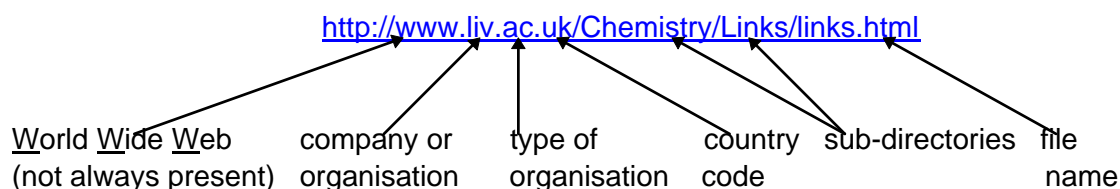


APPENDIX 2: STRUCTURE OF URLS

Addresses on the Internet are called URLs - Uniform Resource Locators. Web addresses start with <http://>, the [http](http://) standing for hypertext transfer protocol. The rest of the address gives the location of the page on the Internet. For example <http://www.liv.ac.uk/> is the Home Page or front door to Liverpool University's web site. You may be given a more complex URL, for example Liverpool Universities Links for Chemists page at <http://www.liv.ac.uk/Chemistry/Links/links.html>

Some search tools have options that allow you to search for terms within a URL, to search for links to a specific URL, or to limit your search to a type of organisation - for example academic or commercial - by using part of the domain name. Most of the options for searching by country or continent work by analysing the country codes at the end of the domain name.

All URLs have a standard format, for example:



Directories and file names are usually case sensitive. If you have a URL that is a mixture of upper and lower case letters it must be typed in exactly as given.

The part of the URL representing the company or organisation is not always exactly the same as the organisation's official name. There will inevitably be several, completely separate organisations with similar names, for example RBA Information Services, RBA Design, RBA Software. Only one of these can have [rba.co.uk](http://www.rba.co.uk) for their *domain name* and, with a few exceptions, it is "First come, first served". If one of your strategies for locating company web sites is to guess the URL by adding [.com](http://www.com) or [.co.uk](http://www.co.uk) to the company name, always double check when you get to the web page that you are looking at the right company.

After the company or organisation "name" is the type of organisation:

- org** - trade, industry or research association; charity; not for profit, but can be registered by anyone
- co** or **com** - commercial, private company, individual
- ac** or **edu** - academic
- net** - Internet service or network provider
- gov** - government
- mil** - military

These are referred to as TLDs - top-level domains. New TLDs that are now coming into use include **.name** for personal sites, **.info** and **.biz**.

Some countries do not include the organisation type in the URL so it is not always present. UK companies have the additional options of **.plc.uk** or **.ltd.uk** depending on their legal status. Very few companies have registered these domains because of the restrictions placed on their use, for example the name has to match exactly the name that appears on the company registration documents.

The country code is the standard ISO two letter code for example **.de** for Germany, **.jp** for Japan. Note, though, that **uk** is used rather than **gb**. If there is no country code, for example <http://www.shell.com/> it can be based anywhere. Anyone, and I do mean **anyone**, can register a **.com** address provided that it is not already taken and they pay the registration fee. The registrant could be an international organisation, a large corporation or a private individual, and they can be based anywhere in the world. A **.com** URL does not by itself confer authority on a web site.

The country code of a web site does not tell you where the organisation is physically located. A company may decide to register its domain name in several countries. Neither does it tell you where the web server is located.

Each country has its own domain name registry that is open to public view and most can be searched via the web. Allwhois (<http://www.allwhois.com/>) is a good starting point for finding out if a domain name is already taken and who owns it.