Points that were explained and/or written in last class

Aside: Note the need to make progress on your project.

1. Brief about the language all web browsers understand i.e. HTML. It’s basic format.
2. Examples of web browsers (know at least, three popular ones)
3. Things needed to get your website hosted (ISP, username, password, FTP site or address).
4. How to establish dialup connection from your PC.
5. Search Engines
   - How they work
   - Examples
   - Adverts
   - Submitting URLs i.e. your website address
6. eLibraries (becoming electronic)

-----------------------Supplementary notes-------------------------

**HTML** – This can be compared to a computer programming language. It is more or less a series of ‘Tags’ that can be applied to regular text to make it more presentable.

EX: If I am writing something that needs to be bold, I would use the Bold tag: `<b></b>`. Anything that sits in the middle (`<b>bold</b>`) follows the instructions of that particular tag. The word bold would be **bold**.

There are many tags, too many to name here. They can be used to make text bigger, smaller, bolder, underlined, and italic. Then can make paragraphs, lists, tables. It can be used to show pictures, link documents with ‘Hyperlinks’, email addresses, etc.

**General format:**
```
<html>
<title></title>
<body>

......
</body>
</html>
```

The book has an exhaustive HTML section. If you want to learn how to make web pages, read it. You can also use a program (such as Microsoft Front Page, Macromedia Dream Weaver) to create web pages. These programs also have webpage design wizards built in them.

Although most tags in HTML are paired (to signify which section or word to begin and end the format e.g. `<body>` & `</body>`), note that some tags are not paired e.g. `<hr>`
Note:

**HTML – Hyper Text Markup Language**

**HTTP – Hyper Text Transfer Protocol**

These are two tools that are used to make the World Wide Web. As stated before, the World Wide Web (WWW) is made up of Web servers, dedicated computers that display content for your information, and amusement.

### Storing Web Pages:

Web servers have what are called ‘directory structures’. It is just like files and folders on your PC. If you go to a web site it may organize things by content:

EX: [www.csam.montclair.edu/~sannih](http://www.csam.montclair.edu/~sannih) is my web site, and I have several folders that organize things better:

[www.csam.montclair.edu/~sannih/images](http://www.csam.montclair.edu/~sannih/images) is a folder inside of the above folder that contains images.

You will notice that each directory is preceded by a `/`. This is pretty much universal. Files that are in a directory come after the directory name:

[www.csam.montclair.edu/~sannih/images/mainlogosm.gif](http://www.csam.montclair.edu/~sannih/images/mainlogosm.gif)

This is the image file for the MSU logo on my website; there may be other files in the directory that can be accessed the same way.

When you go to [www.csam.montclair.edu/~sannih](http://www.csam.montclair.edu/~sannih) you don’t specify a file name, but the file index.html is loaded for you automatically. This little feature was introduced to save you from typing a lot. This is called a ‘default page’.

Now, let’s review some Internet Techie areas like IP Addresses and Domain Name mentioned in last class

### Addresses:

**IP – Internet Protocol address consists of 4 numbers separated by periods (dots). Each number can be between 0 and 255.**

**EX:** 130.68.1.12  
127.0.0.1

[Quick math: 256 possibilities for each number, 256 * 256 * 256 * 256 = 4 billion possible numbers. It is a lot, but we are running out! In a couple of years we will go to 6 numbers to add a ‘few’ more combinations.]

### Domain Names:
Most people don’t want to remember a number, for example if you were looking for a ‘Wal – Mart’ you wouldn’t necessarily need the exact address, you could see it from miles away. **We can identify web sites by names usually.**

**EX:** Montclair’s Web site is [www.montclair.edu](http://www.montclair.edu).

Now, let’s discuss how to put them together:

**DNS – Domain Name Service (Server)** is a quick way of converting the numbers to the names. A computer (actually many computers world wide) stores the ‘translation’ between the two.

EX: when you type [www.montclair.edu](http://www.montclair.edu) a ‘server’ converts this to 130.68.1.12. Then it takes you to this address. You should then see Montclair’s web site.

**E-Mail:**

*E-mail works in a similar way, let’s say you want to send an e-mail to Bob Smith, and he works at BigCorp.* You know they have a website, [www.BigCorp.com](http://www.BigCorp.com). So you send your e-mail to BobSmith@BigCorp.com. You are sending the email to USER Bob Smith AT the DOMIAN BigCorp.

All of the conversion is done automatically; all you need to know how to do is spell.

It is important to note here that each service provided via the Internet has its own server...e.g. Email Server, Domain Name (DNS) Server, Webpage (HTTP) Server, FTP Server, and NNTP Server among others. We gave examples of some of them in class.

**Domains:**

.com, .org, .net, .mil, .gov, etc. are all DOMAINS that you can have a site in. Some are more restricted than others (Unless you are a Government agency, you can’t use the .gov or .mil domain).

You can ‘rent’ an address in a domain. Say you want [www.hakeem.com](http://www.hakeem.com). Go to [http://www.verisign.com/nds/naming/](http://www.verisign.com/nds/naming/) and see if it is available. If it is, you can rent it on a yearly (or longer) basis for a fee. The fees don’t stop there though! Now that you ‘own’ it, you have to have someone host it for you, to display your content. That will cost you more money per year (per month). Just think if other products you buy worked this way? (Actually they do, think of cable, wireless, internet service, etc.)

**Transporting information:**

*We learned that ‘data’ is nothing more than 1’s and 0’s. How does it move around the internet though?*
TCP/IP – Transmission Control Protocol / Internet Protocol

The author’s example: Imagine mailing a book you are writing from Tahiti to New York. Instead of putting it in box, you decide to write it on post cards, hundreds of thousands of post cards. Is this tedious? Is there a guarantee all will get there? If a couple of post cards get lost, is this better than loosing the whole book?

The designers of TCP/IP didn't want you loosing the whole thing, so they decided that breaking it into manageable chunks – IP Packets – was the smart thing to do. Each Packet is labeled and numbered, so you know where it is going and where it will go in the end when it has to be put back together.

Most everything uses packets, web browsing, e-mail, and instant messaging to name a few. If you lose a packet, your computer may request another to replace the lost one. This allows a pretty reliable system.

Where do the lost packets go? Well, they end up in other people’s hands usually. What do they do with them? Depends on what it is. Let’s say it’s an instant message to your friend about what a homework assignment. That is not that interesting, but a credit card number, that may be useful...More on this when we do security.

Systems:

**LAN – Local area network** – If you are using a hub or router at home, school, or office to connect multiple computers, you are on a LAN.

**WAN – Wide area network** – Computers connected over a large distance, corporate networks fall into this category. This may span through all the entire continents.

**Ethernet** – This is the underlying technology, see ‘party example’ in the book (p 75 – 77). **Note the difference between the communication style the Ethernet and Internet uses. Ethernet used a Broadcast Network to implement Point-to-Point Communication while Internet uses Point-to-Point Network to implement Point-to-Point Communication.**

Connecting to the Internet:

Usually you must buy this from an ISP – Internet Service Provider

You can always be connected (cable, DSL) or you can connect when you want to (dial up).

**World Wide Web vs. Internet**

We’ve already discussed this in a previous class. A lot of people think they are the same, they are not. There is an important distinction. The World Wide Web is web sites, and is powered by Web Servers. The internet is the underlying system to connect computers (think of it as the wires and computers themselves).
Search Engine Basics

Search engines; name your favorite 10 (increasing every day)

Google, Yahoo, Excite, AOL, MSN, Lycos, Dogpile, Ask, etc, etc, etc…

Why are there so many? Do we need that many? Should there be more?

Think about watching the news. There are approximately 10 channels in the area that have news of some form or another (some may do news all day long). Is this necessary?

Is seeing things from a different perspective good or bad? Would you want to only watch news coming from just one source? (Think communism)

When it comes down to it though, its just information – shouldn’t all information show up in the same manner? When you do a search on different engines, you’ll notice that the result displayed looks the same. Although many of you will notice that ads differed from page to page.

Do we really need to see advertising on a search engine? We are after all just looking for information (be it commercial or not). Google needs to pay the bills some how...

Technology behind search engines:

There are basically 2 parts that you should be concerned with:

Front End (GUI, Web Site, etc.) – this is what the user sees. It is (usually) pleasing to the eye and easy to use. You basically tell it what you want (be it information from menus or simply typed in) and it will give it back to you, in a nice neat manner.

Back End (Databases, crawlers [spiders], algorithms, etc.) – This is the stuff you don’t see. In fact, no one really sees it except the people who designed / run the search engine. The parts:

Database – stores all the information that is gathered from the web. It is categorized in many different ways, and will respond to the user’s questions.

Crawler(spider) – physically searches the web, taking ‘snapshots’ of the information on every webpage it comes across. There are millions of these that search the web, 24 / 7 (Although apparently they search less than ½ of all information). They will store what they find into the database.

Algorithms - Basically the way the information is searched for, the method. Each search engine has its own way of doing this; this is what makes for the different results you may get, even for the same question.

All of these items are nothing more than software, written for the sole purpose of serving the searching public. They run on lots and lots of
expensive pieces of hardware. Thousands of servers are dedicated to the task of running the crawlers, powering the databases, and running the website so people can search.

Libraries:

Remember libraries? Big (or small) storehouses of information; in the past may have just contained printed works; they now contain everything from microfilm to DVD’s of information.

Libraries Becoming Electronic:

Because we now have the technology, why can’t we just electronically scan every printed work, and have it available in digital format? We are, little by little.

Is this a big job? Of course it is. Start slow though...begin with current works, they should be easy to get to electronic media (if they are already not in that form). Then scan the oldest (most fragile) things before they disappear for good.

Interesting sites:

Project Gutenberg - [http://promo.net/pg/](http://promo.net/pg/)
The project is dedicated to having individuals transcribe entire books. These books are then available for download – for free. Of course you are not going to find the latest Stephen King books here, mostly just the old classics that you were forced to read in high school. Why? Remember those copyrights.

University of Texas Ransom Center –
The university has invested a lot of money into equipment to convert books to digital format. One of their proudest accomplishments is digitalizing an original bible, printed by Johann Guttenberg in ~ 1454. This is one of the first books created using movable type. There are only 5 of these in the US at this time.

Other Libraries, Online:

Pick a major school, they probably have one; Penn State, University of Washington, etc.

Online resources are far from complete. These are really more of a way to wet your appetite. You will still have to come to the library to get most sources.

Most magazines and newspapers store extensive archives on their own websites. Some are free, others are not.

Some scientific journals are available online, most often these require a membership fee or a ‘pay per use’ fee.

Government agencies and departments house lots of free information, The Census bureau, the Department of Labor, Department of transportation, etc. The library of congress is also online: [http://www.loc.gov/library](http://www.loc.gov/library)