INFORMATION

OVERLOAD

With the proliferation of databases, on-line services, wireless news feeds and other sources of information, trying to locate relevant facts can be like searching for the proverbial needle in a haystack. A number of mostly small companies have developed software tools that simplify the tasks of search, retrieval and storage. While the market for such tools is still embryonic, it promises to grow substantially.

— BY —

D A V I D
BRODWIN

— AND —

S A N D R A
BERNARDI

THE INFORMATION AGE has lived up to its name: It takes ages to sift through all the information that is now available to organizations. The overabundance of information has created a time-consuming and difficult challenge for busy executives and professionals searching for that key piece of data in an increasingly competitive world.

Until recently, companies managed this overload by hiring dedicated information professionals, typically in a corporate library or research position, to scan on-line CD-ROM, microform and print-based services, and find the few specific pieces of information, among everything published, that would make the difference. These search professionals were necessary for several reasons. First, the services were hard to use: The information retrieval language took time to learn, and you had to know it precisely to avoid overlooking some key data source. Second, the services were expensive. By keeping a single control point, companies were able to limit their budgets.

Now that world is chang-

ing. User organizations are increasingly demanding tools that put information directly into the hands of end-users who need it and which allow the information to be integrated with internal data from the company's ongoing activities.

The emerging world looks more like this: End-users have software tools on their own personal computers and work-stations which they use to scan for information. In some cases, scanning takes the form of a one-time request; in other cases, the user creates a profile of the information needed, and the profile is refreshed continually as new stories appear on the wires and else-where. An on-site LAN-based content server holds often-used databases, for which the company pays an annual subscription fee.

The database provider ships updates to the company using various high-capacity delivery media such as CD-ROM or DAT (digital audio tape). When it is received, the information is usually transferred to fast magnetic drives so that access times are improved. For frequently changing databases, updates will be

provided over high-speed phone links or direct broadcast technologies.

Importantly, the software tools provide a consistent user interface across these diverse sources of content and across different vendors of purchased content. The tools are location-independent, allowing users to search the same way regardless of whether they are searching a local database or on a system across the country. Ultimately, these tools will provide the companies that sell their content with a consistent delivery platform environment to target.

DRIVERS OF CHANGE

A combination of technology changes and new economic incentives and pressures on key vendors are driving this transition.

Technology changes
• Proliferation of LAN installations and E-mail services, which make it possible for end-users to search using a shared content server and to have information distributed readily around the company;
• Development of software standards that provide a com-

mon way to access information from multiple sources;

• Development of powerful, end-user-friendly search software, which allows precise and accurate searching of large and complex databases;

• Improvements in price/performance of workstations and mass storage, which make it economically feasible to store reasonably large databases on a LAN-based content server (typically, multiple gigabytes are required);

• Improvements in price/performance of telecommunications (wireless and wireline), which make it possible to update databases frequently, without a prohibitive investment in expensive leased lines and/or broadcast receivers.

Economic factors

Many prospective players will be fighting for a piece of the action.

• Content owners (primary publishers like McGraw-Hill, CMP Publications, Ziff-Davis and others) want to extract as much value from their ownership as possible, and not give away half or more of the revenue to the information repackagers that now stand between them and the end-customer. As the balance between print-based revenue and electronic revenue shifts, primary publishers are scrutinizing these issues more seriously than they did in the past, when electronic revenue was meager compared with the robust print business.

• Corporate end-users want to put much more information into the hands of individual professionals and managers, but can't accept a situation where bills can fluctuate widely each month. They need a way to have fixed (or at least managed) monthly costs for information access.

• Software manufacturers and telecommunications vendors, facing declining margins in many segments, are looking to the content market as a way to restore margins and create growth.

GETTING THE RIGHT STUFF

For end-users, one of the most difficult aspects of accessing information is finding what they really want. Indeed, for most end-users, using a sophisticated electronic information service is like shopping at Home Depot with the lights out: You know it's got to be there, but you'll never find it. Information scientists discuss this problem in terms of "precision" vs. "recall."

High precision means that every article the system finds for you is about your topic of interest, whereas high recall means that every conceivably relevant article in the database is found. High precision coupled with low recall means you get a few relevant items back when you search, but you may have missed something important. Low precision combined with high recall means you won't miss anything, but you'll spend hours sorting the good from the bad. The challenge is to create a system that delivers high precision and high recall and is still easy to use. As a simple example, if you ask for articles on *IBM Corp.* activities in *personal computers,* you might miss articles on *I.B.M.* activities in *personal computing.*

Most search tools look for word occurrence in text strings, coupled with some restrictors as to type or date of document. The user reviews the results displayed after such a search and, if there are too many articles, too few or just the wrong ones, modifies the search specification until the right result is obtained. Software vendors are addressing this issue in several ways. With a few exceptions, these vendors are small, little-known startups, usually private, with revenue under $20 million.

Both Verity Inc., Mountain View, Calif., and ConQuest Software Inc., Columbia, Md., simplify the process by allowing users to search databases for concepts or meanings rather than individual words. Verity's software called Topic utilizes what the company describes as "concept retrieval." This allows the user to define an overall concept—such as corporate development, which may be associated with a range of specific words like *mergers, acquisitions, strategy* and *shareholder value.* Once a concept has been defined, the user need not specify every conceivable word to get a good search.

ConQuest's approach is to supply users with a semantic network—a ready-made knowledge base of information built from a dictionary and including all possible word definitions. Users can carry out a search based upon every available definition, pick out only the definitions they consider to be relevant or add new definitions.

Another method is called relevance feedback. Here, a user does not need to keep adjusting search terms, but looks at the first 20 or so articles that are produced and says, in effect, "I like these, but I don't like those." Relying on comparisons of word frequency or other parameters, the software then determines what particular characteristics mark the articles that the user wants and refines its internal representation of the search request accordingly. The Smart technology from Individual Inc., Cambridge, Mass., incorporates this.

The user reviews the results displayed after such a search and, if there are too many articles, too few, or just the wrong ones, modifies the search specification until the right result is obtained.

Smart sets up profiles, based on user specifications, which are used to filter stories from a variety of on-line feeds. The articles that match the profile are then sent by E-mail to users, who "grade" the results in terms of relevancy. Smart gradually refines the data profile so that all of the articles captured are very relevant (which usually takes several weeks of data feeds).

Hoover from Sandpoint Corp., also in Cambridge, helps users interact with the information sources in a different way. The software, working with Lotus Notes, provides the user with a single interface that can be applied to multiple on-line services as well as locally stored content. Hoover can search such sources as Dow Jones, Dialog, Newsnet and other on-line sources, as well as CD-ROMs and corporate databases. And Hoover integrates and organizes the information based upon user specifications. For example, it can look through designated external sources for information on companies and present the information in up to 25 different formats, such as financials, joint ventures, subsidiaries and divisions, new

product announcements, etc. Virgil Corp., San Francisco, is working on an unannounced product with similar functionality.

Not to be outdone by these tiny upstarts, Apple Computer Inc., Cupertino, Calif., recently entered the arena with AppleSearch, a product due in the third quarter of 1993 and based on an engine from Personal Library Software, Rockville, Md. AppleSearch is said to provide a natural language-based technique for expressing search requests. In addition, once searches have been defined, they can be stored as agents—to run again on demand.

DELIVERING INFORMATION

How can information get to the customer's LAN-based content server? There are three basic approaches: First, continually broadcast a stream of information over the airwaves (using, for example, Vertical Blanking Interval, FM-Sideband or a satellite service' and capture the information locally. Second, make available on-line data that end-users can capture at will. Third, ship a physical storage medium, such as CD-ROM, tape or DAT, to the customer at periodic intervals.

Desktop Data Inc.. Waltham, Mass., offers NewsEdge, a live news management and processing tool, which allows users to subscribe to individual news feeds from more than 100 data sources such as Associated Press, Reuters or Dow Jones. NewsEdge intercepts these feeds at the user's site via FM, satellite or leased line transmission, and analyzes and distributes the information to users based upon individual profiles.

Another company, Mainstream Data Inc., Salt Lake City, Utah, goes a step further. combining multiple-source news feeds into a single signal and rebroadcasting the information on unused portions of FM bandwidth or via satellite. On the user's end resides Mainstream's Intelligent Data Receiver, which monitors the datastream, decoding and capturing only the information to which the user has subscribed.

Northern Telecom Inc., Nashville, Tenn., has an arrangement with Dialog Information Services, Palo Alto, Calif., that lets the telecommunications com-

---

The Truth about Verity

In the highly competitive software industry, quality customer service is key. Companies like Borland International Inc., Scotts Valley, Calif., and Cisco Systems Inc., Menlo Park, Calif., are using information retrieval tools to bolster their efforts. Both companies tapped the Topic retrieval system from Verity Inc., Mountain View, Calif., to streamline problem resolution.

Borland and Cisco wanted to implement systems that would allow their customer support engineers to retrieve relevant documents over a broad range of information sources, as well as create a "knowledge base" to leverage the expertise of senior engineers. They were looking for a cutting-edge client-server solution that would still be easy to use and maintain.

Cisco is currently managing 2 Gbytes of information, or 200,000 documents, using Topic. Within two weeks of implementing the system, Cisco says, support productivity improved by 10 percent. For a department of 70, that was like adding another seven support engineers. For Borland, the system allowed the company to cut training time of support personnel by two-thirds.

The customer support market is one of the main areas Verity is targeting. "In the 1970s, organizations were automating financial and accounting systems, and the 1980s saw the automation of engineering design and manufacturing systems," says Barry Goss, Verity's director of marketing. "Today, organizations are looking to automate their customer service functions in an attempt to differentiate themselves."

In the past, technical support engineers spent a lot of wasted time "reinventing the wheel"— solving problems that had already been solved. Because they lacked good automation tools, it was easier to start from scratch than work back through hundreds of E-mail messages, product documentation and customer call reports.

For technical support people, products like Verity's open the door to a whole new world of information access. Organizations can index such things as call manuals, release notes, bug reports, technical notes and data sheets to empower the support personnel to provide the best service and response times. Some organizations are also including information from other functional areas, such as marketing and sales, to enable their support engineers to provide more comprehensive answers.

Several software developers, such as Lotus Development Corp., Adobe Systems Inc. and Frame Technology Inc., have chosen to embed the Topic search engine in their products to address the horizontal market with broad capabilities. In addition, some of this technology is finding its way into vertically focused applications, such as human resources management. The Topic engine allows human resources managers to match applicants' skills with unfilled positions. —D.B. & S.B.

---

pany capture data from Dialog's on-line services and redistribute it internally. Northern Telecom has designated 15 to 20 indexed terms to home in on strategic information for competitive analysis. The one-year-old service has grown in popularity, according to Northern Telecom marketing specialist Tim Raga. Other parts of the worldwide company want the service, but access is restricted by copyrights.

Oracle Systems Corp., Belmont, Calif., is developing a service called Data Freight that will distribute information one way at very high rates (0.5 to 1.5 Mbits per second) via satellites and ground-based stations. Oracle is

planning not to compete head to head with existing information distributors, but to focus on print information unavailable in electronic form. This service could also be used to distribute software and multimedia programs.

Another large vendor, Adobe Systems Inc., Mountain View, Calif., recently announced a product, Acrobat, that allows users to communicate electronic documents across platforms, regardless of the application or computing environment used to create the document. Adobe licensed Verity's Topic search retrieval engine and incorporated it to enable users to search for content across diverse documents.

64

Adobe is one of many companies supporting a 20-year-old standard, Standard General Markup Language (SGML), a method for defining document structures. If SGML or an equivalent standard is adopted, this will provide content publishers with a common language to write to, much as Windows provides today's software developers with a common environment.

CAPTURING THE DATA

Traditionally, received information wasn't stored in the user's organization at all; it was simply printed and then lost. Now, there are several tools to maintain a repository of LAN-based information for ongoing use.

Sandpoint's Hoover allows local storage of both internal and external information within the Lotus Notes framework. Users can direct Hoover to conduct searches in multiple sources and store the findings in a local Notes database. SoftArc Inc., Scarborough, Ont., offers local content management with an electronic bulletin board soft-

ware (BBS) product, FirstClass. Users can post and store information and carry out simple text searches from both Macintosh and Windows environments.

Popular groupware products, such as Lotus Notes (acting on its own), also have some features that can maintain content. The prospective user should look carefully, however. Some of these products may offer the advantage of better integration with other office functions, but at the cost of limited retrieval features or slower performance than dedicated software products. Those same trade-offs are inherent in all the information software products—there's no single solution to the problems of obtaining highly relevant information, easy storage and universal data retrieval.

WHO OWNS THE DATA?

As new technology makes it possible to deliver a wide variety of information directly to the end-user, it raises questions about who owns the rights to this content and how they will be paid—questions that, if not properly

resolved, could crimp the growth of this new industry. Take the perspective of a popular publication, say *The Wall Street Journal. The Wall Street Journal* probably has many paid subscribers in a large Silicon Valley company like Apple. If Apple were to ask *The Wall Street Journal* for an electronic version of the publication which Apple could distribute freely and without further charge over the internal network, some of those subscription revenues would be at risk. But if *The Wall Street Journal* charged enough for "*The Wall Street Journal* on Disk" to make up the potentially lost revenues for the print version, the disk version would be far too expensive for smaller companies.
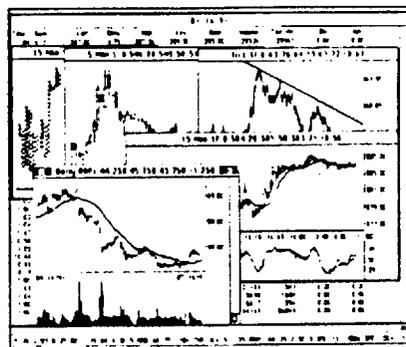
For this reason, print publishers have mostly taken a "go slow" approach to electronic distribution, limiting redistribution rights and restricting fixed-price unlimited use contracts. In fact, some of the software and services vendors named above say their contracts with the original content supplier limit how much information they can allow

The following table compares several software and services vendors in the emerging LAN-based electronic information industry. Each company's products offer varying functionality. The indicated functions include abilities added by partners.

	Content Provision	Content Maintenance	Transmission Provision	Local Solution	Media Supported	User Interface & Search Approach
ConQuest Software Inc. Columbia, Md.	Limited or no involvement	Database maintenance	Limited or no involvement	LAN server	CD-ROM, hard drive, dial-up	Only for content in own data structure
Desktop Data Inc. Waltham, Mass.	Works with content providers	Database maintenance	Helps customer buy and integrate	LAN server	CD-ROM, hard drive, wireless	Same
Individual Inc. Cambridge, Mass.	Same	Capture and forward	Provided through E-mail, Lotus Notes	Only through Lotus Notes	Fax, E-mail, Lotus Notes	Same
Mainstream Data Inc. Salt Lake City	Same	Database maintenance	Provides transmission	LAN server or single user	Wireless	Same
Sandpoint Corp. Cambridge, Mass.	Same	Database maintenance	Helps customer buy and integrate	LAN server or single user	CD-ROM, hard drive, wireless, dial-up	Same
SoftArc Inc. Scarborough, Ont.	Limited or no involvement	Within bulletin board	Limited or no involvement	LAN server or single user	Dial-up	Same
Verity Inc. Mountain View, Calif.	Works with content providers	Database maintenance	Helps customer buy and integrate	LAN server	CD-ROM, hard drive, wireless	Single user interface for multiple-source searching

their customers to retain for future use—hardly an economically efficient situation!

For this industry to take off, all participants will have to get much more sophisticated and flexible about how to price and delimit source data rights. The end result may be about as complex as airline seat pricing or it may emulate financial securities information. For example, the delayed "The Wall Street Journal on Disk" may cost a lot less than the real-time "The Wall Street Journal On-line."

SHRINKING THE MIDDLEMAN

As end-users are able to negotiate better access to more diverse sources of data, a power shift will take place within the information industry: away from vendors that have historically repackaged information and delivered it on-line (if they don't adapt) and toward both the customers and the original suppliers of content.

Two kinds of customers will be the early adopters—first, those that have a large base of professional employees and have historically incurred high annual charges for on-line and print-based information services. Investment banks, consulting firms and many technology- and R&D-intensive businesses, among others, fall into this category. These compa-

nies will move fast both because better access to information makes their professionals more productive and because they can get more value for their investment by bringing the delivery capability in-house and site-licensing the most frequently used content.

Other early adopters will be newer firms staffed by young executives who are already very comfortable with computers, and who constantly seek ways to leverage information technology to avoid or defer adding headcount. However, prospective sellers of the new information products and services described here should keep in mind that, unless a company has already invested heavily in local networks and has a high penetration of PCs among its managerial and professional staff, it is probably a poor sales prospect.

The suppliers who hold source rights to original content will gain power as they capture a greater percentage of the continually increasing revenue stream for electronic information, but they will see increasing pressure on revenues for their print-based products. These suppliers—primarily media companies—will need to invest heavily to come up the learning curve of this new industry. They need to master new technologies, new marketing and distribution practices, and new pricing struc-

tures. They need to learn to confront a new class of competitors as diverse as AT&T Corp. and Microsoft Corp., which have substantial ambitions and capabilities in the information-content industry.

Finally, for the emerging software vendors discussed in this article, we see substantial opportunity, both in selling information retrieval tools directly to end-users and in partnering with content providers, telecommunications vendors and "old technology" intermediaries—many of whom want to cash in on this market but who lack the specific software know-how to do so. Ultimately, we think the winning companies will be those that concentrate on a narrow range of technologies but which propagate their products (and the potential standards they imply) across a broad range of channels and partners (as Verity is doing). The also-rans and losers will be those that persist in selling proprietary bundles of content, connectivity, access and storage to narrowly defined vertical and functional markets. ∎

David R. Brodwin is director, management consulting, in the computer and information industry practice of Arthur D. Little, San Francisco. Sandra Bernardi is a research associate. This article is based on a more extensive report which is available on request.

66

Now that the hype over Windows NT has subsided, it's time to take a sobering look at the impact it will have on other vendors and the computer industry as a whole. Despite Microsoft Corp.'s heavyweight champion status on the desktop with MS-DOS and Windows, Windows NT is likely to remain in the contender category for some time.

Sun Microsystems Inc. and Unix are not going to get knocked out in the first bout. Neither will Novell Inc. and NetWare. I'm less sure about IBM Corp.'s OS/2. If IBM's new CEO Louis Gerstner is seriously looking for a way to save money and bring IBM back into the thick of the fight, he would do well to give up on OS/2 and put his effort behind one of the operating systems that has a better chance of winning.

Both Sun and Novell have proven technology, satisfied users, strong software partners, and established distribution and support channels. These are all attributes that will keep them on the leading edge of the contest. It will be some time before Microsoft and Windows NT can match their expertise.

For the last decade some of the best minds in the industry have been at work improving and enhancing Unix. It has proven its ability to run large-scale operations at financial institutions, retailers, government agencies and manufacturing sites. It has been ported to numerous platforms from portable computers to the world's fastest supercomputers. While I am anything but a Unix bigot, I believe that Unix is going to be around for some time.

Novell's NetWare has proven it is a better network operating system than anything Microsoft has been able to develop. LAN Manager has been markedly unsuccessful when it has gone head to head with NetWare, despite massive marketing efforts by Microsoft. If this shows us anything, it's that Microsoft is not invincible. Everything it produces does not automatically win.

Many people in the industry have short memories. They forget that it took Microsoft years before it introduced a version of Windows that users wanted. The adage that it takes Microsoft three tries to get it right is true more often than not.

Although I have no doubt Windows NT will eventually be a strong contender, I think it will take a few years before it is robust enough so that large companies will feel secure about running mission-critical applications on NT. Until then they will wait and see.

Besides dealing with the technical difficulties, Microsoft will face problems building up the required support and distribution infrastructure for an operating system as complex as Windows NT. Early indications are that Microsoft still does not fully grasp the kind of services that it must provide if it expects corporate users to run large-scale applications on Windows NT.

Neither will Windows NT provide an immediate opportunity for Microsoft to unseat a host of application software vendors as it did when it introduced Windows 3.0. When Windows 3.0 was introduced Microsoft unleashed Windows versions of Excel, Word, Works and other applications before its competitors did, vaulting Microsoft to the top in a number of categories.

With Windows NT, Microsoft does not have any applications designed for the client-server world. Instead it is relying on its software partners to provide them. And many of those companies are already selling their applications on Unix and NetWare.

The real importance of Windows NT is not that it is going to allow Microsoft to immediately extend its desktop hegemony to the server, minicomputer and mainframe. Rather, the introduction of Windows NT is significant because it puts the final nail in the coffin of host-based computing. From now on the ruling paradigm is client-server computing.

In the short term, the endorsement of client-server computing by Microsoft will probably help Unix and NetWare more than it hurts them, because vendors like Sun and Novell can provide the technology to do this now, and Microsoft can't.

In the long run, however, Microsoft will enhance Windows NT, and build the required support and service infrastructure. That, combined with Microsoft's massive installed base and marketing prowess, will allow Windows NT to emerge as the dominant, though not sole, client-server operating system.

Eric Nee

SECOND INTERNATIONAL SYMPOSIUM: NATIONAL SECURITY & NATIONAL COMPETITIVENESS: OPEN SOURCE SOLUTIONS Proceedings, 1993 Volume I - Link Page

Return to Electronic Index Page