Rustam Al-Shahi

MRC Clinical Training

Fellow, University of

Edinburgh Department

of Clinical Neurosciences

If you want to find anything on the World Wide Web (WWW), you need to know about search engines. Hitherto, both Internet sceptics and 'techno-phobes' have criticised the various search engines for having the same failings as the WWW itself: that the overall quality and relevance of sites are poor. After all, what search engines retrieve for you is merely a reflection of the information available.

However, as the WWW expands and consumerism drives its improvement, search engines are also becoming better at finding the most relevant high quality information available. The challenge now is to identify the best of the profusion of search engines!

At the outset, it is important to understand the design of these utilities. They may be **search engines** (e.g. www.google.com) which scour the Internet itself for sites of relevance, **directories** of sites posted by webmasters and reviewed by the authors of the directory (e.g. www.yahoo.com), or **hybrids** of the two (e.g. www.altavista.com). This inevitably means that search engines are more likely to offer what you want. Search engines are composed of three parts: the device that searches the WWW (variously called 'spiders', 'crawlers' or 'robots'), the index of what the engine has found on the WWW, and lastly software
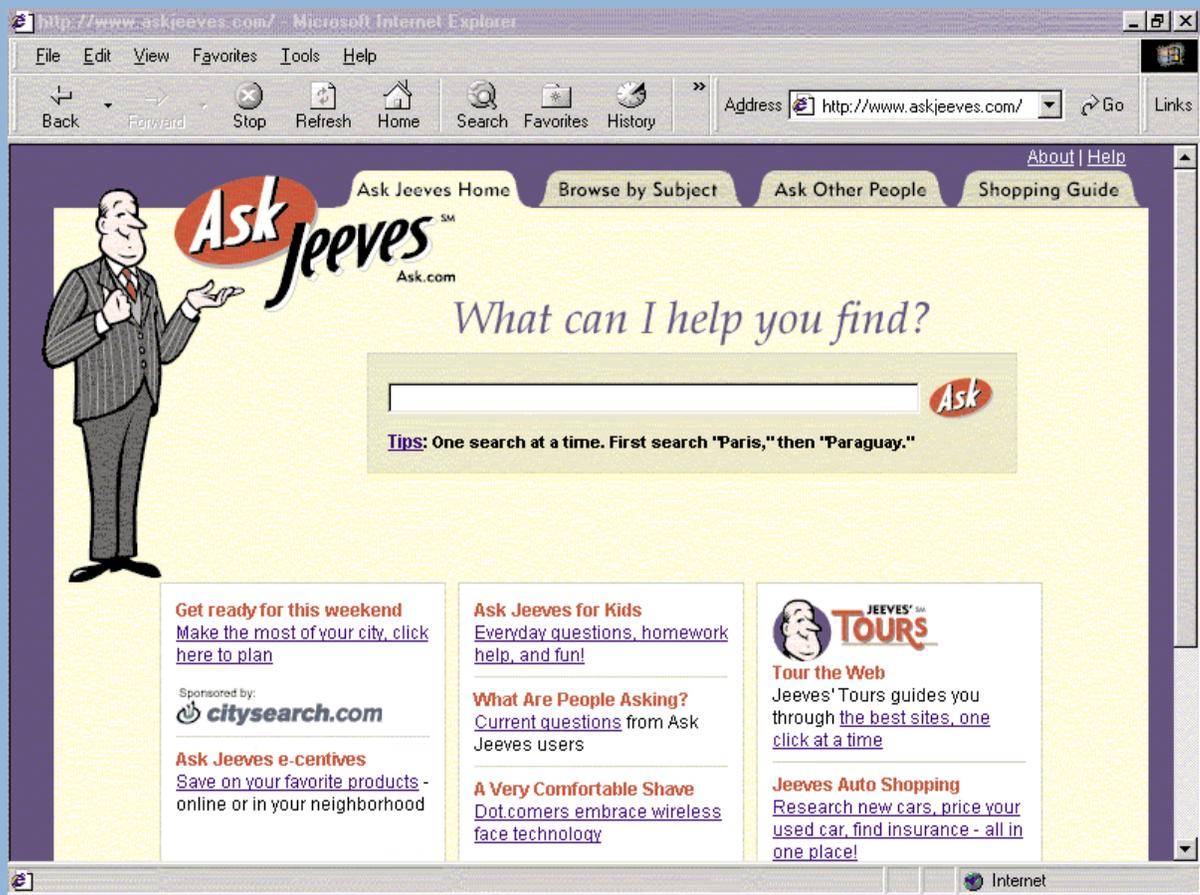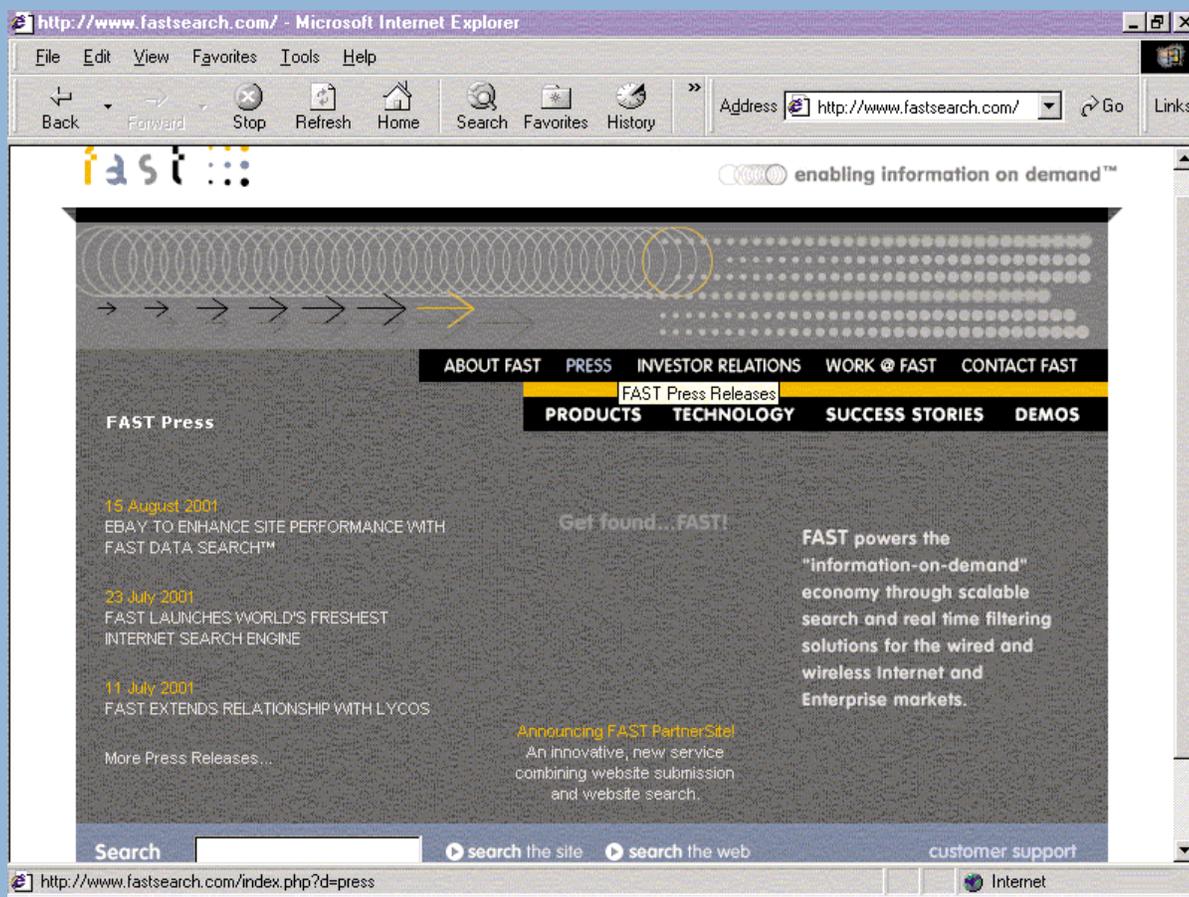
Search engines

which ranks the sites it finds for you according to their relevance to your search criteria. Whilst the greatest discriminator between the engines will be the extent of their spiders, and so the size of their index, the ranking software is shared by some engines, making their results almost identical for the same search.

There are two reasons why many search engines have now improved. Firstly, their indexes are more comprehensive, although what proportion of the WWW they cover may never be known, as the size of the WWW is almost impossible to gauge. Secondly, the ranking software they use is increasingly sophisticated. The software lists sites according to the location and frequency of your search terms on a page, the extent to which other sites on the WWW 'link to' the site, as well as whether the site is listed in their directory (in the case of 'hybrids'). Fewer sites now look at the 'meta' tags (invisible pieces of information about a page's content, written by its author).

It is also your responsibility to ensure you enter the terms most relevant to what you are looking for. Every engine offers a box that is tempting to complete the instant you see it, but the increasingly prevalent advanced search capabilities must not be ignored. You may search for particular files (e.g. web pages, sound files and graphics), in certain languages, use suggested keywords and even Boolean commands. This makes searching the WWW increasingly like using 'Medline', albeit a little simpler and quicker!

If you follow these tips, which engines will find your desired site in the first 10 of the plethora of 'hits' they are guaranteed to offer you? Useful up-to-date information is available from a site called Search Engine Watch (www.searchenginewatch.com, accessed 29th May 2001). Because of the way it compiles its index, Google (www.google.com) is by far and away the largest of the engines, closely followed by Fast Search (www.fastsearch.com) and AltaVista (www.altavista.com). They all offer advanced searching capabilities, and it will be personal preference as to which of the three you prefer if you are not swayed by Google's sheer size alone.