Search Skills on the World Wide Web

What's out there?

The Internet gives you access to many types of information. This information can be found various forms (e.g.):

- Mostly text-based resources
- Numeric data
- Interactive/interrogative resources
- Images
- Audio and Video
- Sometimes these can overlap: *e.g. a text page containing numeric data, images and embedded videos.*

Text-based resources include:

Traditional information in digital format – e.g.

- Online books
- Online reference sources
- Online journals

There are other text-based sources of information on the web, *for example*:

- Simple web pages and **downloadable text based documents**, e.g. Word, Acrobat, PowerPoint, etc.
- **Forums and other social media** which are used for discussing academic topics.

Numeric/data resources include:

- **Tables** set out in web pages
- **Downloadable data** in the form of files such as Access, Excel, SPSS, Minitab, etc
- Web pages which incorporate **scripts which calculate** (e.g. numeric/currency conversion)

Other types of web resource include

- **sound**, e.g. audio files which can be downloaded or 'live feeds'
- **images**, e.g. simple images, animated images or interactive images (e.g. image maps)
- **Video**, e.g. video files which can be downloaded or 'live feeds' including webcam
- **interactive resources** (e.g. online tutorials, forums, chat rooms and webcam).
- Web pages which are the 'front end' of **databases**

These latter, database-organised resources, are possibly the most important of all as they are behind many other key resources,* (e.g. to search for items to buy, words in an online dictionary, thesaurus or items in an *Online Bibliography).

Now we've completed our review of what is out there, we'll look at how we find the material we want on the topics we are interested in.

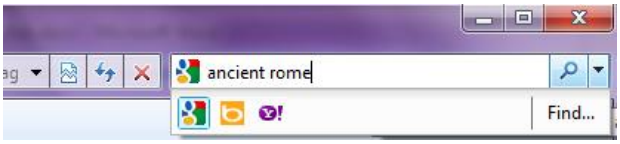
Finding Internet Resources

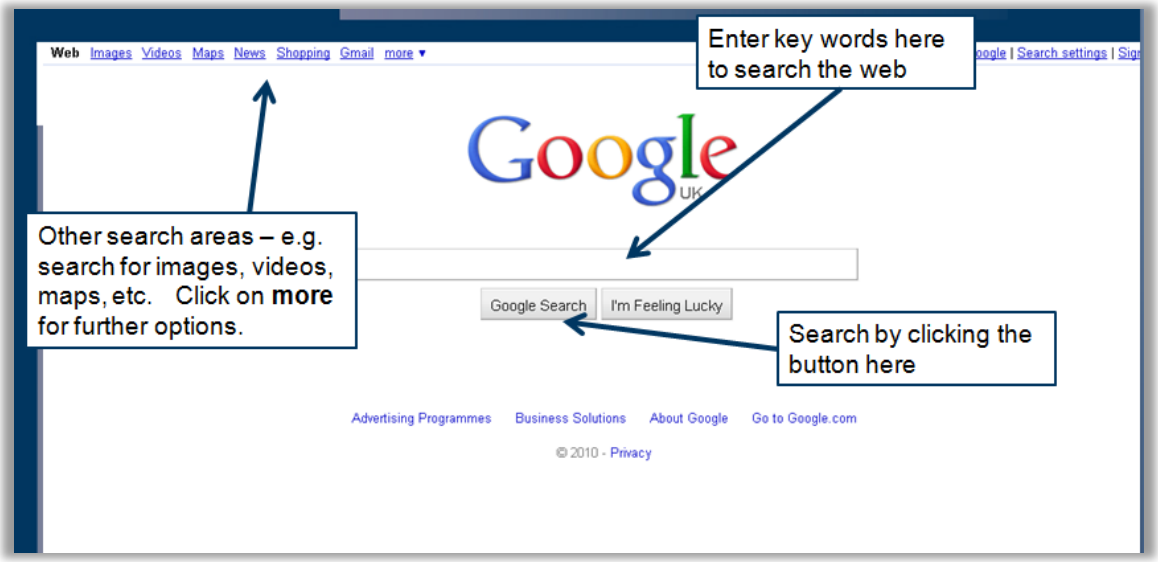
There are a variety of online tools you can use to find information on the web.

- You may already be familiar with **Search Engines**. These are systems which search and create databases of a wide range of accessible resources on the web.
- **Subject Gateways/Directories** are a type of classified directory. They are usually held on databases which can be browsed or searched.
- **Bibliographic Databases** list articles in journals, books, theses and research papers.
- But you must know how **to use the above resources properly**.

Search Engines

- **Search Engines** regularly scour the web for new material and add references to web sites it finds to a large, **searchable** database.
- These web sites can contain any kind of material and all resources are available, apart from a few 'hidden' resources.
- Most browsers have "search boxes" linked to popular search engines like Google:





At the bottom of a **Google** search page you will find suggested **related** searches.

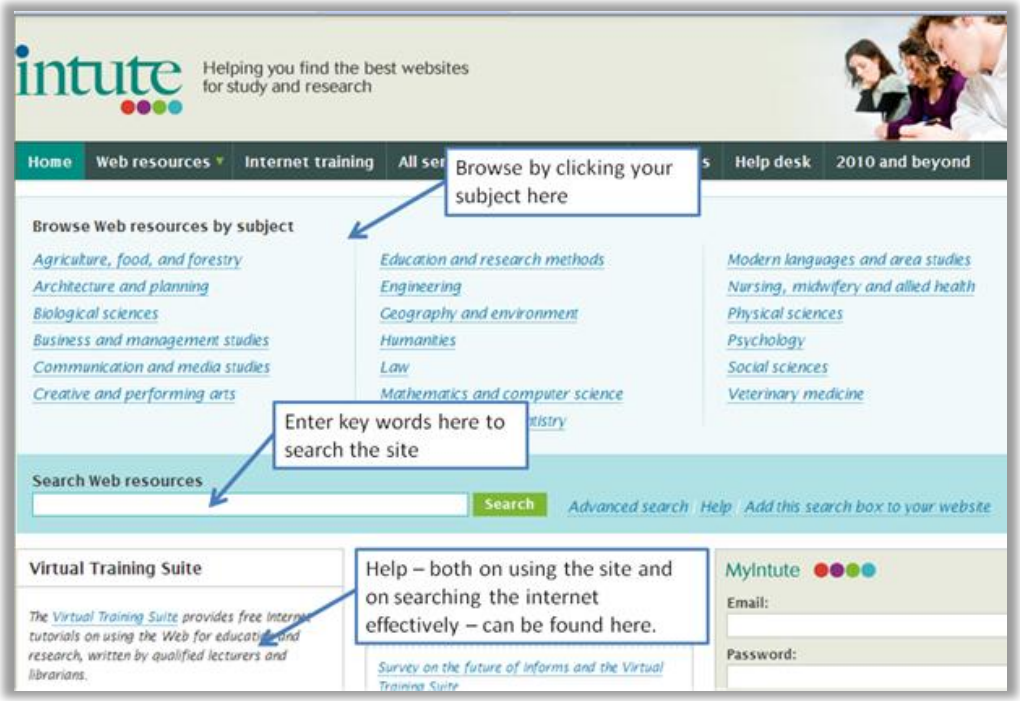
Below this are links (1-10) to more "hits".

Finally, below this is the link to the **Advanced search** page.

- Search engines can very quickly find you useful and often very up-to-date information.
- The downside is that much *irrelevant information* may be included in the search engine databases.
- This means you must use **effective search techniques** to ensure you get relevant results when you perform a search. You can't just Google it!
- You must also understand how to **evaluate** the worth of web sites to your studies.

Subject Gateways/Directories

- Offer 'value-added' approach to information seeking as subject specialists are involved in their compilation and design.
- They concentrate on organising high quality resources, often on topics of interest to users with specialist interests. An example of this is **Geo-Guide** - http://www.geo-guide.de/
- They often combine resources covering a number of different subjects.
- An example of this is **Intute** - www.intute.ac.uk



- These can be searchable within individual subject areas or across several subjects at once.
- They are usually searchable by browsing topics or by entering key words in a search box.

Online Bibliographic Databases

- These are usually subject based though they often cover a wide range of related subjects.
- They will cover traditional, printed material as well as online material (e.g. online journals). *Often they will give links to online versions of sources referenced*.
- The databases will be searchable using subject terms or by author and/or title.
- *Citations found may also include brief abstracts of the item's contents*.

To recap:

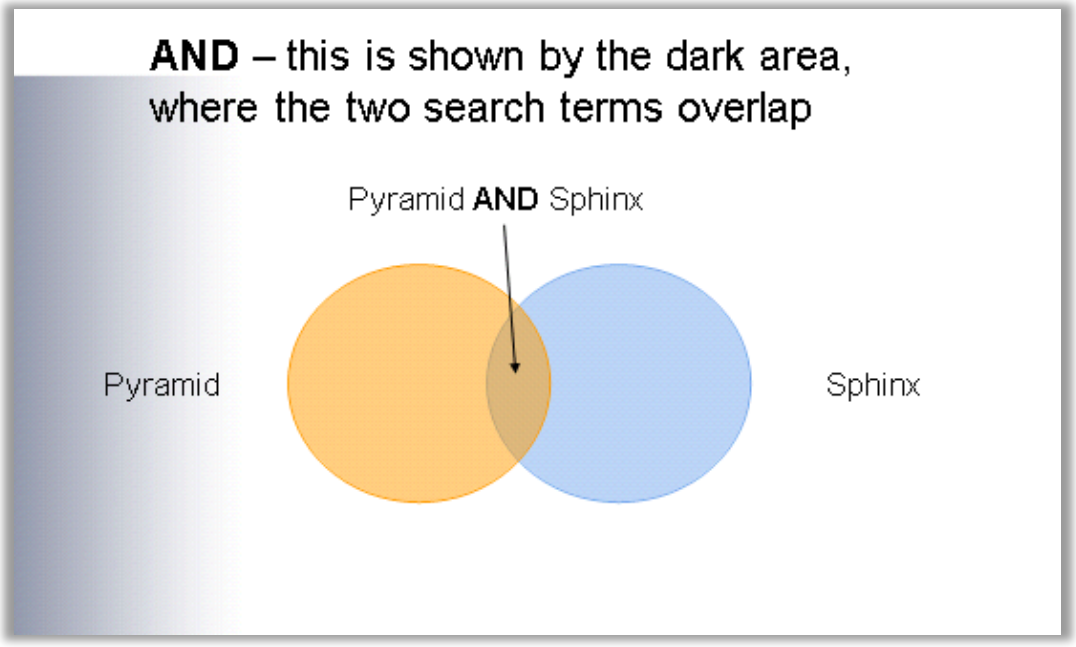
- ✓ *Many internet resources can be found through online databases.*
- ✓ *They very often have searchable databases of the information they contain.*
- ✓ *So knowing proper database search techniques is an essential skill if you want to access these resources.*
- ✓ *We'll now look at basic database search techniques.*

Boolean Operators

- Boolean operators were developed as a mathematical and logical tool in the 19th century (by George Boole).
- One important application of it is in computer operations and, in particular, **databases**.
- You don't need to understand all about Boolean logic to use them.
- The basic operations commonly used are: **AND**, **OR**, **NOT** (or **AND NOT**)
- These few basic operations are all that are required to be able to search a database efficiently.

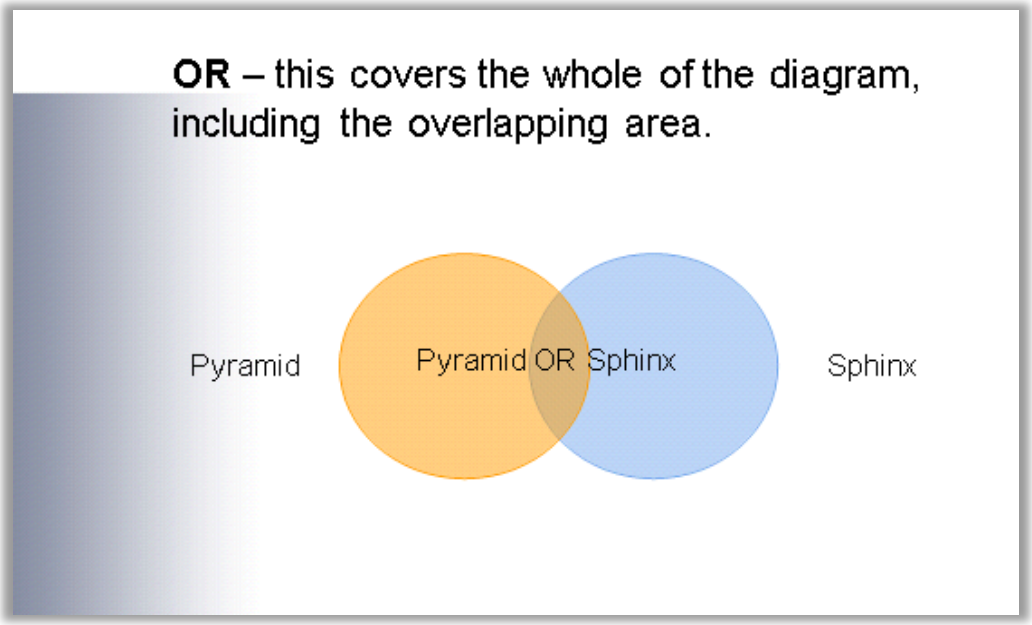
Now a brief explanation of the concepts behind them.

- To help explain the concepts described, examine the diagrams below.
- To follow this, note that:
  The overlapping circles form a Venn diagram.
  Each circle represents a single search term
  The area where they overlap is where both terms are present
  Outside the overlapping area only one term is present



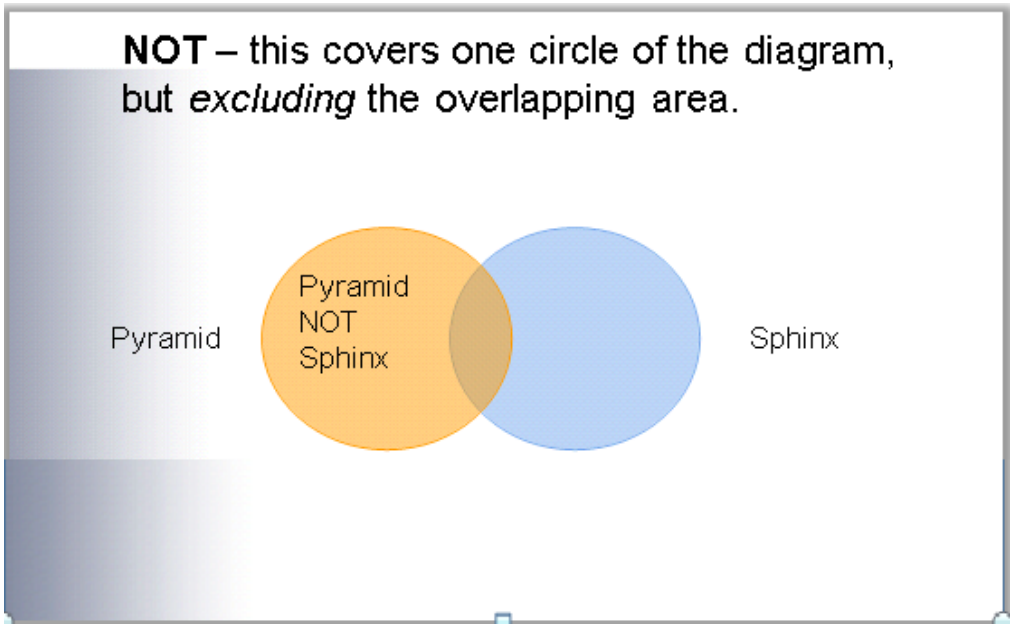
- **The Boolean Operator - AND**
  - Used between two terms or phrases, this will force search results which contain both terms.
  - You can combine several terms or phrase, not just two. Just keep adding 'ANDs'.
  - **For instance**, *'English* AND *Language* AND *Chaucer* AND *Style'*
- **NEAR**
  - A variation on AND which is occasionally allowed is **NEAR**
  - **NEAR** is not found in Boolean logic but was created for searching databases (although not all databases make use of it).
  - If you searched using the terms, "*Chaucer* **AND** *Style"*, you would be hoping to capture variations like

    "*Chaucer's style", "the style of Chaucer", "Chaucer developed an original style", etc.*
  - Many databases will search for the two terms throughout the whole text of a record.
  - So you can get results where the two terms occur, in different sentences or even different paragraphs. This may not be useful to you.
  - If the option **NEAR** is available you can choose to search to find the two terms only where they are separated by no more than 5, 10, 20, etc. words.



- **The Boolean Operator - OR**
  - This operator is used to specify that either one term or another is present in your retrieved results:
  - '*English* **OR** *Scots'* means that either one term or the other – or both – is present.
  - **Both** could be present but you will get results that contain just one of the terms as well

- This is also used for synonyms or alternative spellings, for instance:
  - '*Computer* **OR** *PC* **OR** *Workstation*'
  - '*Color Coding* **OR** *Colour Coding'.*

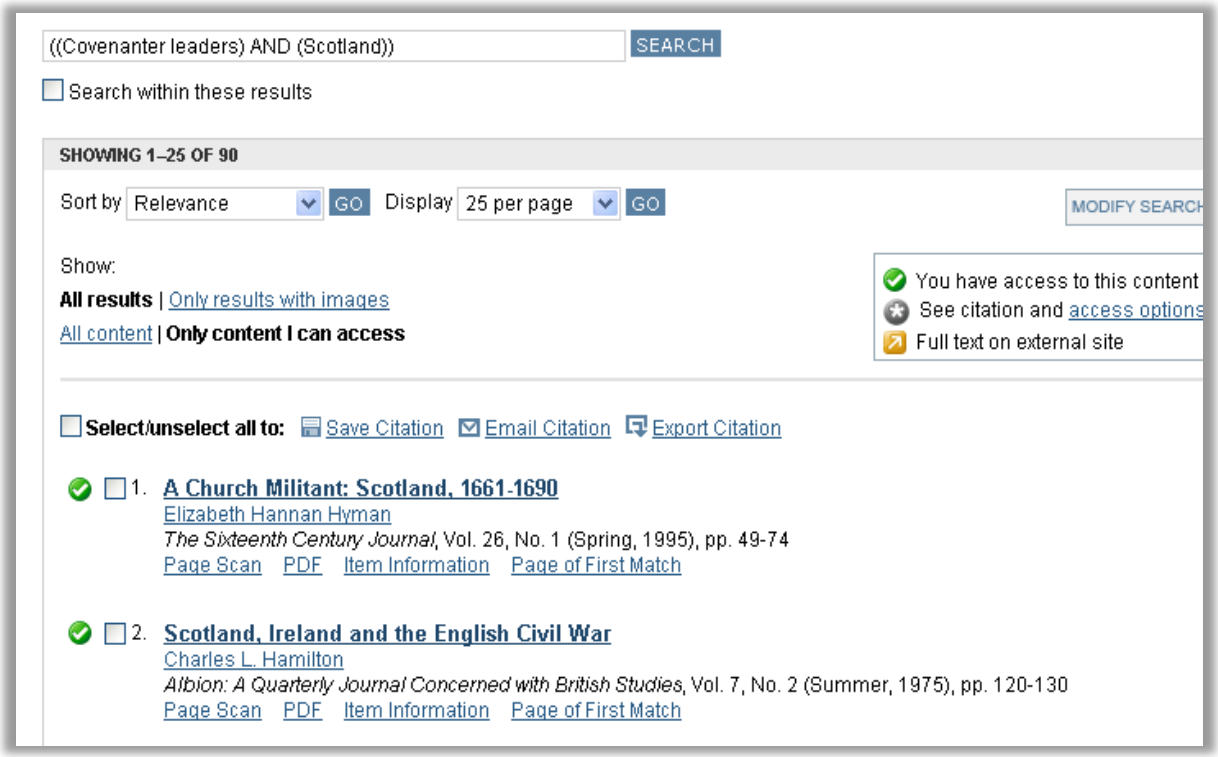


- **The Boolean Operator – NOT (**sometimes **AND NOT)**
  - This operator is used when you want to specify that a term is ***not*** present in your retrieved results.
  - Suppose you wanted to know about the Egyptian sphinx and used this search:
    '*Egypt'* **AND** *'Sphinx'* **NOT** *'pyramid'*
  - This would give only references containing the keyword '*Egypt'* and references containing the word '*pyramid'* will ***not*** be retrieved.
- ***However, this should be used with caution.***
  - The previous example would get you many 'hits', but it might also lose you valuable results because *pyramids* and *sphinxes* are sometimes discussed together.
  - You might find it better not to exclude *pyramids* and consider other ways of reducing the number of hits you get.
  - Perhaps '**NOT** *tourism*' to exclude some non-academic information.
  - Or perhaps by adding extra search terms, e.g. '**AND** *Old Kingdom*'.
  - Now we will look at using Boolean terms in practice.

Searching Within Databases

- Most databases tend to have the same structures and can be browsed and searched in similar ways.
- So we'll look at one database which will be a useful guide to using many other databases. This is JSTOR. Access it from here: http://www.jstor.org/

- Since the initial search option is very simple, we'll look at the **Advanced Search**.
- Note that you have plenty of space for adding more than one term in ether search box. You can also add more search boxes by clicking the **Add a Field** button.
- Search terms (keywords) are connected using Boolean operators: AND, OR, etc. (these are discussed later.
- Here are the search results (or part anyway):



- Note the search box at the top of the search results page - the search options you choose have been combined with the keywords you entered to create a **Boolean** phrase.
- Compare this with your original search options: it should help you understand how this works.
- You could have entered a Boolean phrase like this into the simple search box. However under the **Advanced Search box** the **Narrow by** option allows you to focus on particular areas – type of source, subject and journals.

- Every bibliographic database works along these lines. There will be:
  - *a basic search option*
  - *an advanced search option which allows search using Boolean operators*
- Some may have no more than this, others may like JSTOR allow other ways of refining your searches – e.g. by topic, format, date or language.

Finding out more …

- It is essential you check out the **Help** on any database you use so you understand how to make best use of it.
- There is also a very good guide to finding electronic (digital, web) resources on the library web pages
- There is a topic-by-topic guide on web-based research at: http://www.vtstutorials.co.uk/
- And the more exciting step-by-step guide to the web: The Internet Detective.

Evaluating Web Resources

- Web resources found via Bibliographic databases are usually validated by subject experts.
- However there are many other resources on the web which don't get this validation – e.g. web sites you find via a search engine.
- In these cases you have to be able to do your own validation.

How good is it then?

- How do we decide whether particular resources and information found on the web are worthwhile?
- There are several aspects to this question:
  - *How easy it is to access the site*
  - *How easy it is to find your way around the site*
  - *How regularly the site is updated*
  - *How reliable the information content is.*

- The final point is the most vital, although all of the points are important.
- A good collection of guides is here: http://bcdwebsearch.wikispaces.com/Web+Site+Evaluation
- This can be judged by the following criteria …

Five Basic Evaluation Criteria

- **Accuracy**: Web sites don't always have editors and fact checkers.
- **Authority**: Check the internet source - is it an academic institution; a government web site; a professional association?
- **Objectivity**: Academic web sites are likely to be more objective than the web site, say, of a political party.
- **Coverage**: The range and depth of the discussion of a topic is important.
- **Currency**: Date of update is always given on a well organised site.

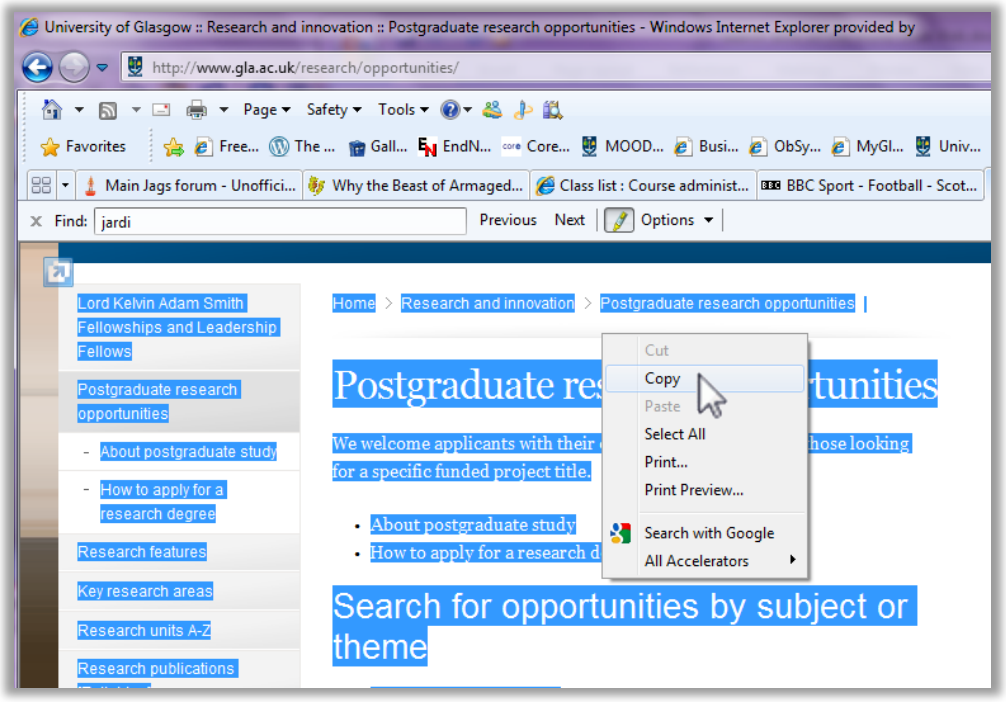
On the Internet, nobody knows you're a dog.[1]

Some more guidance can be found at:
http://www.lesley.edu/library/guides/research/evaluating_web.html
http://21centuryedtech.wikispaces.com/Web+Page+Evaluation

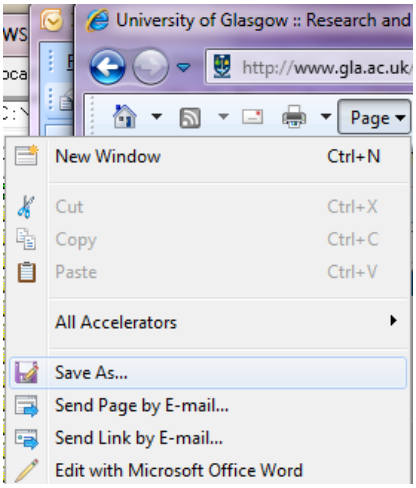
[1] New Yorker cartoon 1993.

Copying Information from the Web

- **Copying Text:** select it then copy and paste into (e.g.) a *Word* document. If the page is a long one, save it as a text file and insert it into your *Word* document.



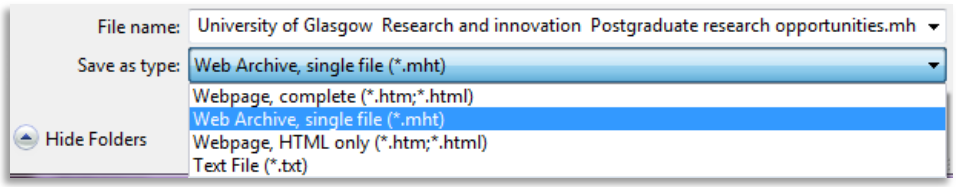
- **Saving a Web Page as a File:** For single pages, this is straightforward. Choose **Save As...** from the **File** menu (or from the **Page** menu) in the browser. You are offered a choice of file types:



a) A **text file,** with a.**txt** extension: the file will lose graphics and formatting, retaining only the text, but it can be easily inserted into most applications.

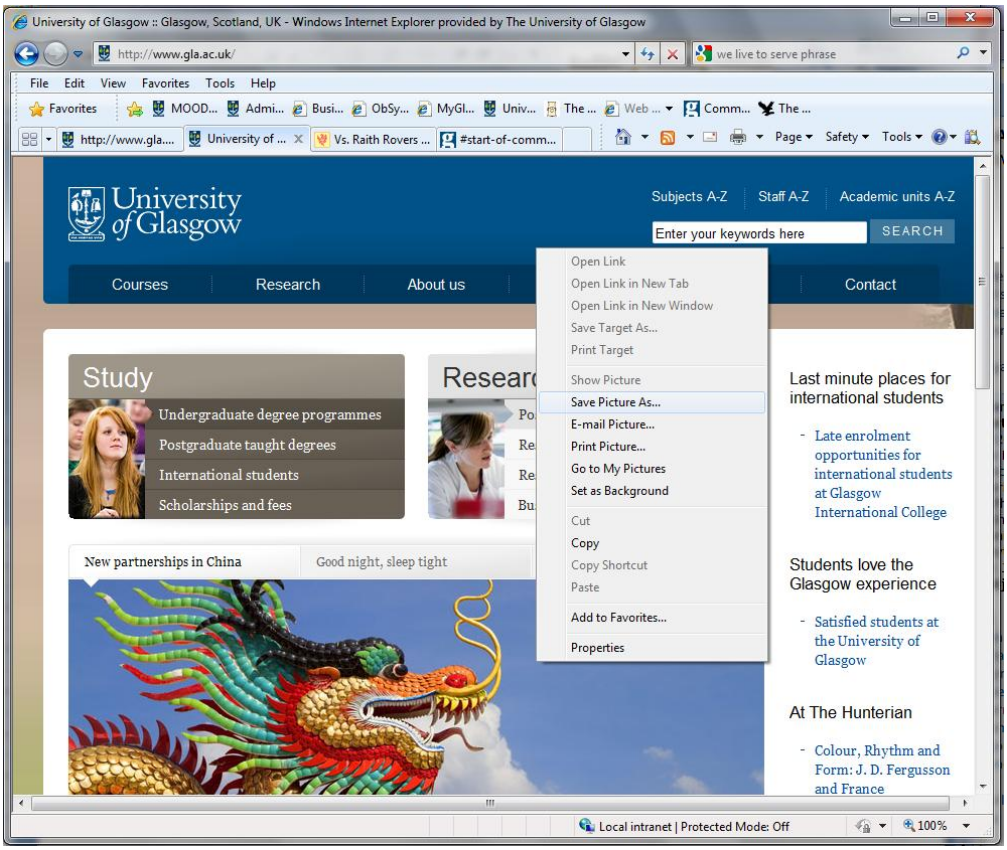
b) An **HTML file,** with a **.html** extension. In this case the HTML code is retained in the text, preserving all the formatting, links to graphics, but not the graphics themselves, and hypertext links to other web sites.

c) A **web archive, complete,** with a.**html** extension (in *Internet Explorer* this option is simply named **Web Archive**). It saves the page you are on, including any images and links on it, so is displayed just as it appeared originally.



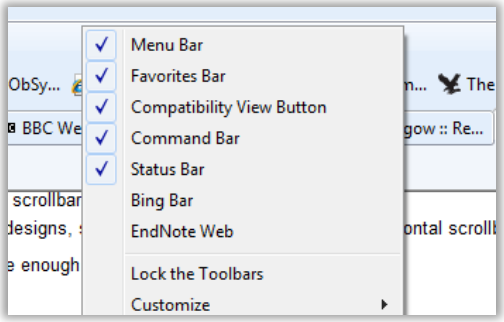
Downloading files:

- Web pages often offer the option of downloading (via a link) text or other files in a variety of formats: e.g. as a *Word* or *Excel* document or a **PDF** file which can be viewed using a program called *Adobe Acrobat.*
- When this option is available you can have more control of where the file is downloaded to by right-clicking on the link and choosing (from the menu that appears) **Save Target As** (in *IE)* or **Save Link As** (*Firefox).*
- **Copying a Picture:** Put the mouse pointer on the image and then press the **right** mouse button. In the menu that appears you will see an option to **Copy** the picture. Select this then move to the application you wish to copy to and **Paste** the picture into the document.
- You could instead choose **Save Picture as ...** (if using *Internet Explorer*) or **Save Image as ...** (if using *Firefox).* The usual **Save** dialog box will appear. Note that images on the web are usually **.gif** or **.jpg** files. Then it can be inserted into a *Word* document using the **Picture** button on the **Insert** tab.



- Most browsers now allow you to **Copy** the image after which you can **Paste** it into another application. *Word* allows you to **Insert** a wide range of graphic file types into a document but not all types. For some you will need to use a graphics program, which will read lots of file formats and do conversions, such as *Paint* or *IrfanView.*

- **Printing out a Page**
- Most browsers now allow you to print out the current page. Use **Print** from the **File** menu.
- When printing from a web page in most browser types, it is advisable to use **Print Preview** option from the **File** menu.
- If you don't see the **File** menu, right click on the toolbar: from the menu, select the **Menu Bar** option.



Some Words of Warning

Copyright:

Don't assume that anything you copy from the web becomes yours to do with as you wish. The right to copy any text whether on the web or on paper remains with the author or any person or agency to whom the author has assigned the copyright.

If you use the text purely for your own study purposes, and copy excerpts or images for quotation in essays or other works produced solely in connection with your own study, you should have little to worry about.

But, if you claim authorship of text that is not yours, or if you publish or broadcast copied text or images without permission (e.g. by incorporating it in your own web page or handing out hard copies), or if you copy material from the web and sell it, you may be doing something illegal.

Note that ignorance of the law is not regarded by the courts as a reasonable excuse.

Online Security:

Stay safe by following these guidelines:

- Treat all unsolicited email with suspicion, regardless of who the sender appears to be.
- Avoid clicking on links, or opening attachments in unsolicited email.
- Never send passwords by email (the University – or any other organisation, like your bank – will NEVER ask you to send your password by email)
- When visiting a web site that requires a password, always carefully type the URL address by hand, or better still, use a 'known-good' bookmark.
- Be very careful about any information you give on a web form: never give password or any bank or credit details.
- See also: http://www.gla.ac.uk/services/it/informationsecurity/

Viruses:

Networks enable you to get interesting and useful material from a huge range of sources. Unfortunately you can also get things you don't want too.

Junk mail and **chain letters** are not just irritating, but clog up lines and slow down the Internet; you should immediately delete and ignore them. A nastier import is a computer **virus,** which could have attached itself to a page that you access and be arriving at your computer as you read the page.

You should make sure you are protected. The only way of dealing with this problem is to run a virus checking program which is regularly updated.

All computers on university networks are protected in this way. If you have a computer at home you should make sure it is similarly protected. Even if it isn't connected to a network, you could bring to it a virus from another machine on a disc or USB pen.

Offensive material:

Another area of legal concern is the transmission of offensive material across the Internet.

It may not be an offence in some other country to put offensive material on a web page, but it could well be illegal in Scotland for you to copy that material to your computer terminal.

Whether you look at it on the screen or print it out makes little difference. "Offensive" is a definition which changes according to what is currently unacceptable but it would include hard pornographic and violently explicit material, especially including children, and material which is likely to incite racial hatred.