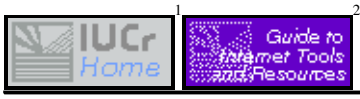

Search Engines

The Internet without **Search Engines** is like a large warehouse without a manager, where you find TV's next to vegetables, and where there are no signs or directories to help you find what you are looking for. *One could wander indiscriminately and make many useful discoveries. Just as easily, though, hours could pass with nothing of value to show for the effort.**[3]

While it is impossible to sort the Internet itself, *e.g.* moving all crystallographers in the world to the same place, next to chemists and physicists, it is still possible to put in signs and compile directories.

This idea is challenging enough to give rise to more than just one independent attempt to put it into reality. Actually, it is pretty hard to put together a comprehensive list of Search Engines, and no attempt is made here to do so. Instead, three outstanding services – *Yahoo!*, *AltaVista*, and *WhoWhere?* – are presented.

These three alone are by far sufficient to meet most needs you will ever have when searching for specific information on the Internet. Furthermore, starting with *Yahoo!* or *AltaVista*, you will easily discover numerous other ways to get what you want – *if* present somewhere on the Internet.

- *Yahoo!* **– A Hierarchical Search System**[4]
- *AltaVista* – Non–Fiction Hitchhiker's Guide to the Galaxy[5]
- *WhoWhere?* – An E–mail Index[6]

Quality of Search Results

The quality of the search results is of course strongly correlated to the way the databases are built. On the one hand, *Yahoo!* is an *edited* directory tree constructed with human intervention. This means, the entries are added in a reasonable way and presumably checked for consistency.

However, the Web is a rapidly evolving institution, and it is simply impossible for humans to keep all links up to date. For example, the link titled "Index – Crystallography – WWW Virtual Library" in the *Yahoo!* subdirectory "Science – Physics – Crystallography[7]" currently points to a page which just tells you that the page has moved. While this is not a catastrophe, you could well have received a message like "The requested URL was not found on this server".

On the other hand, the *AltaVista* and *WhoWhere?* databases are automatically built by so–called *Web Robots*, also known as *Web Crawlers, Wanderers, and Spiders*. Comprehensive information on Web Robots[8] is available elsewhere on the Web.

The enormous amount of information collected by the Web Robots can not be digested

[1] http://www.iucr.ac.uk/iucr–top/welcome.html
[2] int.index.html
[3] #Cite1
[4] int.yahoo.html
[5] int.altavista.html
[6] int.whowhere.html
[7] http://www.yahoo.com/Science/Physics/Crystallography/
[8] http://info.webcrawler.com/mak/projects/robots/robots.html

by humans in any way. Instead, the raw data are run through more or less intelligent "data reduction" routines before they are finally stored in the database. Presumably, the stored information is validated at regular time intervals in order to remove outdated references. However, even the best existing Web Robots cannot keep track fast enough to avoid the famous "URL was not found" messages.

When you search and finally find e−mail addresses, you will often experience problems like in one of the previous examples. When searching for "h flack" with *WhoWhere?*, there were two e−mail addresses[9], one of which (*flack@cui.unige.ch*) was more or less accidentally released to the public in Howard's postings to the newsgroup *sci.techniques.xtallography* (12th Sept. 1995 and 9th May 1996) and should not be used. But how can you know? −− Still, in these cases your best options are to find the personal Home page[10] or to use the telephone![11]

* Citation taken from the About AltaVista[12] page. The page has changed since then.

[Index][13] [Yahoo!][14] [AltaVista][15] [WhoWhere?][16] – 10th Sept. 1996 – © Ralf W. Grosse Kunstleve[17] – **Not to be copied or reproduced without permission** − Author's current manuscript[18]

[9] int.whowhere.html#WhoWhereHFlack
[10] int.altavista.html#FindHomePage
[11] int.whowhere.html#Telephone
[12] http://altavista.digital.com/cgi−bin/query?pg=about
[13] int.index.html
[14] int.yahoo.html
[15] int.altavista.html
[16] int.whowhere.html
[17] http://xplor.csb.yale.edu/~rwgk/
[18] http://xplor.csb.yale.edu/~rwgk/scnet/search.engines.html