Information Technology and Software

NETMARK

An advanced XML database integration technique for managing unstructured documents

NASA's Ames Research Center offers for license its NETMARK software, a unique innovation designed to seamlessly integrate structured, semi-structured, and unstructured data and documents across enterprise organizations. Originally developed to integrate the vast quantities of complex, heterogeneous documents existing within NASA, this schema-less integration technique and framework offers a highly scalable, open enterprise database architecture that eliminates or reduces the need for database design and administration, and converts information from a wide range of data types into a single, universal data type for storage, retrieval, and content and context-sensitive query and search. A production-ready, enterprise-level application, NETMARK rapidly assimilates and retrieves gigabytes of disparate information and can be easily integrated with existing applications as well as accommodate new data formats—fitting into the legacy data network while growing with evolving technologies and business practices.

BENEFITS

- **Economical**—eliminates the need to design, develop, and maintain expensive, highly structured relational databases, lowering both software and administrative costs

- **Flexible**—combines information from heterogeneous structured, semi-structured, and unstructured data sources, and enables easy and unstructured data queries

- **Adaptable**—enables query-based composition of environments that support http and https protocols

- **Secure**—limits query results to the information that users and groups have permission to access

- **Custom**—includes configurable databanks for tailored query workflows in diverse applications

technology solution

THE TECHNOLOGY

NETMARK takes advantage of an object-relational model and the eXtensible Markup Language (XML) standard, along with an open, extensible database framework to dynamically generate arbitrary schema stored within relational databases and an object relational database management system. NETMARK maps XML-encoded information into a true data model by employing a customizable data type definition structure, defined by an SGML parser to model the hierarchical structure of XML data regardless of any particular XML document schema representation.

By achieving a true XML data model, NEMARK can help enterprise organizations make better use of the information they need to make business decisions by converting Web pages, text documents, PDF files, spreadsheets, presentations, and other document types into a single, universal data type, then storing it in an objectrelational database. Users can query this database with searches that are based on content or contextual associations. Query results then can be composed into different data types, including presentations, spreadsheets, and text documents, enabling rapid reuse of information and broadening the scope of data from which users can gain knowledge and make decisions.

Most traditional document management systems do not provide an easy and efficient mechanism to store, manage, and query relevant information from heterogeneous and complex data types. To do so, database management systems need a standard for common data and exchange. The industry standard, XML, places structure within documents. The traditional mapping model is limited because the hierarchy is different for each set of XML documents. In contrast, NETMARK's SGML parser models the documents themselves, and its structure is the same for all XML documents, providing independence of any particular XML document schemas.
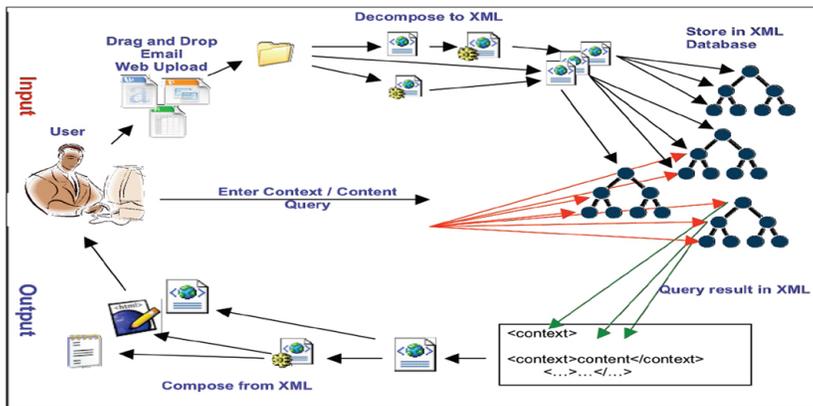
PMT Document Workflow

NETMARK's schema-less integration technique converts information from many different data types into a universal data type for unprecedented information assimilation and retrieval across the enterprise.

APPLICATIONS

The technology has several potential applications:

➲ Enterprise knowledge management applications

➲ Document and content management systems

PUBLICATIONS

U.S. Patent 6,968,338

"Managing Unstructured Data With Structured Legacy Systems," 2008
DOI: 10.1109/AERO.2008.4526666
Conference: Aerospace Conference, IEEE