# Your guide to better searching on the Net

By Paul Myers

Strand Producer: Research, Audio & Video

# Contents

| Title | Page no |
|---|---|
|
|

# Better Searching

For many people searching the Web is a pretty random process; pick a famous search engine, type in what it is you are looking for and hope for the best.

How often though do we find that the results do not match our expectations? You type in Tony Blair and the search engine suggests a million plus pages including material about Lionel Blair!

When we do find material that suits our purposes it is sometimes difficult to know if the site is biased or objective, professional or amateur. Faulty information can have a profound effect on the eventual quality of our work.

If you follow a few golden rules your searching will be fast, efficient and effective. You will also be confident about the quality of any information found.

## Rule 1 – Choose the correct search method

## Rule 2 – Use the right syntax

## Rule 3 – Refine your search effectively

## Rule 4 – Choose your source carefully

# Choosing the correct search method

In any kind of reference book there are two indexes. The front index lets you find major subjects within the book and the index at the back lets you see if specific words are mentioned within the pages. We instinctively know whether to look at the front of the book or the back. It depends on whether we are looking for a major subject or specific information. Similarly there are two kinds of searches that you can perform on the Net.

The first question you should ask yourself when conducting Internet research is "am I searching for a website related to a certain subject or am I searching for a Web page with certain information?" This is a vital question as it determines which search engine you should use.

You would use a subject search if you were looking for sites relating to a broad subject, a specific institution's website or if you didn't know any keywords related to the subject.

You would use a keyword search for obscure subjects, cross-category searches (e.g. cheeseburgers and Elvis) and specific information.

## Subject Guides or Directories

[http://uk.yahoo.com](http://uk.yahoo.com)

[http://www.dmoz.org](http://www.dmoz.org)

[http://www.vlib.org](http://www.vlib.org)

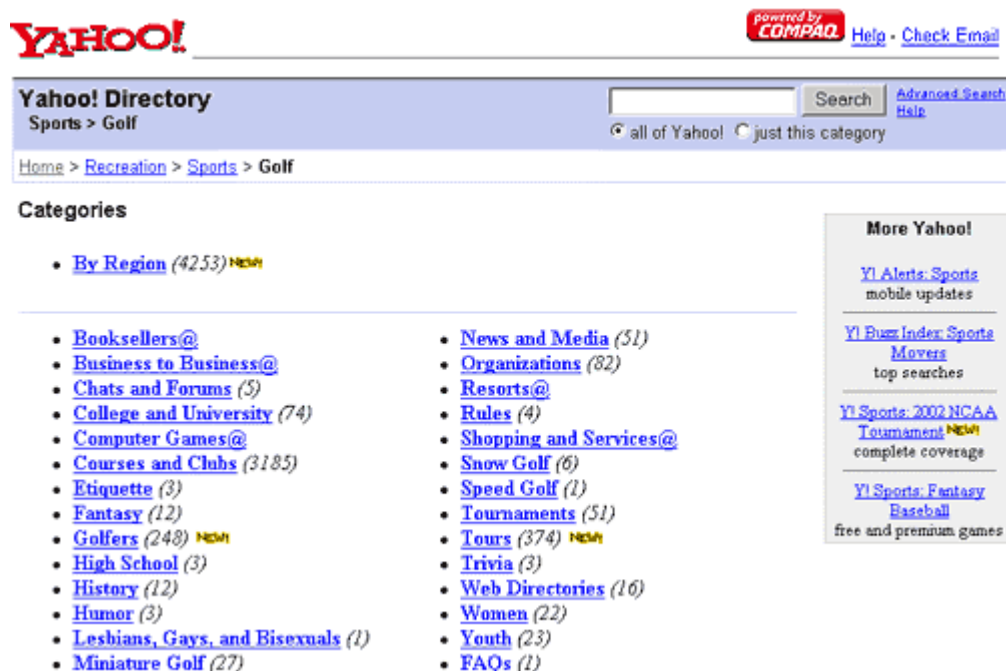[http://directory.google.com](http://directory.google.com)

If I was a golf fan I may find myself typing the word "golf" into Google. It finds nearly 14 million pages but who's to say the best ones are in the top 10? Largely these "hits" are irrelevant pages that just happen to have the word "golf" on them somewhere (e.g. a review of the Volkswagon Golf). Our keyword search is inefficient because we are actually looking for a subject.

If I were to use a "subject guide" like Yahoo and go to their Golf section. I could then take my pick, safe in the knowledge that all the listed web sites are about the subject. I could also browse through sub-categories like "Golf Courses", "Golfing Etiquette" or find golfing magazine websites in the News and Media sub-category. The research is much more structured and eliminates irrelevant pages.

If I wanted to find the name of Nick Faldo's ex-wife, typing "Nick Faldo's ex-wife" into Yahoo would yield disappointing results.

Many people make the big mistake of formulating complex keyword searches in Yahoo. This often ends in failure because Yahoo only knows the words within its category names and the titles, web addresses and descriptions of the websites it lists. If it can't find your keyword in any of these fields it bats your search over to Google to have a go!

Subject guides are compiled by human librarians that review suggested sites and put them into subject lists. This gives great quality control, but inhibits the breadth of coverage - Yahoo has an index of about 2 million sites and it can be months before a new site is indexed. Yahoo is, however an excellent resource if used correctly – as a Yellow Pages of web sites.



## Keyword Search Engines

http://www.google.com

http://www.av.com

http://www.alltheweb.com

http://www.hotbot.com

If you were trying to find specific information like the names of Nick Faldo's ex-wives, It would be better to use a Keyword search engine like Google or Altavista.

These are actually enormous databases indexing the text content of billions of Webpages. You would simply ask them to provide links to any web pages that mentioned the words "Nick Faldo" and "Ex-wife" and hope that the suggestions provide the information you are looking for.

Keyword search engines do not know the subject of the Web pages they index and often you need to refine your search with extra details.

Sports: Time gives Faldo new perspective
... PALM HARBOR -- **Nick Faldo** was Britain's most accomplished 20th century golfer. We
can argue ... His children (ages 14, 11, 7) live with **ex-wife** Gill in England. ...
www.sptimes.com/News/101900/Sports/Time_gives_Faldo_new_.shtml - 50k - Cached - Similar pages

Golf Anecdotes - Great stories, funny anecdotes, golf news, ...
... talent Nick Dougherty is coached by **Nick Faldo**. He has a fitness and mental ... tried
to kidnap his good-looking **ex-wife**. More importantly, perhaps, Stephenson was ...
www.home.zonnet.nl/cveldkamp/anecdotes.htm - 26k - Cached - Similar pages

CNN/SI - Golf - Most experienced Ryder Cup players in deep slump ...
... **Nick Faldo** and David Carter of England are the defending ... to be getting married on
Wednesday to Hilary Watson, the **ex-wife** of former PGA Tour player Denis Watson ...
www.cnnsi.com/golf/news/1999/08/31/golf_notes/ - 11k - Cached - Similar pages

# Exercise 1

Which would be the best choice of search (subject guide or keyword) for the following tasks

1. Find Arsenal fan sites

2. Find a report about the 1973 FA Cup Final between Leeds and Sunderland

3. Find a list of health sector employment agencies

4. Did Elvis Presley meet Richard Nixon

**\* see next page for answers**

# Answers 1

1. The best place to find Arsenal fan sites would be in the "Arsenal Football Club" section of a subject guide.

2. The best place to find a report on the 1973 Cup Final would be a keyword search engine. This is likely to be a single web page rather than an entire site, and it is doubtful that a subject guide like Yahoo would have a section dedicated to the 1973 Cup Final.

3. A subject guide would be the best place to look for health sector employment agencies. Simply go to the Health section of a site like Yahoo and look for the "employment" subcategory.

4 Whilst subject guides may have general sections on Elvis and Nixon it is doubtful that they would have a list of sites specifically about a meeting. I would therefore suggest a keyword search for any page that mentions Elvis, Nixon and "meeting".

# Using the right syntax

Many frustrated researchers complain that search engine suggest too many irrelevant websites.

A Google search for John Major yields 3,360,000 "hits" (i.e. pages in its index).
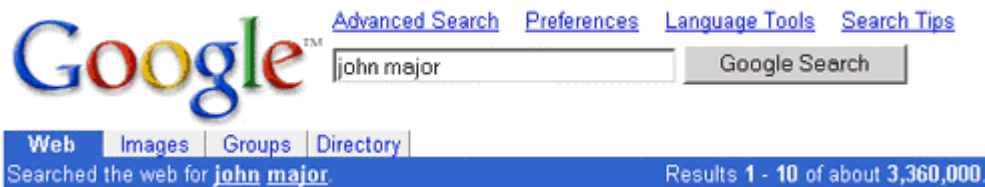


Are there really over three million pages about John Major?

A closer look at the results reveals that Google has also found pages about Major John Wesley Powell and John Skilton's Major League baseball links.

We shouldn't be too hard on Google because it has actually done exactly what we asked it to do. We entered the words John and Major into the search box, and Google found us pages with John and Major on them. We need to specify that the word John has to be directly before the word Major.

The correct syntax for expressing this is to put the phrase in quotation marks and search for "John Major". Google find 56,400 pages with the exact phrase we were looking for.

**Phrasing**

Sometimes you have to be very careful when wording a search.

A search on the phrase "John Major's resignation speech" is too restrictive and only yields a couple of a couple of pages.

By doing a "combined search" for any page that contains the phrases "John Major" and "resignation speech", we can be a little more flexible in our approach. This search yields 125 pages including the official John Major website.

125 pages is still too many to trawl through and we might find that many of them have irrelevant pages about Margaret Thatcher's resignation speech.

Search engines handle combined searches in different ways, but the best, all-round syntax for this search would be to precede each included element of the search with a plus sign and each excluded element with a minus sign.

+"John Major" +"resignation speech" –"Margaret Thatcher"

yields fifty good pages for us to look at – all of which have the phrases "John Major" and "resignation speech", but none of which mention Margaret Thatcher.

As we can see, search engines can be rather fussy. They lack the intuitive understanding of human beings and need things spelt out to them. Most search engines have good "help" sections, though, and you can quickly check up on their syntax rules.

# Exercise 2

Which would be the best keyword search to find Margaret Thatcher's middle name?

+ "Margaret Thatcher" + "middle name"

+" Margaret Thatcher" + "full name"

or

+" Margaret Thatcher" + biography

**\* see next page for answer**

# Answers 2

"Middle name" is not a particularly common phrase and a search on Google will only yield two pages (neither of which have Margaret Thatcher's middle name").

"Full name" is a commonly used term, but we have not ensured that the phrase directly refers to Margaret Thatcher on the page.

"Biography" is the best choice. If we can find a decent biography of Margaret Thatcher we can be sure that it will mention her middle name.

# Refining your search

Sometimes the best choice of search engine and perfect syntax will yield too many or too few sites. Sometimes the results will be from the wrong kind of source or will place emphasis on irrelevant details. If our first search doesn't work we probably need to refine it by adding more criteria or by using "advanced search" tricks.

**Likely Keywords**

As we can see the key to successful searching is to choose the right words. The most important thing is to think about which words will be on the kind of page you are looking for.

A search for "Oasis" will give you sites about the pop group, the drink, the shop and an oasis in the desert. If you wanted pages about the pop group Oasis you will need to add keywords.

What do keywords Oasis themselves think you will use?

Website producers often give great thought about the keywords users might type into a search engine when trying to find their sites. They are then careful to incorporate these words into the Webpage's HTML code. The keywords used on the official Oasis site are:

**Oasis, Liam, Noel, Gallagher, Gem, Andy Bell, Alan White, Bonehead, Guigsy, Ignition, Official, Standing on the Shoulder of Giants, Be Here Now, Morning Glory, Definitely Maybe, Discography, Competition, Tour, Videos, Photos**

It is a good idea to have as much information at your disposal BEFORE you conduct an Internet search. If you have a newspaper article, highlight potential keywords. If someone has asked you to look for something, ask for as much detail as they can provide. The more keywords you can use, the more options you'll have for searching.

# Exercise 3

Which of the following searches would we most likely to get you general, high quality resources dedicated to the pop group Oasis

Oasis + "Pop Group"

Oasis + Music

Oasis + Wonderwall

Oasis + Gallagher

**\* see next page for the answer**

# Answers 3

Adding "pop group" would not yield high quality results as the phrase "pop group" is not likely to appear on pages dedicated to a loud alternative rock band. A search for "pop group" is more likely to bring up wishy-washy pages about rock music in general.

Adding "music" would produce unpredictable results. You might get good quality pages that contain the word music, but you would also get a lot of spurious results.

Adding "Wonderwall" might produce pages that are too focussed on one aspect of the band Oasis. You may get too many pages with lyrics to the song rather than background and information.

Adding "Gallagher" is probably the best technique. It will rule out all spurious sites about soft drinks and shops and will return results that are not too focussed on one particular aspect of the band.

# Geography, Politics and Language

The language we choose often determines the type of site that the search engine will suggest to us. On a basic level this could mean use of regional spelling, local slang or specialist words. A search for "Tony Blair" and "Labor" party would get American sites about British politics

Language also has a political ramifications. If we are after Northern Ireland Republican groups it might be better for us to search for the phrase "North of Ireland" rather than "Northern Ireland".

Many general news sites talk about the Middle Eastern group Hamas, however if we are after sites that offer in depth profiling we may want to use their full name "The Islamic Resistance Movement". If we added the word "terrorism" into the search we might get sites that are more sympathetic to Israel, whereas the term "freedom fighters" might steer the results to Palestinian sources.

**Advanced Search Techniques**

Many keyword search engine have an "advanced search" option. This will enable you to specify where keywords should appear, allowing for greater control over the results.

In addition to general searches you can also specify that certain words appear in:

a) the title of a web page.

Most Web designers give their pages titles that reflect the main subject. Therefore if you search for the word biography and specify that "Margaret Thatcher" should be in the title, you can be assured that the sites suggested will be about her.

b) the web address

This is most useful for focussing your search on a country or origin, a specific type organisation or a company. A search for Yasser Arafat from Israeli sources could be controlled by insisting that ".il" (the country code for Israel) appears in the web address. A search UK Government pages about genetically modified food" could be a lot less painful if we insist that the phrase "genetically modified" appears in the title and "gov.uk" appears in the web address.

# Choose your source carefully

## Pitfalls

The Web provides an easy, open publishing route for everyone, from massive corporate entities to previously disenfranchised minority groups, everyone can have a voice on the Web. It is also a wonderfully democratic medium where "Bob's Mickey Mouse Fan Site" can sit happily above the official Disney site in any search results.

However, there is a downside of this. Being such an open, easy, unrestricted medium the web is open to abuse and there are quite a few pitfalls awaiting naïve researchers. Anyone conducting serious information research should look out for the following:

- biased sites

- spoof websites

- amateur websites

- deliberately malicious sites

- out of date sites

## Evaluating the Source

Amateur websites are often built on "web space" provided by their Internet service provider. Most ISPs will allow their members to build websites on their servers. You can spot these sites by looking for the name of an ISP in the web address. Also look out for "free web space" sites like Geocities. These offer web space to anyone who wants it and fund themselves through copious advertising banners.

You can often assess the bias of a website by looking at its "Links" section. Motivated sites nearly always offer links to other sites that they feel further their cause and it is rare to find a political website that offers links to their opposition.

Spoof websites can be difficult to detect if you are in a hurry. Do not take any site at face value. The website [www.whitehouse.org](www.whitehouse.org) looks official enough but it is in fact a spoof. Check the ownership of the site and try to corroborate any suspect information with reliable sources.

Malicious sites will often hide behind an authentic look, but provide little or no contact information. Owners of libellous websites will publish anonymously – often taking advantage of free web space providers as they seek to evade the legal consequences of their content. If you have doubts about a site, why not do a background check by searching for any mentions of it on other websites.

You can check the date of a page by typing:

**javascript:alert(document.lastModified)**

into the address bar.

This will prompt a box displaying the date the page was last saved on its server. Be careful though – this change might have been a minor alteration. Also, be aware that some pages are generated automatically when requested (although the box will display todays date, this is not a reliable indicator of the content's age).

# Checking ownership

http://www.easywhois.com

Reliable websites should always provide good contact information. This can often be found on the home page or in a section marked "contact us" or "about us".

If an organisation has registered its own .com, .net or .org domain name, we can check contact details by searching domain name registration database. A convenient site for running this search is www.easywhois.com

```
Domain Name: BBC.COM

Administrative Contact, Technical Contact, Billing Contact:
    Butterworth, Brandon  (BB231)  brandon@RD.BBC.CO.UK
    British Broadcasting Corporation
    Kingswood Warren
    Tadworth
    Surrey
    KT206NP
    UK
    +44 1737 839592 ext. 374 (FAX) +44 1737 839516

Record last updated on 26-Jan-2001.
Record expires on 16-Jul-2010.
Record created on 15-Jul-1989.
Database last updated on 7-Apr-2002 20:43:00 EDT.

Domain servers in listed order:

NSO.THNY.BBC.CO.UK          38.160.150.20
NSO.THDO.BBC.CO.UK          212.58.224.20
```

Want to look up another domain? Just enter it below.

```
look up a domain
    Domain name: [                    ]
```

An Easywhois lookup on the BBC.com domain name

# Always double check

Information gleaned from a website is as reliable as information taken from a piece of paper found in the road. You can feel a lot more confident if you can trace the authors of a web page, but even the most reliable sites can make mistakes.

To ensure the integrity of your research always double check by performing new searches for any information found. Try to corroborate information by seeing if it is repeated on reliable websites.

Look for information from responsible news sources or credited academics. If you can't corroborate the information, try to trace the owners of the site – speak to them and satisfy yourself that they are reliable.

# Useful Internet courses

We have updated our portfolio of internet research courses.  While we continue to offer a short course to get complete beginners up to speed - **Introduction to the Net** - we now also offer three specialist courses for those using the net for research purposes.  These are:

- **Investigative Journalism on the Net** - an advanced research course that is suitable for both researchers and journalists
- **Undercover Journalism on the Net** - an exciting new course for conducting anonymous research on the net
- **Statistical Journalism** - a course in manipulating and sorting figures to reveal trends and find stories

For  full details on all our New Media courses visit our website
www.bbctraining.co.uk
email training @bbc.co.uk or call us on


# + 44 (0)870 122 0216