Dialect

Keeping the industry better informed

InfoSort

Dialog

Profound

DataStar

Enter the Dragon

Maintaining a competitive edge in Asia

Live Intranet

- ## Drug induced vision
  The pharmaceutical industry in the 21st century

- ## 66% of the web is missing
  Why do search engines only cover
  a fraction of the internet?

66% of t

*By David Green,
Head of Publishing at
Informed Business Services*

RANK	NAME	URL	% COVERAGE
1	Hotbot	www.hotbot.com	34%
2	Altavista	www.altavista.com	28%
3	Northern Light	www.nlsearch.com	20%
4	Excite	www.excite.com	14%
5	Infoseek	www.infoseek.com	10%
6	Lycos	www.lycos.com	3%

Traditionally, retrieving external information was the domain of information professionals who primarily used paid-for online databases with sophisticated search languages. The ascendancy of the Internet, and in particular, the web, has resulted in an explosion of alternative information sources that are currently relatively cheap or free. However there are key differences between these sources.

Information is increasingly accessed via the web, not on it

Database integration involves the user visiting a web site and retrieving the information they require from a database, using what is often a rudimentary search language. This information is presented in a temporary, computer generated web page. Therefore search engines can't find the valuable information on these temporary (or dynamic) web pages and the information remains hidden in the database until retrieved by the user. Also database web site integration is expensive and online advertising has not been recouping the development costs. Consequently there is an increasing trend towards charging for information on the web.

Nobody ever searched the web

When using a search engine you do not 'search the web'. In fact you are searching a database of indexed web sites. These databases have been compiled by programs called 'spiders' which search the web for new web pages. However, they cannot keep up with the phenomenal growth of the web. Research published in the April 98 edition of the journal 'Science' revealed just how much ,or rather how little, of

the web
is missi g

Why do search engines only cover a fraction of the internet?

the web each of the major search engines cover - see table above.

Web directories

So why doesn't Yahoo! feature in this table? Because it is not a search engine! It is a web directory. Unlike search engines which are compiled automatically by computers indexing keywords on web pages, web directories are compiled manually by human editors. They are pre-defined lists of web sites which are categorised by subject. However, as they are compiled manually, web directories only cover a fraction of what's available on the web.

Also, inclusion in a directory is often entirely at the discretion of the editor(s) - so someone else determines what constitutes a useful web site - which may not always be what you want.

Search languages

Online search command languages are currently more powerful than their web search engine cousins. Not only do they offer a greater range of options for identifying and retrieving the information, they also allow the user to manipulate the results in ways that search engines just can't. RANK and SORT would be examples of such commands on Dialog. These issues have several consequences:

- There is an increasing need to manage the burgeoning choice of sources and to quickly identify which is most appropriate
- The expansion of information retrieval skills beyond the exclusivity of information professionals to end users (high volume consumers of information who are aligned to non-information functions within the organisation).
- Managing subscriptions to a myriad range of web sites versus single integrated invoicing for online databases.
- Cost of information versus cost of time - there is a greater appreciation of the speed and power of online databases.

The web is a compliment to, not a replacement for, commercial online databases. Time has a premium value in today's business environment and "free information that takes too long to find and format is expensive information". (Information Today, Feb 98).

Links to over 750 useful web sites for researchers, plus a range of in-depth articles on the information industry, are available for free from the Reviews section of the Informed Business Services web site at www.informed-ibs.com