

# High-speed searching

The web is an incredible resource but finding what you want can often be down to chance more than skill. Rosemary Haworth takes the guesswork out of online searches and shows you how to access the information you require more quickly

Although every web page in existence has some sort of referencing associated with it, there is no standard way of recording its content so the internet may be an excellent reference tool but, unlike other sources, finding what you want is not a simple matter of looking in an index and turning to the relevant page.

The internet breaks down geographical barriers enabling, for instance, a surfer in Birmingham, West Midlands to find out about job opportunities and courses of study in Birmingham, Massachusetts, as well as anywhere else. Similarly, it means academic researchers are no longer restricted to the books and information

archives housed within their own institution – now they can access data and share it with colleagues across the globe.

Unfortunately, the web's sheer size makes it unwieldy, while the number of disparate locations in which information is stored, and the way it is catalogued, make finding what you want a hit-and-miss affair.

Over the following pages we'll look at the structure of the internet (such as it is) and the driving forces behind it. We'll also take a look at the mechanics of search engines and show you how to make better use of their capabilities, ensuring your search is more likely to end with a crock of gold rather than a load of mouldy old socks.

## Basic keyword searches

A search engine operates by sending out 'spiders', which crawl around the web looking for information and making associations between related words and phrases. Spiders look for keywords or metatags embedded into individual

web pages, which give a basic idea of the information to be found on a specific page.

Suppose you're curious about a particular brand of notebook – we'll call it Brandname Notebooks. The most obvious route is tapping in [www.brandname.com](http://www.brandname.com) or [www.brandname.co.uk](http://www.brandname.co.uk). However, if this doesn't return results, your next port of call may well be [www.pcadvisor.co.uk](http://www.pcadvisor.co.uk), since you know that *PC Advisor* reviews notebook computers.

Type in the keyword 'Brandname' in the Search the site dialog box on our home page and, within a few seconds, any pages on the site containing keyword references to Brandname will be listed on the Results page. Bingo, you now have a news story and two reviews of Brandname notebooks and have only had to type in two or three URLs to get them.

Searches like this are simple and ideal if you have a vague idea of where to find the information you need. Beyond these basic tools, you need to use a dedicated search engine. When you use the search facility on your service provider's home page (the web page that, by default, first appears when you log on to the internet, unless you've specified a different home page setting), it may use more than one search engine to locate the information you request.

For example, BTopenworld ([www.btopenworld.com](http://www.btopenworld.com)), which is shown in the screenshots left, uses LookSmart as its search facility, and also offers Quick Links. Simply choose the topic that interests you – such as IT – and you are taken to a relevant web page within the BT site.



↑ BTopenworld uses a search facility called LookSmart. Alternatively, you can click on the relevant topic under its Quick Links feature – we chose IT...

→ ... and were taken to a relevant web page, with useful features, news, software offers and sponsored advertorials





← You can glean UK-based results in Google by clicking on the Search pages from the UK checkbox

isn't a single, universal search engine tasked with cataloguing and retrieving everything that there is on the web.

### Refined web searching

A site-based search engine, such as BT's trawls through the text contained on its website, including information about the images. But it also uses referrals and references from dedicated search engines about commonly returned results for the terms you entered. You're probably familiar with one or more dedicated search engines such as MSN (www.msn.com), Yahoo (www.yahoo.com) or AltaVista (www.altavista.com).

You may be wondering why you need more than one search engine or why there

The internet is vast and no single search engine, however powerful, could possibly hope to keep track of everything – and keep abreast of any changes. In addition, the web is a competitive, commercial medium, driven by the forces of sponsorship and advertising as well as information cross-referencing and provision.

### Location, location, location

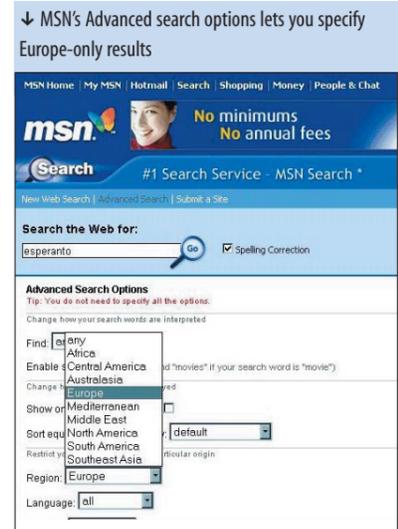
By dint of America's sheer size (as well as the fact that its inhabitants got dirt-cheap internet connections and embraced the web years before us), most English-

language sites returned by search engines will probably be US-based.

If the information you seek is specific to Europe or the UK, tell the search engine this: start by tapping in the name of the search engine you wish to use, using a .co.uk tag rather than .com. Depending on your ISP and the search engine you're using, the search engine may automatically recognise that you're a UK user and offer you the option of finding UK-specific web pages or performing a worldwide search. Type in the keyword(s) for your search and, once the results page appears, you will be given advanced search options letting you refine the content by language, location, domain name and file type.

To specify UK-only results using Google, go to www.google.co.uk and, under the search keyword field box, click on the Search pages from the UK button (as shown in the screenshot above). AltaVista's location setting is similar – directly beneath its search field is a Find Results option, letting you choose either UK or Worldwide. The same is also true of MSN, as shown in the screenshot, below.

Though many search engines recognise terms in other languages, they may not help you understand the content of the pages. You can cut and paste portions of foreign-language text into online translation engines such as Babelfish (www.babelfish.com/translations.html) or SysTran (www.systransoft.com). In our experience, Google is the most user-friendly of the mainstream engines. You can ask it to



automatically translate the information on web pages by hitting the Translate this page option next to the page heading.

### Smooth operators

Single keyword searches usually result in a vast number of pages being listed for you to trawl through. You can narrow your search by using multiple keywords, and marking words you wish to be included with a plus sign, and those you wish to be excluded with a minus sign.

Such terms and markers are known as Boolean operators. The words 'and' and 'not', as well as asking the search engine to look for particular dates (2001 or not before 1999) can help cut down on the number of dud pages you have to scour. Similarly, putting quotation marks around phrases or names to indicate that they represent a single unit can bear fruit.

Phrases such as 'near', indicating that other forms of the keyword can be searched, or, in AltaVista and Google, advanced operating terms – for example, inurl (search words in web addresses) or relate (related terms) – can be helpful, too. For search engine-specific terms, go to the Advanced search or Help options.

If, having tried all these tricks, you're still no closer to getting the information you need, try synonyms for the keywords you enter. After all, a computer can't read your mind and can only go on the information you feed it.

### Natural language searching

Sites such as Ask Jeeves (www.ask.co.uk) use a different tack based on full-sentence queries. For instance, 'Where can I find a plumber in SW19?' will return interpretations of your query, from which you can refine your search to get the exact information you're after. You'll often find cached



## Safe seeking

It's no secret that the internet houses vast quantities of irreverent, offensive and, in many cases, obscene images and movie clips. Many search engines give you the option to exclude sites with spurious content from appearing in your list of search results. Any search engine worth using will provide an option to filter out offensive material, usually with a single click before you begin the search request or as an advanced search option. Sites such as MSN (see the screenshot, right) let you save these settings (using My Preferences, Save My Preferences) for use in subsequent searches.

Depending on your ISP, you may have opted for family-friendly content filtering when you set up your online account. Antivirus and personal firewall software and your PC's security settings can also be used to block any online nasties appearing on your desktop. If you have any of these facilities installed, you don't need additional content filtering when performing online research.

## Top five surfing tips

**Finding that needle in a haystack is easy when you know how. All you need to do is hone your queries and suddenly, search engines will be positively inundating you with useful pages.**

1. Enter phrases or questions **You'll get better results if you enter a few related words rather than a single one. Thus, 'Robin Hood merry men' will be more effective than simply 'Robin'. You may also enter a question: 'Who led the band of merry men?' Most search engines ignore the word 'the' though.**
2. Choose your words wisely **A generic search phrase like 'Sherwood' will yield too many results if you're trying to find out about Robin's famous dwelling place. A better phrase would be 'Robin Hood Sherwood Forest'.**
3. If at first you don't succeed... **If you don't glean what you're looking for on the first attempt, don't give up too quickly. Why not try varying the original query by substituting synonyms or related words? Too many results? Have a go at adding more descriptive terms to make your query more precise.**
4. Quote marks for key phrases **Place quotation marks around specific, verbatim phrases that you want to find exactly as written, such as "invention is the mother of necessity".**
5. Help me operators! **Common tools for linking search words together include Boolean operators 'and', 'or' and 'not', and the + and - signs. These operators, which are usually recognised by most search engines, can help you craft more clearly delineated queries.**

'And' serves to link words that must be contained in each returned result: 'Robin and Marion'. 'Only one of the two words must be found in the results, use 'or': 'Marion or Marian'. 'Not' or the minus sign, inserted before a word, can be used to exclude one or more words that must not appear in the results: 'Robin Hood -bird'. The plus sign used before a word indicates that the word must appear in the search results: '+sheriff +outlaw'.

results from previous users, together with the answers they found most helpful (see the screenshot below).

A number of services such as this now exist, though their partial dependence on a legacy of successfully answered queries means it takes a while for newer search services to build up a useful database.

### Reinvigorating dead links

The sheer volume of web content means that not all of the results returned by your search engine will be current. You may click on a web link and get the response Page not found. This may be due to the website owner failing to update the information recently, the page having moved elsewhere or the search engine's spiders having found a site that effectively no longer exists.

Some search engines, such as Google, get round this by letting you access their own cached versions of the pages in question. To request a copy of the cached web page stored on Google's server, add the word 'cached' towards the end of your search request string.

← Ask Jeeves differs from other search engines, as you have to enter a question. You then define your search by choosing the most relevant returned question, or clicking on one of the listed sites

### Using multiple search engines

Although the six major search engines – Yahoo, AltaVista, Google, MSN, AOL and Lycos – are all pretty effective, if the information you're after is elusive it's worth trying the services of parallel search engines – such as MetaCrawler (www.metacrawler.com). These search a number of general and specialist search engines concurrently, providing more results from more sources. This is handy in case the site you seek hasn't submitted its details to many search engines.

From a business' point of view, such facilities highlight the importance of announcing your online presence. As stated at the beginning of this article, with so much content and so many competing websites vying for surfers' attention, you need to proactively market your site with search engine submissions if you want to be found.

### The power of advertising

For all its seemingly liberal – and liberating – ethos, the worldwide web is still about money. Okay, so we'd all love to say farewell to annoying pop-up ads, but the mighty dollar directly affects the results returned when you perform a web search. Website owners pay search engine hosts in return for favourable rankings – a deal known as search engine optimisation.

This makes sense: how often have you used a search engine and looked through

each and every result? Frequently there are thousands of pages that contain one or all of the search terms you entered, even if you narrow the search criteria using Boolean operators.

Spending lots of money on developing a good-looking site with e-commerce capabilities makes no sense unless you're then going to market the site itself. It's a bit like paying for a TV ad campaign, but then only booking an off-peak slot on an obscure channel. The adverts cost the same to produce irrespective of viewing figures, but you can affect the number of people that see it and, by extension, the number of potential customers.

Get a respectable number of site visitors and you, in turn, can start charging other companies for advertising themselves on your site.

### Political persuasion

You should also be aware that, despite claims to the contrary, the internet is a political tool. After all, information is, and always will be, about power. Some pieces of data are easily and transparently available – simply tap in the most likely web address or enter the most obvious keywords into a search engine and, in a

flash, up pops the website or the information you were after.

Other information, however, may not be quite so forthcoming, no matter how wide or narrow your search and no matter how many seemingly relevant but not-quite-what-you-were-after pages you trawl through. Sometimes, it's down to the person doing the searching; other times it's simply that the person who coded the web pages you're after hasn't metatagged (described with keywords) them intelligently enough. Alternatively, and this is quite common, someone may not want you to find the information in question.

### Gaining access

If you've clicked on a link that appears to be a dead end – page not found or unable to access this page – it could be because the web page the search has found for you is stored in a secure part of the website in question. For example, links to parts of the *PC Advisor* website may be accessible only to registered site users – registration is free, and as soon as you have a password you can access those pages.

Other times, the type of coding (commonly, URLs that end with a .asp suffix, standing for active service page)

ensures you have to manually type in the last part or part of the web page's string. If the URL the search engine returns is lengthy and separated by a number of forward slashes – such as [www.pcadvisor.co.uk/registered/downloads/utilities.html](http://www.pcadvisor.co.uk/registered/downloads/utilities.html) – copy and paste the first few elements of the string into your browser window. You'll find [www.pcadvisor.co.uk/registered](http://www.pcadvisor.co.uk/registered) will get you to the registration form for our site and, until you have done this and are logged in, you won't be able to access the free utilities download section.

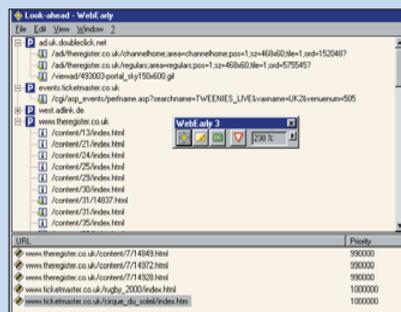
It could also be that the website developer ascribes access privileges to a select group of people and wishes to verify who you are before granting you access. Research companies, for example, may be willing to give away nuggets of information, but expect you to subscribe and pay for the information they have spent time and trouble collating.

Alternatively, the page in question may contain frames or Flash elements, which not all browsers or search engines are able to locate or display. Website owners should know that such fancy formatting can actually be a hindrance to web surfers finding or accessing your site. ■

## Save your online research

**Y**ou can create your own cache of web results by asking the search engine to save the search you've just performed. If your preferred search engine doesn't feature this facility, you can easily lay your hands on inexpensive and often free software applications that specialise in web caching.

SurfSaver, for instance, works in tandem with your browser, running in the background but springing into action whenever you find a site or web page that you'd like to store. Pages are saved complete with the original layout, graphics and formatting, and SurfSaver includes support for frames and retains any hyperlinks contained in the page.



Online researchers will find its ability to save to designated folders and perform full-text and date searches of stored

← WebEarly cuts down on preloading pages by first comparing them to your areas of interest

pages very useful. You can download the free software from SurfSaver's website at [www.surfsaver.com](http://www.surfsaver.com). Alternatively, find it on this month's cover disc.

You can speed up your dialup surfing easily by using WebEarly 3.0, a program

which automatically caches link information contained on the current page (and sometimes the pages those links lead to). This means that, should you decide to click on them, the page appears instantaneously. The more you use WebEarly, the more it learns about your surfing habits, tracking keywords you commonly enter and working out your areas of interest. WebEarly costs £16 from [www.goto-software.com](http://www.goto-software.com), or you can find other web acceleration tools at *PC Advisor's* online Download Directory.



↑ SurfSaver is a browser add-on which lets you store web pages from your browser into searchable folders