

Search techniques for the Internet

Boolean logic for non-logicians

There is an awful lot of material on the Internet. Look at the results these searches threw up in April 2002:

<u>Search for...</u>	<u>Engine</u>	<u>Results</u>
Whale.....	Google	1,260,000
Astronomy.....	AltaVista.....	969,983
Terrorism	Lycos.....	2,206,622
Romania	alltheweb.....	2,833,547

Clearly, we are not going to look through two million pages, so searches need to be narrowed. This is done according to Boolean logic. It sounds daunting, but the principles are simple, and once you have understood them you need remember only a handful of symbols of the +, - and " " type.

Boolean logic as applied to an Internet search depends on three concepts:

- OR** If you write **astronomy OR astrology**, you will be shown sites which include either astronomy or astrology (or both). This is the search that will produce the largest number of results.
- AND** If you write **astronomy AND astrology**, the results will show all the sites which include both of those words. (You might, for example, be researching an essay on the early history of astronomy.)
- NOT** If you write **astronomy NOT astrology**, the results will show all the sites which include astronomy except those that also include astrology. (You may wish to exclude sites which are not scientific.)

The search engines have, unfortunately, managed to complicate the issue by expressing these concepts in significantly different ways. However, try to focus on the fact that at the end of the day it all boils down to these three queries.

Popular Search Engines at a Glance

www.google.com	www.AltaVista.com
www.lycos.com	www.AskJeeves.com
www.alltheweb.com	www.yahoo.com

How does this work in practice?

In practice, you will need to look at the helpfile of the search engine you are using to see exactly how it expresses the concepts described above. As the search engine systems differ, most people tend to stick to one or two engines that they get to know well. (But see *Choosing a Search Engine*.) As an example, let's look at *Google*, one of the most popular search engines.

Imagine you are looking for information on how whales breathe.



Stage 1

You go to your normal search engine and type in **whale**. You get 1,260,000 results, the first page of which - whale songs, whale watching, even "the infamous exploding whale", clearly show that your search is too broad.

Stage 2

You decide to narrow your search to sites that include both **whale** and **breathe**. You therefore type in **+whale +breathe**.

This is how *Google* expresses the AND concept above. The search is significantly narrowed and

there are some useful pages within the first ten results. If you want a children's site, this search is adequate. For a more academic search go to...

Stage 3

You decide that you should include the more academic term "respiration" as an alternative to breathe. You therefore want to search for **breathe OR respiration** as well as **whale**. To do this, you need to put the alternatives in parentheses:

+whale +(breathe OR respiration)

As in mathematical calculations, the items in parentheses are dealt with separately, as a unit. So we are searching for documents that have both **whale** and either **breathe** or **respiration**.

Stage 4

We can do the same thing to the word **whale**, adding **marine mammals**. However, we want the engine to search for marine mammals as a unit rather than as two separate words, so we enclose in inverted commas:

+(whale OR "marine mammals") +(breathe OR respiration)

Stage 5

For the purpose of this assignment, our interest in whales is not to save the whale, so we want to exclude all sites with an environmentalist slant:

+(whale OR "marine mammals") +(breathe OR respiration) -green

Now go to *Google* and type in this query and compare the results with the initial query in stage 1. This will show you how a refined search helps you to locate the exact information that you are looking for.

In a nutshell

- +cat +dog** (with no spaces between + and the item it refers to) searches for pages with both items.
- +cat -dog** searches for pages with cat but not dog. (Note that in *Altavista* you do not need a space before the -, but in *Google*, you do.)
- cat OR dog** searches for pages with either or both. You need to know whether **or** should be in capitals: *Google* requires **OR**; **or** is fine for *Lycos*.
- In **"persian cat"** the quotation marks tell the engine to search for the words as a unit. It therefore won't find **cat** or **tabby cat**.
- Similarly, brackets tell the engine to view the query as a unit: **+(cat OR feline) +dog**
- On most search engines, it doesn't matter whether you write in capitals or not, so **Bush** = **bush** = **buSh**. But *Altavista* is case sensitive and will therefore discriminate between the plant and the president.
- Some search engines allow you to do "stem searches". So if you can't remember the exact spelling of the person you are looking for, you can type **Modigl*** and the results will include **Modigliani**. (But not all engines offer this: *Google* doesn't.)
- Be careful of the NOT search. A search for **+cats -dogs** would exclude a site which stated, *This site deals only with cats. For dogs, click here...*
- Most search engines have their own advanced search pages which help you formulate your searches.
- All reputable search engines have a help page which explains their search systems.

What are you searching for?

Search engines allow you to specify whether you are searching for text, audio, website titles or URLs, results in a particular file format (eg .xls) or on a particular subject (eg news). You can generally make these choices either on the front page or in the advanced help.

© Clarity Language Consultants Ltd 2002. May be reproduced only by institutions licensed to use Study Skills Success. See www.ieltspractice.com.

