

Intelligent Web Searching for the Biosciences

Finding high quality information on the World Wide Web to support academic work can be a time consuming and difficult task. This guide gives practical advice on locating and evaluating high quality web resources in the biosciences.

Getting Started

The World Wide Web (WWW) may be accessed using Internet Explorer 6. To open the browser, click on the **Start** menu then select **Networked Applications, General Software, Internet, Web & Email, Internet Explorer 6**

The first document displayed will be the Cardiff University homepage (www.cardiff.ac.uk). You may access other sites or search engines by entering the URL (web address) into the Address box at the top of the screen.

Searching the Web

Searching the Web can be a hit or miss affair for the following reasons:

- The Web is unregulated - there is little attempt at quality control
- Search engines usually provide an unmanageable number of "hits" and vary greatly in how effectively these are ranked
- Search engines are not geared towards academic users and lack the functionality and precision of databases. When searching, you are guessing which words appear in the pages you want or which terms were chosen to organise a site covering a topic. This makes searching extremely unpredictable and imprecise.

More about Search Engines

It is important to remember that when using a search engine you are searching a relatively small subset of the Web. Search engines use

automated programmes called "spiders" and "bots" to crawl the Web hunting for pages to include in their databases. There is no human intervention and hence no quality control. When using *Google* or *Yahoo* for instance, you are searching a database of web pages selected from the billions of pages which exist on servers around the world. Search engines vary not only in the extent of their coverage of the Web, but also in the methods used to index and retrieve pages and the accuracy with which they interpret your search. This accounts for the variation experienced when using different search engines. You may check the extent of overlap/uniqueness between the results of two search engines by running a single search on *Thumbshots* (ranking.thumbshots.com); this analyses your results indicating the number of common and unique sites from the first hundred hits on each engine).

Which search engine is best?

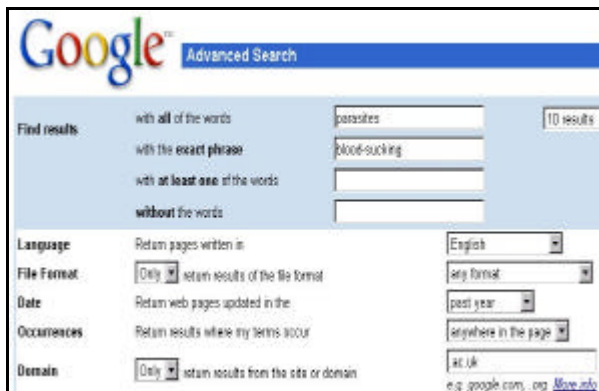
In recent years, *Google* has been regarded as the leader in the field. It currently boasts 8.2 billion web documents, a clean uncluttered layout and excellent ranking capabilities. When you perform a *Google* search the results are ranked according to how many other sites/pages link to a retrieved site. It also takes into account other factors such as the proximity of your keywords within the documents. This ensures greater relevance than some competitors.

However, search engines are gradually increasing in sophistication and most have greatly improved in order to compete with *Google*. *Yahoo* (search.yahoo.com) is by far the biggest search engine with 19.2 billion documents. It compares well with *Google* in terms of ranking capabilities

and currently offers the best image search on the Web. Other recommended engines include **Ask** (formerly *Ask Jeeves*) (www.ask.com) and **Exalead** (www.Exalead.com/search) which is the only engine which allows you to truncate search terms as if you were searching a database e.g. cell* finds *cell, cells, cellular* etc. Rather than make exclusive use of *Google*, it is perhaps time to explore other search engines which have many excellent and complementary features. Up to date information on search engine developments can be found at *Search Engine Watch* (searchenginewatch.com).

Using Advanced Search Options

The most effective way of using search engines such as *Google* and *Yahoo*, can be to use the *Advanced Search* option (a link will usually be provided on the main search screen). Apart from options enabling proximity searching, you are usually able to focus your search by applying filters similar to those found in a bibliographic database. Most allow you to limit your search by language, date and domain. By using the latter option you are able to restrict your search to sites which originated within the UK academic community by specifying *.ac.uk*



Advanced Search options can greatly improve the efficiency and relevance of your search and it is worth familiarizing yourself with the options available. Search engines usually provide **Help** pages providing a wealth of information offering search tips and suggestions. It is worth checking these out to ensure that you are using the best possible search techniques.

What exactly is Google Scholar?

Google Scholar (www.scholar.google.com) is *Google's* much publicised search engine which aims to provide results drawn from academic and scholarly literature. As well as finding web sites, it retrieves results from the “hidden web” such as abstracts from databases, references and links to peer-reviewed papers, theses, books and information from numerous scholarly sources.

Unfortunately, *Google Scholar* is not a “one stop shop”. Many of the items found will only be accessible to subscribers and its ranking capabilities are inferior to those of its parent. *Google Scholar* is in a relatively early stage of development and is not an alternative to using databases as coverage of scholarly material can be patchy.

Meta Search Engines

Meta-search engines such as *Dogpile* (www.dogpile.com) and *Trovando* (www.trovando.it) transmit your search simultaneously to several individual search engines. Within seconds, you get back results from all the search engines queried. They are sometimes referred to as *parallel search engines*, or *mega search engines*.

Meta search engines are useful when you want to retrieve a relatively small number of relevant results. They are excellent when searching for obscure subjects and when you want to get an overall picture of what is available on the Web on your topic. On the down side, their use is limited primarily to simple queries, little or no field searching is available and they return a limited number of results that do not represent the totality of results from any single source engine. Some meta search engines such as *Dogpile* allow you to view the top 10 results from each interrogated search engine at a glance and will also highlight those results which are unique to each.

Subject gateways

These are highly recommended for finding information for academic work. Unlike search

engines they are selective and make a serious attempt at evaluating the quality of internet resources. Searching via a gateway will yield a small number of results but you are guaranteed that the sites retrieved will be of a high standard.

Intute Health and Life Sciences

www.intute.ac.uk/healthandlifesciences/ is a subject gateway created by a network of UK universities and partners. Subject experts and information specialists select and evaluate the websites and write reliable descriptions of their content. The complete Intute database contains around 114,000 records. You can also browse or restrict your searches to one or more of the nine sub-gateways which include Medicine, Bioresearch and Natural History. To search, enter your terms into the Query box then click on the **Search** button. *Advanced Search* offers greater functionality allowing you to limit your search to particular kinds of web resources e.g. images, lecture notes, professional organisations etc.

Intute will retrieve a list of relevant web sites with accompanying abstracts which have been selected using strict quality criteria.

Medical search engines

If you are searching for resources in the medical sciences, as an alternative to the general search engines you could try a specific medical search engine. These search across a range of websites within the medical sciences and can sometimes provide a more focussed result. Examples of these include *MedHunt* (www.hon.ch/MedHunt) and the evidence-based search engine *TRIP* (www.tripdatabase.com).

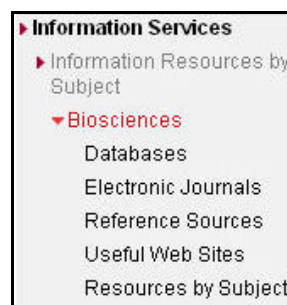
Finding Images

For the biomedical sciences, try using a specialised image site such as *Bristol Biomed Image Archive* (www.brisbio.ac.uk), and the Karolinska Institute's *Medical Images and Illustrations* site (www.mic.ki.se/Medimages.html). For the biological sciences *Biological Sciences Image Galleries* (www.academicinfo.net/bioimage.html) is a useful resource. However, many general search engines provide image searches. *Yahoo's* image search (images.search.yahoo.com) is one of the best on the Web.

Cardiff University Information Resources for the Biosciences

The Library has collected together links to the most important internet resources in the biosciences. These may be found at: www.cardiff.ac.uk/schoolsanddivisions/divisions/in-srv/bysubject/biosciences/index.html

Use the navigable list to the left of the screen to choose your options.



The options are:

- **Databases** Providing links to key bibliographic databases including *Web of Science*, *PubMed*, *CSA* and *BIOSIS*. Use these for tracing details of journal articles on your topic
- **Electronic Journals** Links to information regarding access to ejournals including a link to the Cardiff University Ejournal database
- **Reference Sources**. Links to online reference works such as *Encyclopedia of Life Sciences* and to information concerning printed reference materials in the Libraries
- **Useful Websites**. Provides links to subject gateways and directories of

internet sources, web searching tutorials, universities and research sites and the websites of key local and national library resources

- **Resources by subject.** A list of sites arranged under broad subject headings. Includes links to sites in anatomy, biochemistry, botany, genetics, molecular biology medicine and neuroscience.

Evaluating Web Resources

It is your responsibility to establish the validity, authorship, timeliness and integrity of what you find before you use that information. Aim to use only sources which are of a comparable quality to reputable textbooks or journal articles.

Much can be inferred about a site by looking closely at the URL (web address). For example:

<http://www.research-councils.ac.uk/>

The URL is prefixed with an acronym for the *protocol* used for the transfer of information across the internet. In this case, *http* (hypertext transfer protocol). The URL sometimes begins with the letters *www*. This is followed by the source name (the publisher, agency or 'server' hosting the document). The URL also indicates the domain (server type) on which the site is hosted:

- **.ac.uk** for UK academic sites, e.g. www.cardiff.ac.uk
(All countries except the US have a two letter code which usually forms part of the URL)
- **.edu** for academic sites in the US, e.g. www.nyu.edu
- **.org** for non-profit organisations, e.g. www.rbgkew.org.uk
- **.gov** for official government sites, e.g. www.environment-agency.gov.uk
- **.co** or **.com** for commercial organisations, e.g. www.vicks.com

Before using information from the Web, ask yourself the following questions:

- **What kind of server is it hosted on?**
Be aware that information provided on a **.co** or **.com** site is there for commercial not educational purposes. If the page is a

mask for advertising, the information might be biased. Try to stick to the preferred domains (*.ac.uk*, *.edu*, *.gov* & *.org*).

- **Where has the information come from?**
Information from educational institutions, research councils, and professional organisations should be more reliable than that published by unqualified individuals. It is more likely to have been submitted for rigorous checks for accuracy.
- **Who has provided the information and why?**
If it is a personal page, the information is not necessarily bad but you will need to check out the credentials of the author. Even if the author is an academic it is worth asking if he/she has published widely in that particular field.
- **Are details of named authors included?**
If there is no personal author look for an agency or organisation that claims responsibility for the page. If you can't find out who wrote it – don't use it.
- **Is a publication date given? Is the information current enough?**
- Remember, undated factual or statistical information is no more valid than anonymous material.
- **Is it maintained regularly?**
- Check that the page has been updated recently. If not, the information may be out of date. This will also indicate if the author is still maintaining an interest in the information or has abandoned it.

For information on citing web-based sources see the Library's *Citing References* guide.

For further information or assistance contact the Science Library

Tel: 029 2087 4085

Email: SciLiby@Cardiff.ac.uk

SCI-021-Q-1 / August 2006 / NM