

INFORMATION SCHOOL BOOT CAMP

Session Four: Structures of the Internet

September 27, 2001

INTERNET BASICS HANDOUT

A. The Internet

The *Internet* is a global, decentralized network connecting millions of computers to servers in order to publish and access digital information.

Each computer has a connector, such as a *modem* or *DSL*, to send a request for data over a telephone or cable line. This request is sent to a *router*, a sort of switchboard operator, that tells the request which route to take to find the data. Most all data requests go to a *Server* where data is stored. The Server then transmits the data request back through the routers to your computer.

B. Internet Browsers

A *browser* is a software application used to locate and display Web pages. The computer labs in the iSchool have four different browsers available:

- Microsoft Internet Explorer 5.5
- Netscape 4.76
- Netscape 6.01
- Opera 5.11

Although the school provides a shortcut to Internet Explorer from the desktop, it's possible to access all of the iSchool's browsers by clicking on:

START > PROGRAMS > INTERNET BROWSERS > select a browser

There are not many differences between the four except for differing graphics and/or the presence of programming languages that aren't applied in one or the other browsers. Opera is arranged a little differently than IE or Netscape, but it often loads pages faster than the other browsers, and it portrays multiple windows on the same page. For simplicity's sake, the material presented in this information handout will reflect the utilities available through Internet Explorer. It should be noted that, by clicking on the IE shortcut from the iSchool computer desktop, the user will automatically be transported to the homepage for the Information School.

C. Browser Functionality

Web pages are arranged uniformly no matter what browser is being applied. The elements listed below explain the significant sections that appear on the monitor's screen:

1. Title Bar

At the very top of the browser is a narrow blue band with some text in it. Known as the *title bar*, the words here specify the title for the individual page being viewed.

2. Menu Bar

Below the title bar is the *menu bar*. This contains the words *File*, *Edit*, *View*, *Favorites*, *Tools*, and *Help*. Some important functions using these individual options are:

- **File > Print Preview:**

This feature is important because some pages, especially those with *frames* (i.e., Web pages that consist of separate scrollable windows on the same screen), don't always print properly. Before choosing to print something from the Internet, check the "print preview" to see how it will look on the page.

- **View > Text Size OR View > Encoding**

When working with websites hosted in other languages, different *font* (i.e., the particular typeface design and size) conventions might be used. The standard size is "medium," and the standard *encoding* (i.e., the specific coding that represents the characters) is "Western European (Windows)."

- **View > Source**

This allows the user to see the programming statements and instructions, also called the *source code*, used by the Website creator to write the page seen graphically.

- **Favorites**

This option is a tool for saving important links. It's important to be aware, however, that favorites are applicable exclusively to the particular terminal from which they're assigned. The likelihood that the same computer in the iSchool lab may be used on a regular basis is generally unrealistic, unless the student is consistent and patient.

- **Help**

For answers to questions about what is going on with the browser that's currently being employed, refer to the Help menu, notably the "Contents and Index" portion.

3. Icon Bar

Below the menu bar is what's known as the *icon bar*. Although most of these graphical image links are unlabeled, it is possible to identify their function by hovering the cursor over the icon. Because these functions vary in importance, only a few of the most significant ones will be covered:

- **Back and Forward**

The first icon on the left side of the bar, the left pointing arrow, is the back button. This takes the searcher back to the last Web page visited. Internet Explorer also has a drop box (the small triangle pointing down), from which the user can view a list of the last nine or ten pages visited. The right pointing arrow is the forward button. It is only possible to move forward if the "back" icon has been utilized, otherwise the computer doesn't know where to go.

- **Stop**

The red button with a white "X" in the center is the stop button. Oftentimes when searching the Web, certain pages, due to extensive and complicated graphics, will simply take longer to load than is desired. When this occurs, clicking on the "stop" icon will terminate the loading process and permit the user to move to another page.

- **Refresh**

This feature, the icon that appears as a white page with two circling green arrows, allows the user to reload the same Web page that is on the screen to see if it has changed since the last time it was downloaded. The “Refresh” key is useful to students as they create their personal web pages, since it lets them continuously build on and develop their content.

- **Home**

The little house is the Home icon which is essentially the default first page the viewer sees after opening the browser. The homepage varies depending upon which computer is being used, for example Suzzallo Library’s homepage will be different from that in Odegaard Undergraduate Library. In the iSchool computer labs, the home page is the iSchool page

4. Address Bar

The last line below the icon bar reflects the location for a Website’s URL, commonly known as the *address bar*. This is the space where the user inserts the Web address he or she would like to visit, and subsequently informs the searcher where he or she is on the Internet.

5. Link Bar

Across the bottom of the web page is another bar which normally appears as simply a gray space, except for the IE logo. When the cursor is placed over a hyperlink on a page, however, the URL for the potentially linked page materializes in this area.

D. What is a URL?

URL, the abbreviation for *Uniform Resource Locator*, is the global address of documents and other resources on the Web. In the same way that a library book is assigned an individual call number, every URL is unique and, within the URL, there often exists a massive amount of information. The iSchool’s URL for the student organizations directory, shown below, may be utilized as an example in examining the various parts of the URL:

<http://www.ischool.washington.edu/people/organizations.htm>

1. Schemes and Protocols

“http://”

HTTP, *Hypertext Transfer Protocol* (where *protocol* refers to the agreed-upon format for transmitting data between two devices), instructs the browser how to read the file that’s being opened. The protocol HTTPS is used for delivering secure pages, whereas the FTP protocol is applied for sending files on the Internet.

2. Server Name

a. Host: “www” (the abbreviation for World Wide Web)

This indicates that the file is hosted on the World Wide Web. Not all Internet files are available via the Web; one example of a program used in the iSchool not accessible through the Web is WebCT.

b. Domain name: “ischool.washington”

The University of Washington employs the domain name “www.washington.edu”. Domain names are usually hierarchical: the broadest part of the name is shown at the far right, with the subdomain divisions occurring before it. As such, the established convention at the university is to place schools or departments before the word “washington.” This may essentially be

recognized as the name of the site, because often they are very similar (as in the address for the business 'amazon.com' which is, in fact, <http://www.amazon.com>).

c. Top Level Domain (TLD): “.edu”

This name indicates some information about the type of site being viewed. Only a handful of these exist, and they are controlled by an independent private agency called the ICANN (the Internet Corporation for Assigned Names and Numbers). Major TLD's and their meanings are:

- **.edu: educational institution**
- **.com: commercial site**
- **.gov: government controlled site**
- **.org: organization's site (initially meant non-profit, but not as true today)**
- **.fr/.de/.uk: country designations**

To see new TLD's visit <http://www.icann.org/tlds/>

3. Path

“/people/”

As a searcher delves further into a website, he or she is, in effect, traversing down a digital path or route. A site will likely consist of multiple levels, and the terms that are part of the path take the user deeper into the file structure. If the site is well constructed, the searcher can backtrack easily by following the path in reverse.

4. File

“organizations”

The placement of this term reflects the name of the file contained in this particular path.

5. File Extension

“.htm”

The file extension tells the computer what program is needed to read the file.

- **.htm or .html: readable by internet browsers**
- **.pdf: readable by Adobe Acrobat Reader**
- **.exe: executable program**

In addition, there are other standard MS Office files which launch independently.

E. Exploring a Web Page

To investigate the features and functions of a selected Web site and its pages, a student at the Information School might take advantage of the iSchool homepage (www.ischool.washington.edu) as a suitable illustration. The following terminology descriptions and discussions correlate with the information given on that site.

1. Web Sites and Web Pages

These two terms are often applied interchangeably when, in actuality, they refer to different levels of *granularity* (i.e., the extent to which something contains separate components):

- **Web site:** a site or location on the World Wide Web, usually consisting of a homepage, the first document users see when they enter the site, and additional documents or web pages.
- **Web page:** a single document or file within a web site.

2. Internal Navigation

Many sites will have *internal navigation bars* to assist the user in progressing through the various pages. On the iSchool homepage, there is a navigation bar across the top, directly underneath the heading “Information School” and one on the left side below the iSchool logo. Sometimes these navigation bars will be in independent scrollable frames; when advancing throughout a site, the frame will continually be visible (see www.ask.com as an example).

3. Hyperlinks

Web pages are labeled as “HTML”, short for Hypertext Markup Language because they employ hypertext. *Hypertext* means that a page may contain active links to move the user from one page to another. The words and phrases that are underlined and colored on the iSchool homepage, such as “MSIM Info” and “Portfolios 2001,” are hyperlinks. Links are not always so obvious: If the cursor is placed over the circular image that contains a picture of an envelope from the same page, the cursor’s arrow becomes a finger and the link bar shows the address for the iSchool’s contact information. Nearly anything can serve as a hyperlink.

4. Graphical User Interfaces (GUI) [pronounced “gooey”]

A *Graphical User Interface* is a program interface (i.e., the set of commands or menus through which a user communicates with a program) that take advantage of the computer’s graphics capabilities to make the program easier to use. It makes use of windows, icons, and pop-up menus that free the user from having to learn complex command languages. From the iSchool’s Web site, an example of the manner in which GUI’s can be exploited may be viewed at <http://www.ischool.washington.edu/technology/subscribe.htm>. The blank text box set-up, which is a traditional type of GUI form, allows the user to interact with the system by inserting appropriate terms in order to receive electronic information. They may also be used when querying a search engine or in sending electronic mail.

5. Streaming Media

Streaming represents a technique used for transferring data in order that it can be processed in a steady and continuous stream. The method allows the audio and video data to start being displayed before the entire file has been transmitted. Within the iSchool’s Web site, the page at <http://www.ischool.washington.edu/realmedia> contains links to streaming media files. It’s possible that some courses at the Information School will incorporate streaming as part of the instruction or as curriculum supplement.

F. Search Engines

A search engine is a program that searches documents in its database for specified keywords and returns a list of documents where the keywords are found. Often the term is used to describe particular systems like Alta Vista and Excite, programs that essentially attempt to organize and retrieve the billions of pages on the World Wide Web. The following are examples of some of the most popular search engines:

1. Yahoo! <<http://www.yahoo.com>>

Although not technically a search engine, Yahoo is a valuable resource, being a subject directory index of collected Web pages that may be searched via a GUI query interface. The search engine maintains its esteem seven years after its development because of its discriminating Web site selection into its database and its competency in relevance sorting.

2. *Google* <<http://www.google.com>>

Google is a “relevance ranked” search engine, meaning that the most relevant documents to the query will be retrieved and listed first. This is done using an algorithm that ranks pages according to how many links a certain web document has from other web sites: the more links, the more relevant the document. Google now also incorporates a directory-style structure in addition to its search engine; accessed from a link below the search box, the directory arranges collected links to information by topic, similar to Yahoo. Google’s easy to use format and retrieval of superbly pertinent results have made it one of the most well-liked search engines available.

3. *Other Engines*

◦ *Northern Light* <<http://www.google.com>>

Northern Light is distinctive from the other search engines because it offers the user the ability to search not only the large web page database but also nearly 7,000 full-text published print journals and an array of online news resources, all in one search command. Another feature enables the formation of a “Customized Search Folder” structure to arrange the search results into more manageable categories.

◦ *Ask Jeeves* <<http://www.ask.com>>

This search engine requires that the user submit an actual question, rather than using keywords. It employs sophisticated natural language processing to understand and match the query to a broad knowledge base, which consists of thousands of question templates and researched answer links to Web sites.

◦ *Dogpile* <<http://www.dogpile.com>>

Dogpile is a meta search engine that covers ten search engines at once. This purports to save the user time, but may inadvertently retrieve non-relevant results.

◦ *iLor* <<http://www.ilor.com>>

iLor is a new search service powered by Google. In bringing back Google results to a query, it allows the user to hover over any of the listings that reveal four management options like “Put in My List” and “Go Now,” which opens up a new window while anchoring the user to the original search results.

◦ *Excite* <<http://www.excite.com>>

Excite uses concept-based indexing by applying statistical methods to actually determine the content of a particular Web page for more relevant results. It also presents the results using summaries, which are basically the top few most important sentences in the document.

◦ *Lycos* <<http://www.lycos.com>>

Although results are not listed by relevancy, this search engine does offer comprehensive information about the retrieved documents, including the date of the document, its size, etc.

G. The Invisible Web

Even though Search Engines, such as Google, can index literally billions of pages there is a growing amount of information on the internet that is not, and cannot be indexed with a search engine. For instance, a company can have access to a database on their website. The company’s website may be indexed on Google, but the content in their database will not. In order to find this invisible information there are some valuable resources to use.

1. *Librarians' Index to the Internet* <<http://www.iii.org>>

Librarians' Index is an annotated subject directory of internet resources that are selected and evaluated by librarians.

2. *Invisible Web .Com* <<http://www.invisibleweb.com>>

This is a corporate endeavor similar to LII, using both a directory structure as well as search engine technology. Invisible web though points directly only to databases, and does not search general html pages.

3. *Invisible Web .Net* <<http://www.invisible-web.net>>

The invisible web net is the online companion to the book The Invisible Web by Chris Sherman and Gary Price, two leading activists for invisible web resources. This is perhaps one of the finest resources on IW.

H. The UW Libraries Information Gateway

The information gateway <<http://www.lib.washington.edu>> is the University Libraries' web presence, and all of the UW's electronic resources can be found there. Among the most important are:

1. *The UW Catalog and Cascade*

The libraries OPAC (Online Public Access Catalog) contains all the books, e-books, journals, e-journals. Cascade is the Washington State Cooperative Library Project that combines the catalogs of seven major institutions to allow for statewide borrowing. The service is free to all students.

2. *Digital Reference Tools*

Numbers of electronic reference resources, dictionaries, encyclopedias, language dictionaries, etc. are organized alphabetically and by subject.

3. *Databases*

The UW subscribes to approximately 365 databases. These are organized in three ways.

1. Straight alphabetical list of every database
2. Top 20 database list
3. Subject categories

Contained in these lists of databases you will find two special designations:

Library CD-ROM Network: In order to use a database with this label you must physically be in one of the UW library units.

UW Restricted: In order to use these databases you need to be either 1) physically in a UW library, 2) Using a computer with a UW IP address (this includes through connectivity kits), or 3) by activating proxy server access.

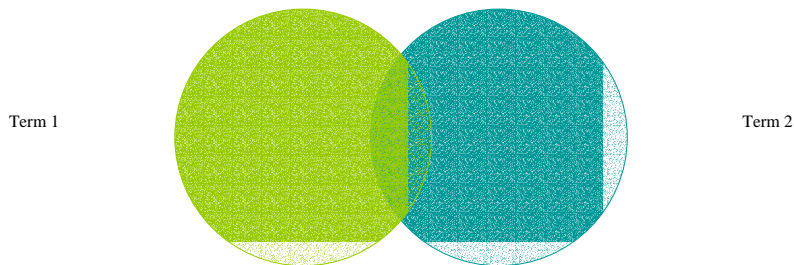
J. Boolean Logic

Boolean logic is a form of algebra in which all values are reduced to either 'true' or 'false.' A concept that will be covered in more depth as part of several iSchool courses, the application of Boolean operators is essential to enable advanced searching techniques available for many Web search engines, indexes, and

other electronic databases. Using the accompanying diagram, it is possible to understand the three most widely used operators:

- **AND** Meaning that the user wants to retrieve all pages containing *both* Term 1 and Term 2. (Documents in the football shaped portion in the center)
- **OR** Meaning that the user wants to retrieve all pages containing *either* Term 1 or Term 2 or both. (Docs in the entire space)
- **NOT** Meaning that the user wants to clarify the definition of Term 1 by rejecting items with Term 2. (Docs on the left, rejecting the overlap)

Venn Diagram



K. Authority for Research

Before citing an Internet resource or recommending another user to a resource, it is very important that some degree of evaluation has been performed on that source. Generally there are approximately six criteria used in source evaluation:

- Authority
- Currency
- Accuracy
- Scope/Coverage
- Objectivity
- Page design

The practice of Web evaluation and resource assessment will be addressed more fully in the required LIS 520 course and other classes within that curriculum “decade.”

Sources used in preparing this document:

- <http://home.earthlink.net/~rosskenw/search.html> -- the online portion to a presentation by Mary B. Ross from Seattle Public Library delivered at the ASIS&T PNC conference September 15, 2001.
- <http://www.webopedia.com/> -- an online basic dictionary of computer terms
- <http://searchenginewatch.com/sereport/01/04-ilor.html> -- web site that offers tips for web searching including analysis of search engines (e.g., this link reviews the iLor search engine)
- <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/SearchEngines.html> and <http://www.lib.berkeley.edu/TeachingLib/Guides/Internet/NorthernLight.html> -- offered by U.C. Berkeley Library, this site, part of their Teaching Library Internet Tutorial, presents a table of recommended search engines
- <http://www.monash.com/spidap3.html> -- another site that describes and ranks various search engines
- <http://dhs.davis.k12.ut.us/library/SEARCH.HTM> -- simplified search tools instruction from Davis High School's (Utah) Media Center
- http://www.english.uwosh.edu/henson/res_doc/webeval1.html -- web site evaluation guidelines offered by Cary Henson, a professor in the English Department of the University of Wisconsin
- <http://www.cnet.com/internet/0-3773.html> - site that reviews and evaluates different web browsers.