

Carnot and InfoSleuth: Database Technology and the World Wide Web

D. Woelk, B. Bohrer, N.Jacobs, K. Ong, C.Tomlinson and C. Unnikrishnan

Microelectronics and Computer Technology Corporation (MCC)
3500 West Balcones Center Drive
Austin, Texas 78759
woelk@mcc.com, Phone (512) 338-3468, FAX (512) 338-3890

1. Carnot Database Technology

The Carnot research project [CARN, WOEL93] at MCC was initiated in 1990 with the goal of addressing the problem of logically unifying physically-distributed, enterprise-wide, heterogeneous information. A prototype has been implemented that provides services for enterprise modeling and model integration to create an enterprise-wide view, semantic expansion of queries on the view to queries on individual resources, and interresource consistency management. Carnot also includes technology for 3D visualization of large information spaces, knowledge discovery in databases, and software application design recovery. The Carnot prototype software has been used by the sponsors of the Carnot project to develop a number of applications. These applications have included workflow management, heterogeneous database access, knowledge discovery in large databases, and integrated access to both text databases and structured databases from a single initial query.

The implementation of the Carnot system has required unique advances in two technology areas. First, innovative techniques for knowledge representation have been developed to capture and maintain an enterprise model and to map operations between an enterprise model and the physical databases. Second, a flexible, dynamic, distributed processing environment has been developed that supports the automatic generation of program scripts that execute on heterogeneous, distributed systems. The scripts control the flow of processing and can be reconfigured dynamically to respond to changes in the hardware environment or to the incorporation of additional information resources. The scripts are embedded in autonomous computing agents that can be dispatched to remote sites.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the ACM copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Association of Computing Machinery. To copy otherwise, or to republish, requires a fee and/or specific permission.
SIGMOD '95, San Jose, CA USA
© 1995 ACM 0-89791-731-6/95/0005..\$3.50

2. InfoSleuth Agents on the World Wide Web

The InfoSleuth research project [INFO, WOEL94, WOEL95] has enhanced the Carnot software to make legacy database systems easily accessible via the World Wide Web. This work is in support of the InfoSleuth goal of developing and deploying new technologies for finding information that is available both in corporate networks and in external networks such as the National Information Infrastructure (NII). The InfoSleuth project will investigate the use of Carnot technology in a more dynamically changing environment, such as the Internet, where new information sources are constantly being added and for which there is no formal control of the registration of new information sources. In this type of environment, traditional techniques for expressing and optimizing database queries are inadequate because of the rapidly changing schema information and the fuzzy nature of the queries.

InfoSleuth will build on Carnot semantic modeling capabilities to enable "deep" descriptions of available information sources. InfoSleuth will deploy semantic agents (based on the Rosette scripting language and the Carnot Extensible Services Switch (ESS) [TOML93]) to carry out distributed, coordinated, self-adapting search algorithms. InfoSleuth agents will collaborate with each other and can remain active in the network to monitor events or the addition of new information.

3. Database Applications on the World Wide Web

A Carnot-based application previously developed for use at Eastman Chemical Company in Kingsport, Tennessee has been modified to support remote access via the Internet-based World Wide Web. The original application provided a forms-based interface and a natural language interface (programmed using C++ and Motif) to multiple Eastman Chemical Company databases. These Digital Equipment RDB databases contain historical information on chemical experiments and the manufacturing of chemical compounds. The

forms-based interface was developed using LDL++, a deductive database system based on the integration of logic programming with relational database technology [ONG95]. LDL++ generates SQL operations that are executed by the RDB DBMS. The natural language interface was developed using the MCC Knowledge Based Natural Language (KBNL) [SING94] software. KBNL was used to interpret English queries and translate them into SQL. LDL++ and KBNL sent messages to Carnot semantic agents which provided access to the databases. The semantic agents also translated semantic differences among databases for the natural language queries.

The C++/Motif interfaces used in the Eastman Chemical Company application have now been replaced by HTML forms. The HTML forms can be executed by World Wide Web client software running on a variety of platforms (Unix, PC, Macintosh) and they enable remote access through the World Wide Web. When a user has filled out a form and submitted the form, an HTTP message is sent to the site where the Carnot ESS software is executing. The HTTP server executing at that site routes the messages containing form data to an ESS agent through a gateway module that conforms to the WWW Common Gateway Interface [McCOO]. The ESS agent then expands the form data into messages to an LDL++ system or a KBNL system.

Information returned to the ESS agent by LDL++ or KBNL is translated into HTML format by the agent. Only a specified amount of information is returned by the agent to the World Wide Web client. The HTML page returned to the World Wide Web client contains a hidden session identifier which can be used in subsequent HTTP messages to retrieve further information resulting from the original query. The Eastman Chemical Company application is being further expanded to take advantage of the development of InfoSleuth agents with specialized knowledge of various chemical and business domains that can collaborate to accomplish design and marketing tasks.

References

- [CARN] "MCC Carnot Project", <http://www.mcc.com/projects/carnot>
- [INFO] "MCC InfoSleuth Project", <http://www.mcc.com/projects/infosleuth>
- [McCOO] McCool, R. "The Common Gateway Interface.", <http://hoohoo.ncsa.uiuc.edu/cgi>.
- [ONG95] Ong, K., N. Arni, C. Tomlinson, Unnikrishnan, and D. Woelk, "A Deductive Database Solution to Intelligent Information Retrieval from Legacy Databases", *Proc. of 4th International Conference on Database Systems for Advanced Applications*, April, 1995.
- [SING94] Singh, Mona, Munindar Singh, and D. Woelk. "Knowledge-Based Natural Language Interfaces for Information Access", MCC Tech. Report in preparation.
- [TOML93] Tomlinson, C., P. Attie, P. Cannata, A. Sheth, M. Singh, and D. Woelk, "Workflow Support in Carnot", *Data Engineering Bulletin*, Vol. 16, No. 2, June, 1993, pp. 33-36.
- [WOEL93] Woelk, D., P. Cannata, M. Huhns, W. Shen, and C. Tomlinson. "Using Carnot for Enterprise Information Integration". *Second Intl. Conf. on Parallel and Distributed Information Systems*. January, 1993. pp. 133-136.
- [WOEL94] Woelk, D. and C. Tomlinson. "The InfoSleuth Project: Intelligent Search Management via Semantic Agents" MCC Tech. Report, Sept., 1994. Also available in the Electronic Proceedings of the Second World Wide Web Conference '94 at <http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/woelk/woelk.html>.
- [WOEL95] Woelk, D. and C. Tomlinson. "InfoSleuth: Networked Exploitation of Information using Semantic Agents", COMPCON, March, 1995.