

# Advanced Data Mining Tool Queries Multiple Databases

## Competitive Advantages

- ☑ Powerful tool for business intelligence and scientific analysis.
- ☑ Makes complex, multiple-database queries efficient.
- ☑ Reveals patterns that would otherwise remain hidden.
- ☑ Adds value for corporate education, and healthcare markets.
- ☑ Low cost to implement in commercial databases.

In both business and research, valuable information can be found by applying complex queries to multiple databases. Typically this is done by querying each database separately and then analyzing the results for possible meta patterns.

The sequential approach is of limited value because the process is time consuming and often fails to recognize important patterns.

## New Capabilities

The new approach proposed here easily performs complex multiple-database queries. It produces results many times faster than traditional methods and delivers more comprehensive results.

Our system finds, analyzes, and develops rules for data patterns as the query proceeds. In effect, it learns as it goes to refine and streamline the search process. This reduces query time while revealing relational data that would otherwise be hidden.

In one recent test, investigators ran queries on various types of U.S. Census data from multiple states. To generate 80,000 patterns, the traditional approach required 10,000 seconds, or more than 2.5 hours. In contrast, the integrated approach proposed here required only 106 seconds, or less than 2 minutes.

## Commercialization

We anticipate strong demand in fields ranging from business and healthcare

to education and government. Specific potential applications include retail sales, population surveys, manufacturing, and sciences such as epidemiology and molecular biology. Potential applications include retail sales, statistics, manufacturing, financial analysis, and sciences such as epidemiology and molecular biology.

A commercial database developer could incorporate this technology into its product line at low cost and with minimal staff or programming time. The developer should realize a significant competitive advantage in each of its markets.

## Next Steps

The technology has been tested successfully on databases of up to 250,000 records. Further testing is planned for databases of up to 1 million records.

## Patent/Licensing Status

Patent pending. Exclusive rights available.

## Learn More

Seminar on data mining algorithms (video)

[www.uvm.edu/~cems/?Page=video/topenalgorithms03-08.php&SM=newsevents/\\_newsmenu.html](http://www.uvm.edu/~cems/?Page=video/topenalgorithms03-08.php&SM=newsevents/_newsmenu.html)

## Primary Investigator

Dr. Xindong Wu  
[www.cs.uvm.edu/~xwu/home.html](http://www.cs.uvm.edu/~xwu/home.html)

## Case Manager

Steve Wernicki  
802/656-9037 (tel) 802/656-8782 (fax)  
[swernicki@uvm.edu](mailto:swernicki@uvm.edu)  
Given Building E201, Burlington, VT 05405

UVM Innovations  
The University of Vermont Office of Technology Transfer

Information Technology

