The Factiva 2004 White Paper Series

Free, Fee-Based and Value- Added Information Services

By Mary Ellen Bates
April 2004

Executive Summary

The key to making good decisions is good information. When looking for information on companies and issues outside their own organizations, today's managers and knowledge workers go online.

Three types of business information providers are available online: free Web sites, fee-based Web sites and value-added information services.

The vast majority of workers seek free information on the Web. But most important business sources are not available for free on the Web. And, because searches on the Web cannot be aggregated, finding useful information is difficult and time-consuming. Even the "free" information on the Web actually comes at a substantial cost to the enterprise in terms of hours spent looking for information. And the cost to an enterprise of *not* finding information can run into the millions of dollars.

Value-added information services, on the other hand, offer high-quality information from a wide variety of reliable sources. Plus, they provide user interface tools to aggregate and speed up information searches.

To this end, these services provide better information with access to both current content and archives for continuous availability. Value-added information services, therefore, enable businesses to make more informed and more effective decisions.

Information — the foundation of business decisions

Every business manager, in every industry, has the same fundamental job responsibility — the responsibility of making decisions. The key to making good decisions — choices that further the goals and objectives of the organization — is good information.

In today's global village, the information required to make decisions often resides not within the organization itself, but beyond its boundaries. When options are being weighed related to competitive, regulatory or product development issues, for example, what managers need most are facts about other companies and institutions. To find that external information, today's managers and knowledge workers go online.

But the online world is a very big place. In its April 2004 *Web Server Survey*, Netcraft found almost 50 million active Web sites, and this does not even count the number of pages *within* a particular site. In Steve Lawrence's and C. Lee Giles' paper, "Accessibility of Information on the Web," they estimated an average of 289 Web pages per server. Applying that multiplier to Netcraft's survey, one can conclude that the Web now contains at least 14 billion pages, which does not include the content within databases or dynamically generated content.

Looking for a particular bit of information among all those Web pages makes the proverbial needle-in-haystack search seem simple. The only sensible way to approach the task is to have the most sophisticated pitchfork and to know in which haystack that needle is most likely to appear.

The high cost of *not* finding information

IDC analysts Susan Feldman and Chris Sherman studied the amount of time knowledge workers spent looking for information, how successful those searches were, and the cost to the enterprise in terms of lost time and decisions based on incomplete or incorrect information. In their 2003 report, *The High Cost of Not Finding Information*, they estimated that knowledge workers spend 15 percent of their time looking for information; some studies estimate that employees spend up to 25 percent of the work day searching for information. However, they also estimated that half of the searches were unsuccessful.

What causes these searches to fail? A number of factors were cited by Feldman and Sherman, including:

- selection of a source not suited for the query; that is, searchers attempted to find information from a source that was not likely to contain the needed information, such as searching for back issues of a periodical from the publication's Web site, not realizing that the site did not include articles more than a month old
- lack of understanding of how to construct a search query; many searchers submit single-word queries, which often return many results, most of which are irrelevant
- poor user interface, which prevents the searcher from identifying and rectifying search errors
- information that exists but is not sufficiently indexed in search engines or other Web finding tools and therefore cannot be found by searchers

Assuming that the average salary of knowledge workers is $80,000, that they spend six hours a week (15 percent of their work time) searching, and that 50 percent of their searches are unsuccessful, IDC concluded that at least $6,000 is wasted per knowledge worker annually. Hence, an enterprise employing 1,000 knowledge workers wastes an average of $6 million per year from workers looking for and failing to find the information they need.

Inappropriate sources, ineffectual search tools

What is particularly striking about the numbers from the IDC study is that much of the problem may lie in workers' over-reliance on search engines as a means of finding information. There is a perception among many Web searchers that a query in Google will retrieve everything on that subject that appears on the Web. Type "marketing" in Google's search box, and you will retrieve over 68 million pages – although you cannot view more than the first 800 or so. Seeing search results in the millions encourages the belief that one need go no further than a search engine to find anything.

Unfortunately, this ignores the problem of what is often referred to as the "deep" or "invisible Web" – that portion of the Web that search engines cannot or will not index. Estimates of the size of the deep Web vary from two to 400 times larger than the "visible Web." None of the content within the deep Web can be found through a search engine, and this content is often richer and more diverse than open Web content. What type of information is hidden from search engines?

- information in databases; search engine spiders –- the indexing programs that scour the Web for content – cannot type queries into search boxes, so cannot retrieve any of the records within a Web-based database
- audio or video content; search engine spiders cannot "read" sounds or images
- content in formats search engine spiders choose not to index, such as Flash or compressed files
- information on sites that require registration, such as many newspaper sites, since search engine spiders cannot fill in registration forms
- content that has changed since the last time the page was indexed by a search engine
- pages that have never been linked to by a page that a search engine has already discovered; search engine spiders cannot index a page for which they do not know the URL
- dynamically generated pages – pages built on-the-fly based on a query or where you clicked from

According to "The Deep Web: Surfacing Hidden Value" by Michael Bergman, the material in the deep Web is, on average, of three times higher quality than the open Web, based on Bergman's document scoring methodology. And Chris Sherman and Gary Price, authors of *The Invisible Web*, comment that "invisible Web content tends to be more focused, and often provide better results for many information needs. Consider how a publication like *Newsweek* would cover a story on Boeing compared to an aviation industry trade magazine such as *Aviation Week and Space Technology.*" Unfortunately, searchers who rely on a search engine never find any of this deep Web information.

Free, fee and value-added

Three general types of business information providers are available online:

• Free Web sites — sites that do not charge for the information available on them, including both open Web and deep Web sites.

• Fee-based Web sites — sites that offer copyrighted material such as articles or analysis, charging for access.

• Value-added information services — Web-based providers that charge for access but offer a wide range of business information — much of it not available on the free Web or through fee-based Web sites.

Free Web sites

Free Web sites include Web pages maintained by companies, publishers, government agencies, trade associations and experts on a particular topic or industry. Most of the content of interest to business researchers is contained within deep Web sites, which cannot be searched using a Web search engine. Examples of free Web sites are the *New York Times* site (www.nytimes.com), CorporateInformation.com, and the CNN financial news site (www.cnnfn.com). Note that most collections of publicly available government documents are inaccessible to search engines, including the U.S. Securities and Exchange Commission database of filings (edgar.sec.gov), the U.S. Patent and Trademark Office database (www.uspto.gov) and the U.S. Census Bureau's reports (www.census.gov).

The most obvious benefit of these sites for searchers is that they are free. But these sites are all individually maintained on the Web. It is impossible to aggregate searches, so researchers must go from site to site, looking for the information they need. Also, free Web sites usually do not offer substantial archives of information — an article available one week may be gone the next.

Fee-based Web sites

Fee-based Web sites offer published articles or high-value information, such as trade statistics or market research reports, which may not be available on the free Web. They often have preformatted collections of material, such as company profiles, and usually provide an archive of materials dating back a year or more. Although they have a better selection of material than free Web sites, they are still limited in focus. These sites charge for information — either a monthly or annual subscription fee or a fee each time a user downloads a document. A drawback of fee-based Web sites is that they encourage ad hoc purchases of information, and researchers must often provide personal credit cards for payment. Most fee-based Web sites do not offer the ability to pay by purchase order. Most offer only rudimentary search tools; output is usually limited to plain text.

Examples of fee-based Web sites include Hoover's Online (www.hoovers.com), the U.S. Department of Commerce STAT-USA service (www.stat-usa.gov), *The Wall Street Journal* (www.wsj.com) and MarketResearch.com

Value-added information services

Value-added business information services offer high-quality information from a wide variety of sources including newswires, newspapers from around the world, trade magazines, newsletters, investment house reports and more. They have extensive archives of periodicals and historical financial data.

These services offer simple but powerful tools that enable researchers to search hundreds of sources — current information and archives — simultaneously. Results can be downloaded, printed or e-mailed. Electronic clipping features keep business researchers informed about topics they select as the online sources are updated.

Value-added information services charge fees, which can be paid within a company's purchasing process. The blanket arrangement can provide access to a specified number of employees within an organization.

The three primary providers of value-added information are Dialog (www.dialog.com), Factiva (www.factiva.com) and LexisNexis (www.lexisnexis.com).

Are resources being spent wisely?

Time — meaning employee work time — and money are two of any organization's most important resources. When workers search for information online, there is an associated cost in both.

One of the prevalent misconceptions about business research is that most business information is available on the Web for free. Granted, often the only direct expense for looking for open Web content is the not-insignificant expense of providing high-speed Internet access to the desktop enterprise-wide and maintaining the firewalls and security features necessary to prevent the spread of viruses, Trojan horses and other "mal-ware" within an organization. However, there is also a significant cost in productivity when a decision-maker is forced to spend hours going through marginally useful information on Web site after Web site. And this does not even take into account the time spent fruitlessly looking for deep Web information through search engines. Using IDC's estimate that 50 percent of searchers' time is spent doing unsuccessful searching, this assumption that most business information can be found through search engines can be extremely costly to organizations. Information that seems to be free is not, in fact, free to the enterprise.

Skilled or muddling through

Log on, plug a few words into a search engine, and bingo — thousands of references pop up. How difficult can it be?

That, essentially, is the attitude of most of the people interviewed in the *Super Information About Information Managers (Super I-AIM)* study conducted by Outsell, Inc., which was commissioned by Factiva, Dialog and KPMG. A whopping 85 percent report that they are skilled or very adept at searching for external information. But at the same time, 60 percent of the study respondents say that having sufficient training is a problem.

Obviously, the findings are contradictory. The reason may be that searching for information online appears to be easy enough, and most attempts to find information probably result in at least partial success. Yet everyone knows the frustration of visiting Web site after Web site looking for something of value, and thinking there must be a better way.

The illusion of easy access

Sixty-two percent of *Super I-AIM* respondents feel that anything is available online, and with 50 million Web sites out there, it's no surprise. But 62 percent of respondents also say external information is too hard to find.

The difficulty in finding information lies in the anarchy of the Internet — every Webmaster is free to do what he or she wants. Freedom is a wonderful thing, but on the Web, it certainly makes finding a particular fact difficult.

With no consistency or standardization of business information on the free Web, it cannot support efficient research. A simple query often returns hundreds of thousands of hits. Some are useful, some are keywords in a completely different context, and some are "spamdexing" — keywords a Webmaster plugs in specifically to grab the attention of search engines, whether or not the site is relevant.

Add to that the problem of the deep Web – the content that exists within databases, hidden behind registration pages, stored in formats not accessible to search engine spiders, or dynamically generated – and it is not surprising that most searchers find the search experience so frustrating.

What appears to be a wealth of information is an illusion. The lack of standardization on the free Web, combined with the inability to aggregate searches, means that much of the information is useless, and the useful information is buried in the clutter or hidden where search engines cannot find it.

Adding up the expenditures

The free Web is free — so far, so good.

But sometimes knowledge workers need proprietary information from fee-based Web sites. Usually, the only way to download an article from fee-based sites is to submit a credit card number. Workers often charge their personal cards, and then submit the expense for reimbursement. Once the time and money required to process the reimbursement payments are factored in, the cost of information goes up again.

So, when it comes to searching for external information, are the organization's resources being spent wisely? The answer is far more complex than the Internet access bill.

Is the right information being collected?

As stated above, 62 percent of respondents in *Super I-AIM* believe anything is available on the Web. But here is another of the study's findings:74 percent believe it's hard to determine what is available online.

The contradiction highlights the difference between perception and reality. The free Web appears to be bursting with information. But the truth is, most of the information business people want and need is not available on the free Web. That is the primary finding of a recent survey I conducted of the top publications in ten industries, ranging from accounting to energy, pharmaceuticals to telecommunications.

Methodology — examining the sources

This survey examined the top five sources from each of ten of Factiva's major industry categories:

- accounting/consulting
- banking/credit
- computers/electronics
- energy
- insurance
- Internet/online services
- investing/securities
- media

- pharmaceuticals
- telecommunications

The first step was to determine whether each source, or its publisher, maintained a Web site. If a site was located, it was checked for:

- Availability of full-text content without a fee
- Access to all current articles or only selected articles
- Availability and extent of content archives

Result — the free Web comes up short

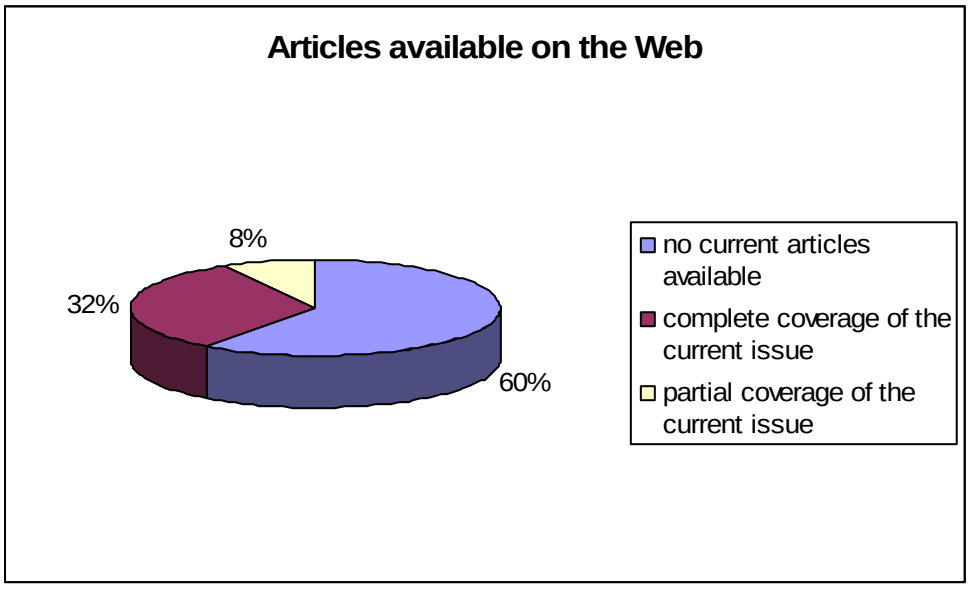
More than two-thirds of the publications studied either do not make the full text of any articles in the most current issue available on the Web for free or only make selected "teaser" articles available for free.

This research proves, therefore, that the information managers surveyed in the *Super I-AIM* study, who believe anything is available on the free Web, are simply mistaken. High-quality business information is not nearly as available as they think.

For organizations that must stay on top of international breaking news, relying on the free Web is even riskier. Most newswires, which are fee-based services from leading news agencies around the world — are not available on the free Web at all. On the other hand, Factiva distributes more than 90 international newswires, more than any other service.

Following is a summary of the findings:

Figure 1



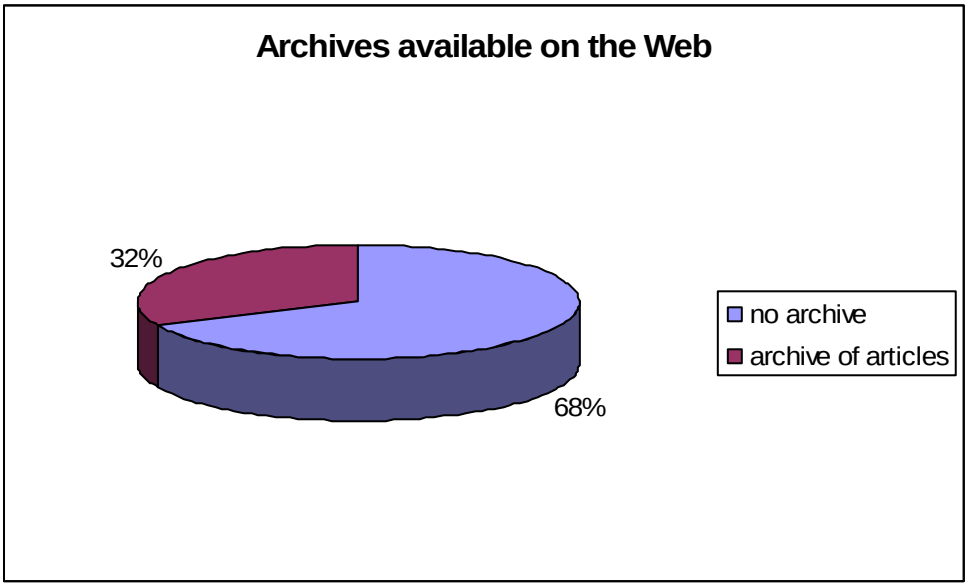
As this research demonstrates, only 40 percent of the titles examined provide at least some current information on the Web for free. But access to the information must also be considered — it's still on a source-by-source basis; these articles are not indexed in search engines so the business researcher is forced to look at individual Web sites, one at a time, to find pertinent

information just from the current issue of selected publications. The search becomes even more difficult if a researcher needs both current and archival information.

When it comes to archives, an astounding majority— 68 percent of the sources studied — do not have a free, publicly accessible archive of articles on their Web site at all.

The reason for this is commercial. Business publishers know the value of both their current information and their archives for business research, particularly for users in the financial services and consulting industries. Therefore, they would much rather make money by selling the information through value-added information services than give it away for free on the Web.

Figure 2



The surprising conclusion of this survey is that less than a third of the key industry sources available through Factiva offer the full content of their current issues on the Web for free or archives on the Web that date back more than a few months. And again, business researchers using the Web must visit each publication's Web site individually — a time-consuming proposition – and hope that the publication makes its archives available for free.

The free Web, therefore, is seriously lacking in important business content, and the information that is available is difficult to access. When knowledge workers search only the free Web for information, it is likely that they will fail to turn up critical facts.

In short, it is very possible that the right information is not being collected. Any manager making decisions based only on information that is publicly available on the Web may be putting his or her enterprise at risk.

Is the information collected leading to good decisions?

The whole point of looking for external information is to be able to make good decisions. Finding appropriate information can be problematic.

Additional research indicates that when information is found, its quality may be unreliable. Most of the people interviewed in *Super I-AIM* — 72 percent —say it is difficult to determine the quality, credibility and accuracy of external information they use in their jobs.

One way to verify information is to crosscheck it. But because searches of the open Web cannot be aggregated, crosschecking is inconvenient to the point of being impossible. In *Super I-AIM*, 62 percent of respondents say that not being able to compare across information alternatives is a problem.

So, knowledge workers are spending a great deal of time and corporate resources searching for information that may or may not be available. If information is found, it may or may not be accurate. Is this external information being used to make decisions?

The answer is yes. Despite the shortcomings of free information on the Web, an astounding number of people who responded to the *Super I-AIM* study say they rely on it for decision-making. Following is a comparison of the attitudes of American workers toward free versus fee-based information:

Figure 3: Free vs. Fee Online Information Services

Question about free vs. fee service	Free Services	Fee-based Services
Which service supplies higher quality information?	44%	47%
Which service provides information that helps you make mission-critical or high-risk decisions?	38%	52%
Which service provides information that you use to make daily decisions?	74%	19%
Which service provides information that you use and trust without verification?	50%	41%
Which service contains information from credible and known sources?	53%	39%

Source: Outsell, Super I-AIM

A conclusion of *Super I-AIM* is that "the Internet has given end users the perception that all information is available for free." Given the problems of free information on the Web — much of it doesn't exist, some of it is time-consuming to find and the rest cannot be verified — managers who want to be sure they are making good decisions should seek stronger, more accurate sources of information. Value-added information services are the solution.

Anatomy of an effective online information source

Business managers have always needed external information. The Internet is simply a technology to deliver it immediately to the desktop.

What are the requirements for an effective online information source? I examined the question in the Factiva White Paper, *Selecting Business Intelligence Sources: The Public Web vs. Value-Added Online Services, August 1999*, and the findings still apply. Following is a summary of my study:

Content requirements for Web-enabled business information resources

• **Reliability and authority.** Information on which decisions are based must come from authoritative, reviewed and edited sources.

• **Updated and archived.** The online resource must provide timely access to the most up-to-date information, along with extensive archives so that information can be retrieved later and so that research can be conducted on material from six months or a year ago.

• **Aggregated information.** All relevant sources should be aggregated and searchable within a single interface.

• **Access to the best Web sites.** Online services should offer important sites that Web search engines cannot access, and distinguish between truly useful sites and those that have little value.

• **Full selection of information.** Newswires, industry newsletters, daily business press, trade journals, industry analysts' reports, historical financials and more — all should be available to search simultaneously.

• **Ready-to-download information.** Information should be in a format that is easy to download, e-mail or print.

• **Updating feature.** Most business researchers have an ongoing need for information on specific topics. Electronic clipping services, in which all relevant new items are e-mailed to subscribers on a daily or weekly basis, are extremely helpful.

• **Auditable payment.** Business information should be easy to pay for within the parameters of company purchasing processes.

• **High value to cost ratio.** The relationship between the direct expense of the information and the time involved in obtaining the information should be favorable.

Factiva meets the information needs of business researchers

Factiva is a value-added information service with a long tradition of journalistic excellence, global coverage and advanced technology. The company was formed in 1999 when Dow Jones and Reuters combined their interactive business intelligence services, Dow Jones Interactive and Reuters Business Briefing.

As an information provider, Factiva offers:

• **Access to nearly 9,000 global news and information sources.** Content is reviewed by subject-matter and industry experts. It is selected for inclusion by an editorial staff whose function is to provide relevant and authoritative information. Nearly 9,000 sources from 118 countries in 22 languages are searchable in Factiva, including more than 190 continuously updated newswires. And, more than 900 sources are available on *or before* the date of publication. Sources include

- National, regional and local newspapers
- Trade publications
- Industry newsletters
- Business newswires
- Press release wires

- Media transcripts
- News photos
- Business-rich Web sites
- Investment analyst reports
- Market research reports
- Country and regional profiles
- Company profiles
- Historical market data

• **Extensive, searchable archives.** Factiva archives date back years — for some publications, 15 years or more. Few free Web sites have archives at all.

• **Aggregated searches through a single user interface.** Factiva provides the tools to search the nearly 9,000 sources simultaneously, current information and archives, rather than requiring business researchers to go from one Web site to the next. Search results are presented in a single list, sorted by relevance or date.

• **Systematic indexing of all Information.** Factiva has developed a proprietary indexing and topic hierarchy which comprises more than 820 industry topics, 520 news subjects, 340 geographic labels and more than 310,000 company codes. All information sources are meta-tagged on a document-by-document basis, enabling users to retrieve information regardless of the wording or language of the original text.

• **Choice of download formats.** Factiva offers consistent download options that enable users to receive and use information in the format they prefer — ASCII, XML, RTF or HTML.

• **24/7 customer support.** With offices in nearly 30 countries and customer help desks covering Europe, Asia and the Americas, Factiva offers plenty of customer support. Assistance is available in six languages via e-mail or telephone.

• **Financial control of information services.** Factiva enables enterprise information managers to determine the number of users, types of resources available and the overall budget for information services through a single contact. Fees are paid through the company's regular payment processes in local currency.

• **Content integration into intranets and portals.** Factiva offers a suite of products designed to provide customers with the best technology for integrating world-class news and business information into their intranets or portals. Workers can stay up to date on topics they select through Factiva's clipping option.

These features and services add up to a robust solution for any organization's information needs.

Factiva — solid support for better decisions

Factiva offers business managers a wide range of reliable, timely information from the most important and reputable sources in the world. The Factiva selection process provides researchers with not necessarily the most business sources, but the best business sources. Factiva's online interface makes access to that information easy and efficient; search results are rich and comprehensive. Therefore, Factiva provides managers with the facts they need, when they need them.

Judging from the responses of people who use Factiva, this value-added information service builds upon the primary purpose of information: Factiva's products, learning programs and support services help users to not only make decisions, but to actually make decisions that are more informed and more effective.

About Factiva

Factiva®, a Dow Jones and Reuters Company, provides world-class global content, including Dow Jones and Reuters newswires and *The Wall Street Journal* - unduplicated in a single service elsewhere. Factiva offers the only single content solution with multiple language interfaces and multilingual content covering 9,000 sources.

Factiva's products and services help companies integrate news and business information into their daily workflow to increase organisational intelligence and leverage external and internal content within the knowledge management function. Factiva's content management and integration services are used by leading organizations around the world.

Built on industry standards and open architecture, Factiva products deliver flexible, extensible, customisable solutions to enable easy integration and use in the enterprise. Editorial and technical consulting, taxonomy application, integration expertise and e-learning programs reflect Factiva's innovative approach to delivering solutions beyond the content.

For more information visit: www.factiva.com or contact us at www.factiva.com/moreinfo.

About the Author

Mary Ellen Bates is the owner of Bates Information Services, Inc., based in Boulder, Colorado, providing business research to business professionals and consulting services to the information industry. She can be contacted at mbates@BatesInfo.com.

Bibliography

Bates, Mary Ellen. "Factiva 2002 White Paper Series: Free, Fee-Based, and Value-Added Information Services." White paper. April 2002. Edited by Donna Andersen.

Bates, Mary Ellen. "Selecting Business Intelligence Sources: The Public Web vs. Value-Added Online Services." White paper. August 1999. [www.factiva.com/infopro/BusIntellletter.pdf]

Bergman, Michael, "The Deep Web: Surfacing Hidden Value" *Journal of Electronic Publishing,* vol. 7, no. 1, August 2001. [www.press.umich.edu/jep/07-01/bergman.html]

Cochrane, Nathan, "The Power of Search", *Sydney Morning-Herald*, April 13, 2004, p.1.

IDC, *The High Cost of Not Finding Information*, April 2003.

Lawrence, Steve and C. Lee Giles, "Accessibility of Information on the Web", *Nature*, no. 400, p. 107, July 8, 1999.

Netcraft, *April 2004 Web Server Survey*, April 2004.
[news.netcraft.com/archives/Web_server_survey.html]

Sherman, Chris and Gary Price. *The Invisible Web: Uncovering Information Sources Search Engines Can't See*. CyberAge Books, 2001.

Super Information About Information Managers (Super I-AIM). Conducted by Outsell, Inc. Sponsored by Dialog, Factiva and KPMG. 2001.