 1 An Agent-based Architecture for Interoperability among Heterogeneous Medical Databases

References

Canfield, K. , Ramesh, V., Quirologico, S. and Silva, M., An Intelligent Information Systems Architecture for Clinical Decision Support on the Internet, *Working Paper, University of Maryland, UMBC*, Baltimore, MD.

Geneserth, M.R. and Ketchpel, S., Software Agents. *Communications of the ACM*, 37(7), 48-53, 1994.

Gruber, T., A translation approach to portable ontology specifications. *Knowledge Acquisition, An International Journal of Knowledge Acquisition for Knowledge-based*

Systems, 5(2), 1993.

Kashyap, V. and Sheth, A.P., Semantics-based Information Brokering. *Third International Conference on Information and Knowledge Management (CIKM'94)*, 1994.

Labrou, Y. and Finin, T., A Semantics Approach for KQML - A General Purpose Communication Language for Software Agents. *Third International Conference on Information and Knowledge Management (CIKM'94)*, 1994.

NLM, UMLS Knowledge Sources: 7th Experimental Edition. *NIH/National Library of Medicine*, 1996.

Ram, S. and Ramesh, V., A Blackboard Based Cooperative System for Schema Integration.*IEEE Expert*, 10(3), 56-62, 1995.

Ramesh, V. and Ram, S. Integrity Constraint Integration in Heterogeneous Databases: An Enhanced Methodology for Schema Integration. *Fourth International Workshop on Information Technology Systems (WITS'94)*, 282-291, 1994.

Sheth, A.P. and Larson, J.A. Federated Database Systems for Managing Distributed Heterogeneous and Autonomous Databases. *ACM Computing Surveys*, 22(3), 183-235, 1990.