The InfoSleuth Project

R. J. Bayardo Jr., W. Bohrer, R. Brice, A. Cichocki, J. Fowler, A. Helal,
V. Kashyap, T. Ksiezyk, G. Martin, M. Nodine, M. Rashid,
M. Rusinkiewicz, R. Shea, C. Unnikrishnan, A. Unruh, and D. Woelk

Microelectronics and Computer Technology Corporation (MCC)
3500 West Balcones Center Drive
Austin, Texas 78759

http://www.mcc.com/projects/infosleuth
sleuth@mcc.com

1 Introduction

The InfoSleuth Project at MCC [7, 9, 8, 1] is developing and deploying technologies for finding information in corporate networks and in external networks, such as networks based on the emerging National Information Infrastructure. InfoSleuth is based on MCC's previously developed Carnot technology [2, 6, 10], which was successfully used to integrate heterogeneous information resources. The Carnot project developed semantic modeling techniques that enable description of the information resources and pioneered the use of agents to provide interoperation among autonomous systems. The InfoSleuth Project investigates the use of Carnot technologies in a dynamically changing environment, such as the Internet, where there is no formal control of the registration of new information sources and the identities of the resources to be used may be unknown at the time the application is developed. InfoSleuth deploys semantic agents [9, 5, 3] that carry out coordinated searches and cooperate with each other to merge the retrieved data into understandable information.

The project is developing technologies to support mediated interoperation of data and services over information networks in a dynamically changing environment, including:

- Tools for advertising, discovering, accessing, and combining information sources, using domain-specific ontologies;
- Mechanisms for consistent update of the information stored in multiple information repositories and to asynchronously monitor the changes to the information sources, and to signal the occurrence of events;
- Tools for data mining, scenario-based data analysis, and knowledge discovery.

The project aims at the development of advanced prototypes of potentially commercializable technologies in the area of agent architectures, agent communication, multidatabase access, data mining, and resolution of semantic heterogeneity. The prototype implementation utilizes MCC technologies in areas of deductive database and extended transactions and complements them with modern paradigms of Java-enhanced web browsers, and agent communication languages. InfoSleuth comprises a network of agents communicating by means of the high-level query language KQML [3]. The knowledge representation language KIF [4], the database query language SQL, and the deductive database language LDL++ [11] are used to represent queries over specified ontologies. The queries are routed by mediation and brokerage agents to specialized agents for data retrieval from distributed resources, and for integration and analysis of results. Users interact with this web of agents via a Java-capable Web browser that communicates with a personalized intelligent User Agent.

InfoSleuth agents advertise their services and process requests for those services either by making inferences based on local knowledge, by routing the request to a more appropriate agent, or by decomposing the request into a collection of sub-requests and then routing these requests to the appropriate agents and integrating the results. Decisions about routing of requests are based on the "InfoSleuth" ontology, a body of metadata that describes agents' knowledge and their relationships with one another. Decisions about decomposition of queries are based on a domain ontology, chosen by the user, that describes the knowledge about the relationships of data stored by resources that subscribe to the ontology. This semantic layer provides collaborating agents with a common vocabulary for interaction.

Detailed information about the InfoSleuth architecture and the functionalities of each of its agents can be found in this same proceedings [1].

2 Demonstration

In our demonstration, the Monitor Agent (depicted in Figure 1) receives copies of agents' messages and displays them graphically, indicating message history by shifting older messages further and further to the right as it displays the newest message at the left of the display window. The Monitor also retains the contents of the messages to provide assistance in debugging agent interactions. These can be selected directly from the graphic display or stepped through in sequence. Communications to and from agents that are not recognized appear as lines that leave the message display screen. Client-side Graphic User Interface (GUI) applets are

SIGMOD '97 AZ,USA
© 1997 ACM 0-89791-911-4/97/0005...$3.50

implemented on top of a simple API that provides communication with back-end user agent services and query functionality via the Java Remote Method Invocation (RMI) communication layer; this means that applet writers do not have to understand anything about KQML, KIF, or RMI, but can write to an API that deals in terms of entity-relationships, embedded SQL, or other high-level logical constructs.

The following demonstrations are planned:

- **IMTS**— an Integrated Management Tool Suite which provides a set of GUI tools for ontology creation and maintenance. It is intended to be used by the *system integration administrator*. The tools assist the administrator in analyzing and consolidating a collection of data resources into a common ontology. The tools also assist in managing and editing ontologies.
- **ontoviewer**— An ontology-based query builder (implemented as a Java applet) that assists the user in learning and formulating queries based on concepts described in common domain ontologies. The user browses through available ontology and choses builds her query incrementally.
- **Dynamic integration**— A demo capturing the dynamic behavior of InfoSleuth. In this demo, data resources will advertize themselves as they become available. The automatic discovery and dynamic integration of these resources will be demonstrated.
- **Subscriptions and notifications**— This demo will capture the openness and selectiveness of query processing. Continuous query processing will be demonstrated by information subscription and notification over existing and non-existing information sources.
- **Information brokering**— Brokering and constrained information matching will be demonstrated. The demo will monitor and display the Broker's decisions on which resources are likely to be relevant to specific user queries (without actually accessing the resource). A health care application will be used for this demo. The Broker's decisions are based on constraint matching (between the data sources and the queries). The demonstration will include queries constrained to expressions on open-ended ranges and set membership and on logical conjunction of simple constraints.
- **healthdm**— a demonstration of a collection of GUI tools to perform data mining and statistical analysis, for both general and application-specific data evaluation. A Health care data mining application will be demonstrated in which elementary Bayesian probability calculations will be performed in connection with high-level data mining queries.

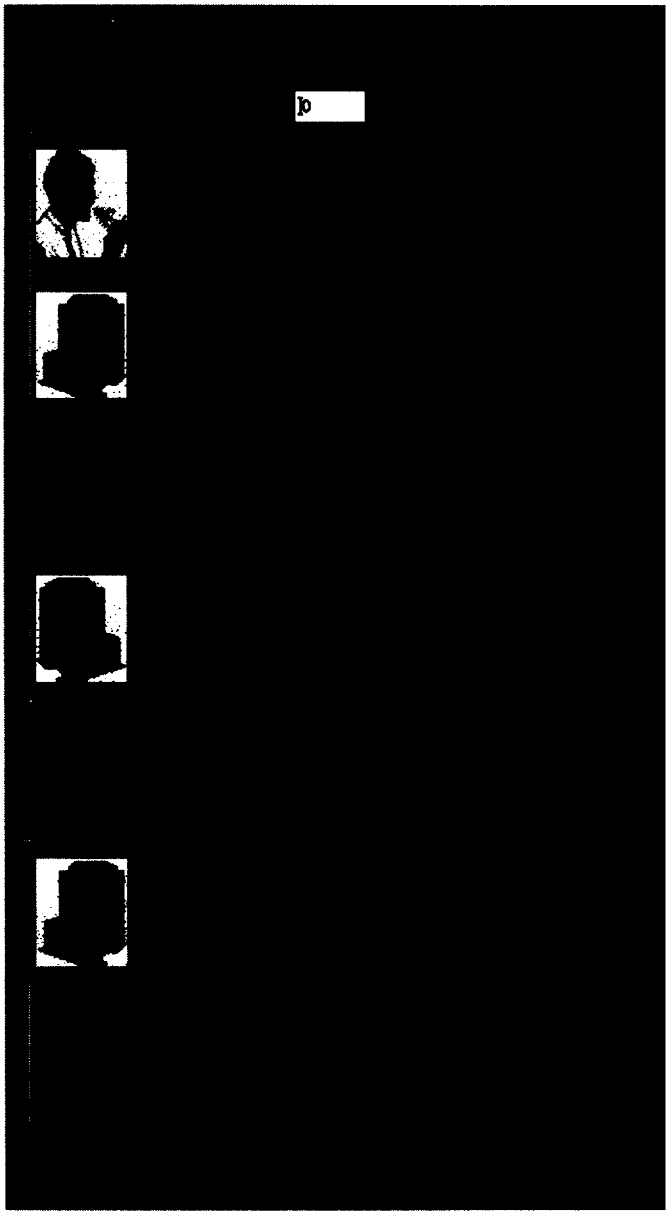

Figure 1: The Monitor Display: An Execution Agent issues a query to a Resource Agent.

References

[1] R. J. Bayardo Jr., W. Bohrer, R. Brice, A. Cichocki, J. Fowler, A. Helal, V. Kashyap, T. Ksiezyk, G. Martin, M. Nodine, M. Rashid, M. Rusinkiewicz, R. Shea, C. Unnikrishnan, A. Unruh, and D. Woelk, "InfoSleuth: Agent-Based Semantic Integration of Information in Open and Dynamic Environments," In Proceedings of the ACM SIGMOD Conference on Management of Data, May 1997.

[2] "http://www.mcc.com/projects/carnot"

[3] T. Finin, R. Fritzson, D. McKay, R. McEntire, "KQML as an Agent Communication Lan guage", Proceedings of the Third International Conference on Information and Knowledge Management, ACM Press, November 1994.

[4] M. R. Genesereth, R. E. Fikes, et al. "Knowledge Interchange Format Version 3 Reference Manual", Logic-92-1, Stanford University Logic Group, 1992.

[5] M. Genesereth and S. Ketchpel, "Software Agents", Communications of the ACM, Vol. 37, No. 7, pp. 48-53, July, 1994.

[6] M. Huhns, N. Jacobs, T. Ksiezyk, W. M. Shen, M. Singh and P. Canata, "Enterprise Information Modeling and Model Integration in Carnot", Enterprise Integration Modeling: Proceedings of the First International Conference, The MIT Press, 1992.

[7] "http://www.mcc.com/projects/infosleuth"

[8] D. Woelk, M. Huhns and C. Tomlinson. "InfoSleuth Agents: The Next Generation of Active Objects", Object Magazine, July/August, 1995.

[9] D. Woelk and C. Tomlinson, "The InfoSleuth Project: Intelligent Search Management via Semantic Agents", Second International World Wide Web Conference, October, 1994.

[10] D. Woelk, P. Cannata, M. Huhns, N. Jacobs, T. Ksiezyk, R. Lavender, G. Merdith, K. Ong, W. Shen, M. Singh, and C. Tomlinson, "Carnot Prototype", in Object-Oriented Multidatabase Systems, O. Bukhres and A. Elmagarmid (editors), 1996.

[11] C. Zaniolo, "The Logical Data Language (LDL): An Integrated Approach to Logic and Databases", MCC Technical Report STP-LD-328-91, 1991.