

Datamation®

Tools To Tame The Data Deluge

On a global basis, data doubles about every nine months, twice as fast as processing capacity. To keep up with the deluge, many companies turn to data mining. Here's a look at some data mining tools and solutions.

Posted November 6, 2002

By [Nathan Segal](#)

Data is omnipresent. Everywhere we go, we encounter it, in our business transactions, subscription lists, email, etc.

To put the growth of data in perspective, every 18 months the processing capacity of the world doubles, but at the same time data has been doubling every nine months. So if an enterprise uses a straightforward approach to processing the data, it'll never keep up. Organizations need next-generation approaches to reduce the data and extract the necessary information. Which is where data mining comes in.

"Data mining uses algorithms to fish through data structures and find you interesting things such as statistical patterns, predictive models, interesting correlations, affinities, etc.," said Dr. Usama Fayyad, CEO and co-founder of [digiMine](#), a data mining services vendor. "It's not about human beings doing it. The reason is that for most large data sets, human beings are not very effective. If you have from one to three variables, people are very good at what we call low dimensions. You can plot the data, leverage your visual system and understand a lot about the data, often much more than an algorithm can understand."

"But the minute you take that number of variables to anything over five or six, humans go from being amazingly good to amazingly bad. When you get into the realm of tens, hundreds and thousands of variables, you need to rely on the algorithm to help you sift through the data. When you have a lot of information about a client, this forms a very high-dimensional space. Taking that data and culling information that is going to be helpful and predictive, such as what the person will buy next, is a very difficult problem. If you leave it to humans to get stuff from the data, they get nowhere."

When looking at the links below, it's important to be aware of a few facts. While there are many data mining tools available, they need to be in a language that the user can understand. If you don't have a thorough knowledge of data mining or statistics, these tools are not for you. In this case, you would be advised to look at the list of software solutions that embed the data mining inside the application.

And in the event that you don't want to do any data mining within your company, you can make use of a managed service model which will do the data mining for you. Such a service is available from [digiMine](#).

The first category contains links to tools designed for data mining experts. Using this software, you would need a data warehouse infrastructure in place.

- [SAS Institute](#)
- [SPSS](#)
- [ANGOSS Software Corp.](#)
- [Unica Corp.](#)
- [Inxight Software Inc.](#)

In this category, data mining is embedded in the database platform for specialized database and mining managers.

- [Oracle](#)
- [Microsoft SQL Server](#)
- [IBM DB2](#)
- [NCR Teradata](#)

This category concerns embedding data mining solutions as a part of business applications, often Customer Relationship Management (CRM). Employing data mining in this context makes it easier for users to learn. However, one would still require expertise and a data warehousing infrastructure, though the mining would take place within a business context.

- [Epiphany](#)
- [Kana](#)
- [Blue Martini](#)
- [Net Perceptions](#)

Parting Remarks

What many people don't realize is how difficult it can be to use data mining within their business enterprise. Another issue is not knowing if data mining will be useful to your business or not. If your enterprise is relatively small, data mining is not likely to be in your best interests. If your business has a limited amount of data, it would be wise to look for opportunities within the data, such as buying trends or fitting customers to different products.

Also, as Fayyad noted, SQL, the language of databases, has proven to be the wrong interface, partly because it cannot handle natural-language queries. SQL was designed to address problems where you know the target and you want the database to quickly retrieve the result. Without an exact description, you're lost. This is one area that could use substantial improvement.

As time goes on, data mining is likely to become part of larger applications, but it is unlikely that data mining will become a consumer product; the intricacies of the technology are too specialized for that. It is difficult to predict how data mining will evolve, but it is likely that the service component for the industry will continue to grow.

Additional Resources

The [Data Mining Group](#) (DMG) is an independent group that develops data mining standards, including the Predictive Model Markup Language (PMML).

According to information on the web site: "PMML is an XML mark up language to describe statistical and data mining models. It describes the inputs to data mining models, the transformations used prior to prepare data for data mining, and the parameters which define the models themselves. Uses include finance, e-business, direct marketing, manufacturing, and defense."

PMML is used by IBM, Oracle, SPSS, Magnify, Angoss, Mineit, etc., with further releases planned. At this point, it is the most widely deployed data mining standard. PMML is part of [XML.org](#).



Property of Quinstreet Enterprise.

[Terms of Service](#) | [Licensing & Reprints](#) | [About Us](#) | [Privacy Policy](#) | [Advertise](#)
Copyright 2014 QuinStreet Inc. All Rights Reserved.