

Key issues

Data mining: does it have applications within the world of libraries?

'Data mining' (with its relatives, text mining and pattern recognition) has sprung into prominence with the advent of mass computing, where almost everything from supermarket purchases to customer telephone queries ends up as data on some computer, somewhere. Banks and airlines use data mining to develop customer profiles and to attempt to tailor offerings to different customer profiles. Supermarkets are extensive users of data mining to generate coupons and special offers, arrange merchandise, plan distribution and analyse the impact of price changes. The medical and pharmaceutical sectors use data mining to discover outcomes of treatment regimes over large patient sampling.

For our purposes, we may turn to a definition of data mining that was developed a few years ago by Stephen Arnold of Arnold IT:

'Data mining is a process of scrutinising a meaningful body of data using specialised tools to discern patterns and relationships (exception definition).'

Thus we have the overlap with text mining (looking just at words and phrases, rather than numeric data) and with pattern recognition (using computer tools to discern patterns within a mass of data). The main contrast between data mining and the more familiar search and retrieval is that data mining is concerned with 'finding what was not sought'. Data mining can suggest things one did not know, and provide answers to questions that were not asked. First of all, however, we need to identify what we are going to be measuring and to decide how to capture, store and represent the data.

Our data warehouses are going to have to be co-ordinated and streamlined, and we are going to need appropriate data mining tools. In particular, the format of the data is vital: data must be normalised, they must be 'clean', and they must be structured – even if the structure be quite



HARRY COLLIER
Managing Director,
Infonortics Ltd.

elementary – so that queries can be run via SQL, AI or via a combination of query tools.

Libraries and data mining

Well, libraries are not supermarkets, nor are they airlines. But they do have a number of areas where they need the kinds of help that data mining can provide:

- Trends in subject popularity to enable better focus of acquisitions and budgets.

- Analysis of usage, borrowing and interlibrary loan patterns to plan branches, mobile stops and to work with schools or faculty.

- Time-of-day traffic (person, phone, web) to plan opening hours and staffing and to deal with seasonal peaks.

- Queues, loan/reserve periods, fines, the opening up of bottlenecks.

- Correlation of in-house transaction data with postal codes or faculty/department codes. Looking for under-represented areas and to discover why, and what the implications may be.

- Benchmarking against similar libraries.

- Preparing 'return on investment' analyses for the annual budget battle and to demonstrate value delivered to constituents.

A number of references within the literature provide some examples of libraries being active in

this sphere; for example, Smathers Library at the University of Florida¹ which describes the complex process of extracting data from NOTIS for analysis by Microsoft Access to discover the costs of buying and storing materials, by method. And Guenther² discusses complexities in capturing data and challenges associated with multiple repositories using different structures – for example, in-house data, data from outside vendors, etc.

The goal of much data mining is customer service which, in a library world, can mean better targeted collections, and product refinement. It can also mean better marketing, with services being better targeted and impact being closely monitored; 'Go beyond stating what was, and into the realm of predicting what could be'³.

Conclusion

Data mining is a complex area. However, extraction of useful information from structured stores of data need not be complex if tackled at a basic level. Libraries are often wary of appearing to monitor their clients; data mining, however, is concerned more with pattern tracking and pattern extraction than with examining individual habits and customs. The literature of data mining is vast. However, some basic investigation such as via⁴ and⁵ could reap real dividends for a go-ahead library sector.

Acknowledgement

For many of the facts and ideas in this short article, I am indebted to Stephen Arnold of Arnold IT in Kentucky whose talks and writings are always an excellent source of elucidation.

References

1. Su, S., Harvesting information from a Library Data Warehouse, *Information Technology and Libraries*, March 2000.
2. Guenther, K., Applying Data Mining Principles to Library Data Collection, *Computers in Libraries*, April 2000.
3. Schulman, S., Data Mining: Life after Report Generators, *Information Today*, March 1998.
4. Nicholson, S. and Stanton, J., *Gaining Strategic Advantage through Bibliomining: Data Mining for Management Decisions, Special, Digital and Traditional Libraries*, to be published (Idea Group, Inc.)
5. Pushpa Ramachandran M., Ragavendra, S. and Turakhia, K., *E-Customer Analytics*. White Paper. Wipro Technologies.

**Harry Collier, Infonortics Ltd, 15 Market Place,
Tetbury, GL8 8DD, England**
Tel: +44 (0)1666 505 772
Fax: +44 (0)1666 505 774
Email: hcollier@infonortics.com
