Summer 2000

Academic Advantage

Wright State University

Dayton, Ohio 45435

Computer Literacy and Applications

Prabhaker Mateti

Department of Computer Science and Engineering

pmateti@cs.wright.edu (937) 775 5114

Contents

Chapter 1

Introduction to AAP Computing

We are sure that you are already into computers! Our goal is to provide a deeper understanding of what you already know. And to expose you to some new things.

1.1 WSU Computing Resources

WSU has extensive computing facilities. Many colleges run their own computer labs.

1.1.1 CaTS

Computing and Telecommunications Services (CaTS) is a University wide facility. They are located in the Library Annex Basement. Help Desk's number is 775-4827. Visit:

- `http://www.cats.wright.edu/catsweb/cs/students/students.html` for a full description.
- `http://www.cats.wright.edu/catsweb/cs/labs/labmap.htm` for a point-and-click interactive map of the CaTS labs in the Library and Library Annex.
- `http://www.cats.wright.edu/catsweb/docs/` for extensive on-line documentation.

- `http://www.cats.wright.edu/catsweb/cs/docroom/unix/personal.htm` to find out how and where you can setup a personal web page and have it accessed from all over the world.

1.1.2 College of ECS

College of Engineering and Computer Science is located in the Russ Engineering Center.

Visit the College's home page: `http://www.cs.wright.edu`.

The College maintains its own computing facilites. There are over a dozen computer labs. Some of them require "card access." Here are a few that are always open:

X-terminal Laboratories, 152B & 152D Russ Engineering Center
PC Laboratory, 152C Russ Engineering Center

1.1.3 Computing in AAP

Each day for the next five days, you will listen to a lecture from a few minutes to at most 45 minutes, followed by hands-on practice in a computer laboratory. Our lecture hall is the same as lab (152A Russ).

Here is a list of our topics.

- WSU Computing Resources
- Unix Essentials
- Introduction to Mathematica
- Exploring the Web
- Structure of Web
- Internet Internals

What you are reading now is reading material for the above topics. You must still take good notes. Follow the Success Strategies.

Our PC Laboratory for the AAP is 152A Russ. There will be student lab assistants to help you in any way they can.

Chapter 2

Unix Essentials

Unix operating system has been in use at universities for three decades that it is a culture. It has matured nicely, and even so is extremely agile compared to Windoze. Because it typically ran, until recently, on expensive machines, the number of Unix installations was small compared to Windows. In 1998, a version of Unix, called Linux, represented 17 percent of new-license shipments of server operating systems.

A good book to read is "Unix for the Impatient" by Paul W. Abrahams, Bruce Larson, Addison-Wesley Pub Co; ISBN: 0201823764.

The rest of this chapter is a mildly edited version of a document I inherited from Joe Slater.

2.1 Introduction to Unix

1. Logging in and out.
2. The shell.
3. `xterm`[1]
4. `mathematica`
5. `Netscape`

Note that Unix is CasE SEnsItIVe

[1] Text that is in `this font` represents actual commands that you type. Some of them take additional arguments.

2.1.1 File management

1. `ls` - list all files in current directory
2. `cd` - change current directory
3. `pwd` - present working directory
4. `rm` *filename*[2] - deletes the file *filename*
5. `cp` *path/filename newpath/newfilename* - copies the file *path/filename* to *newpath/newfilename*
6. `mv` *path/filename newpath/newfilename* - moves the file *path/filename* to *newpath/newfilename*
7. `more` - list contents of a file a screen at a time
8. `script` *filename* - send all terminal display to *filename*
   type `exit` to stop logging everything to *filename* (it will not log you out)

2.1.2 Editing

1. `emacs` - an editor
2. `axe` - invoke axe (an X-editor)
3. `pico` - an editor
4. `nedit` - another editor
5. `cat` - concatenate, list contents of, or create files
   (a) `cat` *filename1* - list *filename1* to screen
   (b) `cat` *filename1 filename2* - lists *filename1* & *filename2* to the screen

2.1.3 Job/Process Control

Any task being performed by the system is called a process. Any process started by a user is called a job. When an application is executed, it becomes a process/job, basically meaning that the computer is thinking about it. Before it is executed, it is just a bunch of instructions on disk waiting to be executed.

1. *application* & - run the application in background. A job running in the background has lower priority than those running in the foreground.

[2]Text in *this font* represents a generic name or command where you have to choose the specifics.

2. <ctrl-z> - suspend current job. *It keeps running in the background! Don't forget to kill it or go back to it!*
3. `bg` - the current job is put in background
4. `jobs` - list running jobs
5. `kill` - kill a job, e.g., `kill %3` kills job number `3`.
6. `logout` or `exit` - logout
7. <ctrl-d> - end of input - can be used to quit most command line applications & log out
8. <ctrl-c> - will break out of the current foreground process and terminate it.

2.1.4 Getting Information

1. `who` - shows who is logged onto the system
2. `finger` - see who is logged on, or get specific information
3. `man` - look for info in a manual page about a command
4. `help` - get help on a specific topic
5. `uptime` - get information about the processor load

2.1.5 Transferring files to other machines - `ftp`

1. connecting
2. `anonymous` connections
3. getting a directory, changing directories
4. binary and ASCII file types. (`binary` and `ascii`)
5. `put`ting a file
6. `get`ting a file
7. getting and putting multiple files - `mput` and `mget`

2.1.6 Printing

1. `lpr -P`*`printername filename`* - send a file named *filename* to the printer
2. `lpq -P`*`printername`* - see where your print job is in the queue
3. `lprm -P`*`printername`* *print job #* - remove your print job from the queue

2.1.7 Command Redirection >, <, |

1. Output Redirection, >
   (a) `finger > user` - puts list of users logged on into file `users`
   (b) `ls > user` - replaces the contents of the file user with the directory
   (c) `finger >> user` - appends the list of users logged on to the directory listing in the file `user`
2. Input Redirection, <
   (a) `wc-l users` - counts the lines in the file users
   (b) `wc-l < users` - does the same thing, but `wc` doesn't ever know where the info came from
3. Pipes, |
   (a) | Connects two commands together
   (b) `finger | wc-l` - Counts the # of lines returned by the `finger` command
4. Complex examples
   - `finger | sort > users` - Lists who is logged on, sorts the list, then puts the list in the file `users`
   - `cat` *filename1 filename2* `>` *junk* - puts *filename1* & *filename2* into *junk*
   - `(cat` *filename1*`;finger) >` *junk* - lists *filename1* and results of `finger` and places results of () into *junk*
   - `cat >` *junk* -Allows you to type into a new file *junk* until a <`ctrl-d`>

Chapter 3

Electronic Mail

E-mail has been the most popular use of the Internet. In the U.S., more e-mail is now exchanged than postal mail. Because e-mail "letters" travel at electronic speed, people sometimes refer to the traditional postal system as "snail mail."

3.1 E-Mail Terminology

E-mail is short for "electronic mail."

Header is a few lines of text about who sent the message, when it was sent, the title of the message (in its subject line), who else received the message, and how it got to your mailbox.

Recipient is the person to whom you're sending the message.

Subject line is a brief description of what the message is about.

Body The body is the content of the letter, generally text but can be a graphic.

Signature A standardized message closing that you can design.

HTML e-mail allows you to create messages with stylized text and pictures so that your note looks like a Web page. The recipient should have a mailer that can display such mail though.

Actions (sometimes called Rules or Scripts) are specified operations that are automatically performed by the mail program upon received mail.

Address book (sometimes called Contacts List) stores the e-mail addresses.

Importing addresses If you've previously used another e-mail program, you can import your old addresses into your new program. First, export the addresses from your old program. Then, import them into the new program.

Download Folder is a directory for the storage of received attachments, allowing you to find the attachments and to scan the selected location for viruses.

Compression shrinks attachments to make them quicker to send. Enable it.

Encoding converts attachments into a sendable format. Consult a local guru before you select a type.

Helper applications are other programs that can handle attachments.

"Emoticons" are symbols at the end of sentences that underscore the sentence's meaning. Examples: `:-)` `:-(` `;-)`.

Abbreviations provide a shorthand for commonly used phrase and expressions.

ASCII art is an illustration comprised entirely of letters and typographic symbols.

> symbols indicate that the text that follows them has been copied, or "quoted," from another message.

Attachment A message may arrive with an accompanying file called an attachment – a word processing document or a picture, for example.

3.2 `pine`

Unix has many many mail programs. One of the easiest, and archaic, to use is `pine`. Run it by just typing its name: `pine` at the shell prompt. It will then present you with a help screen the first time you use it.

3.3 Set Up Your E-mail Program

If you are on a PC or Mac connected to the Internet either via a modem or a LAN, you need to provide your e-mail program (such as Outlook Express, Nestscape Messenger, or pine) the following information.

Your e-mail address identifies your unique address on the Internet for receiving e-mail. At WSU, it's in the form of: yourname@mail.wright.edu.

Your Post Office Protocol (POP or POP3) account name is generally the name of your Unix (e-mail) account.

Your POP server name – If you've got a dial-up Internet account, the POP (usually POP3) server handles your incoming mail. Your Internet service provider will tell you this name. At WSU, it is `mailhost.wright.edu`.

Your Simple Mail Transfer Protocol (SMTP) server name – If you've got a dial-up Internet account, the SMTP server handles the mail you send to other people. Your e-mail program may allow you to maintain multiple accounts – for example, home and business e-mail accounts. Each account must be set up individually within your e-mail program. At WSU, it is `mailhost.wright.edu`.

Once you have this information set up, you don't need to change it again.

3.4 Receive Messages

To receive your e-mail, you must be currently connected to the Internet. Some e-mail programs automatically attempt to retrieve your messages when the program is launched. If you're offline, the program can attempt to connect your computer to the Internet. If your e-mail program is already active, it will provide you a menu item or button that activates the receiving of e-mail. Most e-mail programs provide icons or windows for an Inbox for incoming (and hence unread) mail, and an Outbox for mail messages that are ready to be sent.

3.5 Reply to, Write and Send a Message

In composing a message, use standard capitalization practices. A message in ALL UPPER CASE LIKE THIS is considered "shouting" and is regarded as a sign of inexperience. You should also feel free to incorporate emoticons and e-mail abbreviations. If you can, you should spell-check.

All e-mail programs offer a Reply feature as a menu item or button that you select while reading the message to which you want to reply. When you do this, the program will create a new message addressed to the original sender, and can copy the original message – or any currently selected text. You can then type in your responses wherever you wish and send the message.

When you've written a reply or a new message, you can queue it for sending at a later time, or send it immediately. When you do this, the program will store the message in its Outbox.

To actually send a message you must be connected to the Internet. If you attempt to send an e-mail message when offline, many e-mail program will automatically attempt to connect you to the Internet.

Chapter 4

Mathematica

Mathematica is a software package by Wolfram Research, `http://www.wri.com`. It can do computations in the usual way: numerically – to as much accuracy as needed. But, its real power is in symbolic, and graphical computations. Mathematica can manipulate algebraic formulas, and do calculus and present the answers symbolically as you would in a Math class. Mathematica can also generate excellent two- and three-dimensional graphics.

Mathematica is available on the PCs in 152A Russ as well as the Unix machine paladin. Type `math` at the Unix prompt to run Mathaematica run in text based mode. Type `mathematica` at the Unix prompt to run Mathaematica run in the notebook mode.[1]

4.1 Once Inside Mathematica ...

Mathematica is CasE SEnsItIVe.

4.1.1 Getting Out and Help

- You must hit shift-Return to tell Mathematica to execute a line
- `Quit` - Exit Mathematica
- `Help` - Use the help menu.
- `?`*command* - get help on *command*
- `??`*command* - get extensive help on *command*

[1]The rest of this chapter is a mildly edited version of a document I inherited from Joe Slater.

- `?comm*` - get list of commands starting with the letters *comm*
- `?*` - get list of all commands

4.1.2 Syntax of Mathematica's Expressions

Use {} as delimiters for lists, and `[]` as delimiters for command arguments. We use the familiar `+` for addition, - for subtraction, `*` for multiplication, `/` for division, `.` for matrix multiplication, `^` for exponentiation, `Pi` for the constant π. So, `2^100` gives `1267650600228229401496703205376`.

Following typical books in math, we can also write y next to x, as in $x\ y$ to mean $x * y$.

The trig functions are `Sin[x], Cos[x], Tan[x], Abs[x], ArcSin[x]`, etc. Note the lower-upper spelling and the brackets.

To evaluate *expr* to *n*-digit precision write `N[expr,n]`. `2^100//N` gives the numerical approximation $1.26765\,10\times^{30}$

You can obtain the result of previous line with `%`, and the result of line n with `%n`.

Mathematica can substitute symbolically. For example if you wish to substitute x for y and 2 for z, you use the substitution operator (the "slash-dot") as in `x y^2 z^3/.{y->x, z->2}` yields `8 x^3`

4.1.3 Changing the form of expressions

Multiply out Products and Powers: `Expand[(x+1)^2]` yields `1 + 2 x + x^2`

Reduce to Factors: `Factor[1+x+x^2]` yields `(1 + x)^2`.

To give smallest form of expression use

```
Simplify
```

4.2 Solving equations

4.2.1 Solve polynomial equations analytically

Example: `Solve[A x^2 + B x + C == 0,x]` yields

$$\{\{x \to \frac{-B - \sqrt{B^2 - 4\,A\,C}}{2\,A}\}, \{x \to \frac{-B + \sqrt{B^2 - 4\,A\,C}}{2\,A}\}\}$$

Example: `Solve[4 x^2 + 3 x + 2==0,x]//N` yields
$x \to -0.375 - 0.599479I, x \to -0.375 + 0.599479I$

Solve the following equations for x and y: $xy^2 = 4$ and $y^3 - x = 7$. An obvious solution is $x = 1$ and $y = 2$. Are there more solutions? Solve the first equation for x such that $x = \frac{4}{y^2}$ and substitute the result into the second equation so that $y^3 - \frac{4}{y^2} = 7$. Multiplying by y^2 the equation becomes $y^5 - 7y^2 - 4 = 0$, so there are 5 solutions, we only got one.

Using `Solve[x y^2==4,y^3-x==7,{x,y}]//N` yields the results

$\{\{x \to 1., y \to 2.\},$

$\{x \to -0.46136 + 1.03647\,i, y \to -1.02273 + 1.57473\,i\},$

$\{x \to -0.46136 - 1.03647\,i, y \to -1.02273 - 1.57473\,i\},$

$\{x \to -7.03864 + 0.425486\,i, y \to 0.0227333 - 0.752821\,i\},$

$\{x \to -7.03864 - 0.425486\,i, y \to 0.0227333 + 0.752821\,i\}\}$

4.2.2 Solving equations numerically

`NSolve[{x y^2==4,y^3-x==7},{x,y}]` yields the same result as item 4.2.1 but faster and more directly.

Example `FindRoot[x Sin[x]-1/2==0,{x,1}]` yields the answer $\{x \to 0.740841\}$. The `{x,1}` means solve for `x` with an initial guess of 1. If plot the function using `Plot[x Sin[x] -1/2,{x,0,10}]` we see that there are multiple solutions, and we have to be careful about which one we get.

Example: Solve the following two equations for x and y: $\sin x + \cos y = 0$, $\cos x + \sin y = 0$.

`FindRoot[{Sin[x]+Cos[y]==0,Cos[x]+Sin[y]==0},{x,0},{y,0}]` yields

$\{x \to 2.35619, y \to 2.35619\}$

You must always check your answers in other ways. The actual answer is

$y = -x - \pi/2 + 2\pi n$

where n is any integer. This can be shown by substitution. How did I know this? If I square the two equations, they are still true. However, the left hand sides are always positive. If you do a `ContourPlot` of the sum of the squares of the left hand sides, you can see that there are lines where it is equal to zero. This type of function is called a *cost function* or *error function* and is commonly used to represent cumulative error when you have multiple equations.

4.3 Matrices

1. `A={{1,2},{3,4}}` creates a matrix $M = \begin{matrix} 1 & 2 \\ 3 & 4 \end{matrix}$
2. `MatrixForm` displays a matrix in a conventional form.
3. `B={{1},{2}}` creates a column vector $M = \begin{matrix} 1 \\ 2 \end{matrix}$
4. `C={3,4}` creates a row vector $M = \begin{matrix} 3 & 4 \end{matrix}$
5. `Inverse` inverts a matrix
6. Example: Solve
   $$\begin{matrix} a & b \\ c & d \end{matrix} \begin{matrix} x \\ y \end{matrix} = \begin{matrix} 1 \\ 2 \end{matrix}$$
   `A={{a,b},{c,d}}`
   `B={{1},{2}}`
   `C=Inverse[A].B` gives
   `C=`$\{\{\frac{-2\,b+d}{-(b\,c)+a\,d}\},\{\frac{2\,a-c}{-(b\,c)+a\,d}\}\}$
   Stick in any numbers for `a`, `b`, `c`, and `d` to check that this correct.

4.4 Plotting Examples

1. 2-D: `Plot[Sin[x],{x,0,10}` plots the sine of x from 0 to 10.
2. 3-D: `Plot3D[Sin[x y],{x,0,4},{y,0,4}]`
3. Animation: `Animate[Plot3D[Exp[(-x^2-y^2)*.15]Cos[x y+t],{x,0,4},{y,0,4},PlotRange->{-1,1}],{t,0,2Pi}];`
4. Try replacing the plot command in item 3 with some other variation.

4.5 Calculus

1. `D[`*function*`,{`*variable*`,`*number of times*`}]` takes a derivative.
   Example: `D[x^2,{x,1}]` yields `2 x`
   Example: `D[Sin[x^2],{x,1}]` yields `2 x Cos[x^2]`
2. `Integrate[`*function*`,`*variable*`]` integrates *function* with respect to *variable*
   Example: `Integrate[2x Cos[x^2],x]` yields `8/3`

3. `Integrate[`*function*`,{`*variable, lower limit, upper limit*`}]` integrates *function* with respect to *variable* between the given limits
   Example: `Integrate[x^2,{x,0,2}]` yields `8/3`

4. `Limit[`*function*`,`*variable*`->`*limit*`]` determines the value of the function in the limit.
   Example: `Limit[Sin[t]/t,t->0]`

Chapter 5

Exploring the Web

Web and Internet are two different things. The Internet is the world-wide collection of network-connected computers. The World Wide Web is a collection of interlinked documents that work together using protocols such as HTTP running on computers connected to the Internet.

5.1 Browsing

The Web uses a metaphor of individual pages, usually combined to make up sites. Web pages are written in HTML (HyperText Markup Language), which tells the Web browser how to display the page and its elements. The enabling feature of the Web is its ability to connect to one another text pages, as well as to audio, video, and image files with hyperlinks.

5.1.1 URLs

The following are examples of URLs (Uniform Resource Locators).

- `http://www.cs.wright.edu/`
- `mailto:pmateti@cs.wright.edu`
- `file://U:/pmateti/HTML/PM/pmip0.html`

The first part of the URL (before the two slashes) specifies the method of access. The second is typically the address of the computer the data or service is located. Further parts may specify the names of files, the port to connect to, or the text to search for in a database. A URL is always a single unbroken line with no spaces.

Here are some more examples of URLs.

- `file://www.hcc.hawaii.edu/sound.au` - Retrieves a sound file and plays it.
- `file://www.eit.com/picture.gif` - Retrieves a picture and displays it.
- `file://www.eff.org/directory/` - Displays a directory's contents.
- `http://www.hcc.hawaii.edu/directory/book.html` - Connects to an HTTP server and retrieves an HTML file.
- `ftp://www.xerox.com/pub/file.txt` - Opens an FTP connection to Xerox and downloads a text file.
- `gopher://www.hcc.hawaii.edu` - Connects to the Gopher at Univeristy of Hawaii.
- `telnet://www.hcc.hawaii.edu:1234` - Telnets to www.hcc.hawaii.edu at port 1234.
- `http://mc5.go2net.com/crawler?general=what+is+a+router &method=2&region=0&rpp=20&timeout=5&hpe=10&format=regular&sort=0` This complicated URL is the result of a search request made to MetaCrawler.
- `news:alt.hypertext` - Reads the latest Usenet news by connecting to a user-specified news (NNTP) host and returns the articles in the alt.hypertext newsgroup in hypermedia format. Most Web browsers allow the user to specify a URL and connect to that document or service. When selecting hypertext in an HTML document, the user is actually sending a request to open a URL. In this way, hyperlinks can be made not only to other texts and media, but also to other network services. Web browsers are not simply Web clients, but are also full-featured FTP, Gopher, and telnet clients.

5.1.2 Search Sites

There are three primary types of search sites on the Web: search engines, Web directories, and parallel and metasearch sites.

Search engines such as Excite and HotBot use automated software called Web crawlers or spiders. These programs move from Web site to Web site, logging each site title, URL, and at least some of its text content. The object is to hit millions of Web sites and to stay as current with them as possible. The result is a long list of Web sites placed in a database, which users search by typing in a keyword or phrase.

Web directories such as Yahoo and Magellan offer an editorially selected, topically organized list of Web sites. To accomplish that goal, these sites employ

human editors to find new Web sites and work with programmers to categorize them and build their links into the site's index.

Since both approaches make sense, all the major search engine sites now have built-in topical search indexes, and most Web directories have added a keyword search.

Parallel and metasearch sites ride piggyback on the Web crawler sites. Parallel search programs, such as Vironix Software's WebFerret, launch simultaneous searches on all the popular search engine sites, returning all the results in a single window.

Metasearch sites go a step further. One of the problems with searching on the Web is that the searching vocabulary varies from search site to search site. For example, when you search for Cretaceous Mongolia on Yahoo, the search term should look just like that. But the same search performed at Infoseek would be more effective if you entered `+Cretaceous +Mongolia`; at Galaxy, it should be `Cretaceous AND Mongolia`. Metasearch sites, such as `metasearch.com`, take care of this for you. They let you enter a term in a single field and then automatically account for all the particulars for half a dozen or more popular search sites.

5.2 Newsgroups

Newsgroups are electronic discussion forums. In spite of their name, there is hardly any "news" them. They are opinions, facts, and debates for people with shared interests. The messages are presented in a list, known as a thread, that shows the original message, the responses to the message, and the responses to the responses, so that you can follow an entire discussion or just the parts you're interested in.

Your browser has a companion program, such as Netscape's Collabra or Microsoft's Outlook Express, that will help you read the newsgroups.

Your newsreader lets you check newsgroups the way your browser lets you surf Web sites. The Usenet is the world's largest collection of public newsgroups. The newsgroups go by a complex set of abbreviated names, with the first set of letters of a newsgroup's name indicating its primary subject, such as rec (recreation), soc (society), or comp (computers). Additional abbreviations are separated by periods and are tacked on to indicate subtopics. It's not uncommon for an individual newsgroup to have five, six, or more elements in its name. For example, microsoft.public.inetexplorer.ie4.setup is a newsgroup devoted to people who want to set up Internet Explorer 4.0. .

The messages in newsgroups are stored on news servers owned by ISPs, universities, companies, and other large entities all over the world. Most news servers keep only the more recent posts; they'd soon run out of storage space otherwise.

What happens if you can't find a newsgroup that covers your favorite topic? (Not rose gardening–there are plenty for that.) Well, you could create a new newsgroup–but not without a little effort. If you want your group to be a standard Usenet newsgroup (those whose names begin with comp, misc, news, rec, soc, sci, and talk), you must submit a highly bureaucratic document, called a Request For Discussion (RFD), to the news.groups newsgroup. The group then organizes a straw vote where anybody who wants to can vote on your proposal. (To find out more about the RFD process and how to write an RFD proposal, see Jon Bell's Creating New Newsgroups page.)

The alt newsgroup hierarchy was created because many people felt it was too difficult to create an ordinary newsgroup. (Contrary to popular belief, alt does not mean "alternative topics"; it means "alternative newsgroup management structure.") If you want to create a newsgroup without all the hassle, you post a suggestion in the alt.config newsgroup and leave it up to the news administrators–the ones who make the ultimate decision about carrying new alt newsgroups. For a guide to creating an alt newsgroup, read David Barr's So You Want to Create an Alt Newsgroup.

There are also such things as local and private newsgroups. A discussion group created on a corporate intranet is an example of a private newsgroup. Most ISPs offer a handful of local newsgroups where they make tech support announcements that no one but their customers would want to see.

5.3 Privacy and Security

There are two types of trouble on the Net: threats to security and threats to privacy. Potential security bogies include viruses contained in file downloads, rogue ActiveX controls that can crash your computer, malicious email attachments, and a host of other weaknesses in your TCP/IP software.

For comic relief and to experience serious nuisance, visit `www.digicrime.com`, which bills itself, tongue-in-cheek, "A full service criminal computer hacking organization.", `rootshell.com` which has an extensive collection of material on attack methods, and `www.mc2.nu` which "is a collection of exploits and utilities for your hacking needs."

Despite all the press reports, the odds are against your becoming the random victim of a hacker. You're much more likely to run into a virus. Avoid opening email attachments from people you don't know, and use good judgment about paying with credit cards on the Web.

Threats to your privacy are more subtle. Make your email safe from prying eyes by using an encryption program such as Pretty Good Privacy (PGP). Encryption software translates your message into a secret code so that it can be read only by the person who has the correct decryption key–that is, the person you're sending it to.

Many Web sites are actually programmed to harvest information about any visitor who comes through. They send small files called "cookies" to your hard drive. Each cookie can be read only by the Web server that created it, for up to days or months later when you revisit the site. Among other things, cookies permit Web sites to track your name, your email address, your ISP's name, the last site you visited, your operating system, and your browser's specific make and version number. Of course, they can also help you out by storing passwords so that you can get into subscription-only sites without having to type the password every time. But the savvy consumer will want to control what information is being collected. You can have your browser refuse to accept any cookies or use a program such as CookiePal to track and manage your cookies.

Information circulated on the Web can help a marketer construct a consumer profile of you. Once they have your email address, they can then besiege you with sales pitches. Online telemarketers are annoying, but information collected online can also be used in more sinister ways, such as sending you obscene emails. For example, every time you post to a Usenet newsgroup, your email address becomes available to everyone who reads the group–and some newsgroup postings stick around for years. One way around this problem is to use an anonymous remailer service such as Replay, which forwards your email without your address.

Whenever you enter your name, address, and phone number in a form on the Web, that information could be going to people you don't know, so think twice before revealing personal info.

5.4 Creating your own Web

A web can be of just one page, or many pages linked in a complex way. A " web page" often prints as several printed pages. Its look and feel is governed by tags in HTML (HyperText Markup Language), and little programs written in the programming languages Java, and JavaScript, or in ActiveX, etc.

First, check out a few simple design examples by visiting

`http://www.cats.wright.edu/catsweb/images/default.html`.

You can save these pages and start editing them as you like.

Here is a list of steps from `whatis.com` in Creating a Web Site.

1. Aim, Set objectives, Define your audience.
2. Plan the content, Tips for planning content, Some content ideas. Involve your audience, Build the site structure, Plan navigation and page consistency, Define a creative direction, Some creative direction possibilities, Working with a graphic designer, Select a page creation tool or editor, Select visual tools.

3. Create the HTML pages, Using tables, Creating forms, Add visual elements, Scanning images, Scanning step-by-step, Consider clip art, Creating animated GIFs, Transparentizing images, Making image maps.

4. Keep content dynamic, Add interactivity, Quizes, games, and contests, Discussion groups — Adding sound, Adding video, Adding 3-D, Other interactive approaches, Test your site, Put your site on a server, Publicize your site.

5. Manage your site Creating a Web Site.

5.4.1 Learn HTML – Not?

To create a Web page without learning HTML, use a Wysiwyg HTML page creator, such as Microsoft's FrontPage 2000, FrontPage Express, or Netscape's Composer.

For greater control over your web page, learn HTML, and use a good text editor such as Emacs. Here is a sub-list of HTML tags; the tags are separated by vertical bars.

```
| <A>...</A> | <ADDRESS>...</ADDRESS> | <APPLET>...</APPLET> | <AREA>
| <B>...</B> (bold font) | <BASE> | <BASEFONT> | <BIG>...</BIG> (big
font) | <BLINK> | <BLOCKQUOTE>...</BLOCKQUOTE> (indent for a quotation)
| <BODY>...</BODY> | <BR> | <CAPTION>...</CAPTION> (TABLE tags) | <CENTER>
| <CITE>...</CITE> | <CODE> |comment | <COL> | <COLGROUP> | <DD> (definition
for a term) | <DFN>...</DFN> | <DIR>...</DIR> | <DIV>...</DIV> |<DL>...</DL>
(definition list) | <DT> (definition term) | <EM>...</EM> | <FIELDSET>
| <FONT>...</FONT> | <FORM>...</FORM> | <FRAME> | <FRAMESET>...</FRAMESET>
| <HEAD>...</HEAD> | <HR> | <HTML>...</HTML> | <H1>...</H1> through
<H6>...</H6> | <I>...</I> (italic font) | <IFRAME> | <IMG> | <INPUT>
(a FORM tag) | <ISINDEX> | <KBD>...</KBD> | <LEGEND>...</LEGEND> |
<LI> (list item) | <LINK> | <MENU>...</MENU> | <META> | <NOFRAMES>...</NOFRAMES>
| <NOSCRIPT>...</NOSCRIPT> | <OBJECT>...</OBJECT> | <OL>...</OL> (ordered
list) | <OPTION> (a FORM tag) | <P>...</P> | <SAMP>...</SAMP> | <SELECT>
(a FORM tag) | <SMALL>...</SMALL> (small font) | <STRIKE>...</STRIKE>
(strikethrough font) | <STRONG> | <STYLE>...</STYLE> | <SUB>...</SUB>
(subscript) | <SUP>...</SUP> (superscript) | <TABLE>...</TABLE> (TABLE
tags) | <TBODY>...</TBODY> | <TD>...</TD> (TABLE tags) | <TFOOT>...</TFOOT>
| <TEXTAREA>...</TEXTAREA> (a FORM tag) | <TH>...</TH> (TABLE tags)
| <THEAD>...</THEAD> | <TR>...</TR> (TABLE tags) | <TT>...</TT> (typewriter
font) | <U>...</U> (underline) | <UL>...</UL> (unordered list) | <VAR>...</VAR>
|
```

5.4.2 Publishing

After you've built and tested your Web site on your own computer, you put the files on a computer that has a Web server so that your site will be accessible to the world. This is known as *publishing*. And, once your site is part of the Web, you'll want to update files on some periodic basis.

The final step is to send files. You'll need ftp or web-publishing software to do this.

One of your html files - the first page - has to be named either "default.htm" or "index.htm" for your site to work properly. Server software enforces a rule that once a connection is made to a directory it will automatically load and display one of these two pages. Your ISP can tell you which one works for his setup. (Some servers still enforce the use of a ".html" instead of ".htm" at the end of your files.)

Organize your files into logical subdirectories. In a small site put all your gifs and jpegs in a folder/directory called something like "images". Make sure your html reference code then says "images/picture.jpg". Larger sites require additional category folders (sounds, archive, databases, etc.).

Chapter 6

Structure of Web

The goal of this chapter is to go behind the scenes and explain how the Web works. For a very detailed view, explore the site `http://www.w3.org/`.

6.1 Web Browsers

Web software is designed around an architecture where computers are connected, but may be geographically very far apart. A Web browser is a program which sends requests for documents to a Web server. A Web server is a program that, upon receipt of a request, sends the document requested. Because the task of document storage is left to the server and the task of document presentation is left to the client, each program can concentrate on those duties and progress independently of each other.

Here's an example of how the process works: Running a Web client, the user selects a hyperlink in a piece of hypertext connecting to another document - "The History of Computers", for example.

The Web client uses the address associated with that hyperlink to connect to the Web server at a specified network address and asks for the document associated with "The History of Computers".

The server responds by sending the text and any other media within that text (pictures, sounds, or movies) to the client, which the client then renders for presentation on the user's screen.

6.2 What is in a Web Page

6.2.1 HTML - The HyperText Markup Language

The Web uses HTML for creating and recognizing hypermedia documents. HTML documents are nothing more than standard 7-bit ASCII files with formatting codes that contain information about layout (text styles, document titles, paragraphs, lists, formatted tables, graphics, ...) and hyperlinks. HTML allows document writers to separate information from document presentation - that is, documents containing the same information can be presented in a number of different ways. Users have the option of controlling visual elements such as fonts, and paragraph spacing without changing the original information.

The current standard is HTML 4.0.

6.2.2 Java and ActiveX

Both Java and ActiveX are technologies that let programmers create animated and interactive Web pages–the kinds that move, flash, and play games. HTML, in contrast to Dynamic HTML, can't do much to make a page interactive; Java, JavaScript and ActiveX fill that void.

Sun's Java is a programming language similar to the popular C++ language used to make applications like word processors or spreadsheets. Java programs – known on the Web as applets – have the unique ability to run on any operating system, from Windows to Mac to Unix. While lots of applets are available as shareware and can be plugged ready-made into Web pages, programming Java from scratch is not a Saturday afternoon activity.

Don't confuse JavaScript with Java. JavaScript is a scripting language, a special kind of programming language used to tie other components together or to accept user input. While the names are confusingly similar and both technologies are designed to make Web pages more interactive, Java and JavaScript are two different things.

An ActiveX control is roughly equivalent in concept and implementation to the Java applet. ActiveX is a little harder to define than Java. That's because Microsoft has chosen to make the term "active" a major part of its Internet marketing campaign, but there are so many active things from Microsoft that the word begins to lose meaning. However, ActiveX controls are roughly equivalent to Java applets, in that they run in Web browsers and are designed to enhance Web pages. But where Java is a full-fledged programming language, ActiveX controls are not cross-platform, work only with Internet Explorer, not Netscape's browsers (without a plug-in).

Java, JavaScript, VBScript and ActiveX are not mutually exclusive–they can work together. But Sun and Microsoft are both trying hard to make Web

developers loyal to one over the other.

6.3 Search Engine Glossary

There's searching, and then there's finding. It is important to understand how to formulate efficient info quests.

Boolean search: A search allowing the inclusion or exclusion of documents containing certain words through the use of operators such as AND, NOT and OR.

Concept search: A search for documents related conceptually to a word, rather than specifically containing the word itself.

Full-text index: An index containing every word of every document cataloged, including stop words (defined below).

Fuzzy search: A search that will find matches even when words are only partially spelled or misspelled.

Index: The searchable catalog of documents created by search engine software. Also called "catalog." Index is often used as a synonym for search engine.

Keyword search: A search for documents containing one or more words that are specified by a user.

Phrase search: A search for documents containing a exact sentence or phrase specified by a user.

Precision: The degree in which a search engine lists documents matching a query. The more matching documents that are listed, the higher the precision. For example, if a search engine lists 80 documents found to match a query but only 20 of them contain the search words, then the precision would be 25

Proximity search: A search where users to specify that documents returned should have the words near each other.

Query-By-Example: A search where a user instructs an engine to find more documents that are similar to a particular document. Also called "find similar."

Recall: Related to precision, this is the degree in which a search engine returns all the matching documents in a collection. There may be 100 matching documents, but a search engine may only find 80 of them. It would then list these 80 and have a recall of 80

Relevancy: How well a document provides the information a user is looking for, as measured by the user.

Search Engine: The software that searches an index and returns matches. Search engine is often used synonymously with spider and index, although these are separate components that work with the engine.

Spider: The software that scans documents and adds them to an index by following links. Spider is often used as a synonym for search engine.

Stemming: The ability for a search to include the "stem" of words. For example, stemming allows a user to enter "swimming" and get back results also for the stem word "swim."

Stop words: Conjunctions, prepositions and articles and other words such as AND, TO and A that appear often in documents yet alone may contain little meaning.

Thesaurus: A list of synonyms a search engine can use to find matches for particular words if the words themselves don't appear in documents.

6.4 CGI

When you request a Web page by entering a Web site address, the server sends back the requested page. However, when a user fills out a form on a Web page and sends it in, the Web server passes the form information to a small application program that processes the data. This method or convention for passing data back and forth between the server and the application is called the Common Gateway Interface (CGI). It is part of the Web's HTTP protocol. A CGI application can be written in a number of different languages, the most popular being C++, Java, and Perl.

If you are creating a Web site and want a CGI application to get control, you specify the name of the application in the URL that you code in an HTML file. This URL can be specified as part of the FORMS tags if you are creating a form. For example, you might code:

```
<FORM METHOD=POST ACTION=http://www.mybiz.com/cgi-bin/formprog.pl>
```

and the server at "mybiz.com" would pass control to the CGI application called "formprog.pl" to record the entered data and return a confirmation message.

6.5 Web Servers

Apache is a free of cost Web server that is easy to set up. It's source code is in C. It is distributed under an "open source" license. It runs on Windows 95/98/NT, Unix systems (such as Linux, Solaris), and on other Unix/POSIX-derived systems (such as Rhapsody, BeOS, and BS2000/OSD). According to `www.netcraft.com` web server survey in September 1998, more than 50% of all Internet servers were running Apache.

Chapter 7

Internet Internals

The World-Wide Web is not the Internet! The Internet is the catch-all word used to describe the massive world-wide network of computers. The Internet is comprised of thousands of smaller regional networks scattered throughout the globe. On any given day it connects a hundred million users in over 50 countries. The World-Wide Web is mostly used on the Internet; they do not mean the same thing. The Web refers to a body of information - an abstract space of knowledge, while the Internet refers to the physical side of the global network, a giant mass of cables and computers.

Nobody owns the Internet - although there are companies that help manage different parts of the networks that tie everything together, there is no single governing body that controls what happens on the Internet. The networks within different countries are funded and managed locally according to local policies.

Having access to the Internet usually means that one has access to a number of basic services: electronic mail, interactive conferences, access to information resources, network news, and the ability to transfer files.

The World-Wide Web uses the Internet to transmit hypermedia documents between computer users internationally. Nobody owns the World-Wide Web.

7.1 Infrastructure

Your request for a page from somewhere on the World Wide Web travels through a labrynth that is generally referred to as the Internet "infrastructure." For our tour, let's define it as your particular Internet service provider and all the wires and routing to the ISP for the server where the Web page you're requesting is located.

After travelling through your phone line or other path and then through the telephone company central office (or your enterprise's proxy and firewall servers), your Web page request travels successively through: (1) Your Internet service provider (ISP) server (2) The regional network your ISP is connected to (3) If necessary, through one of the four major network access points (NAPs) in the U.S. (4) Then through the national commercial backbone (5) And then once again through the NAP, regional network, and ISP at the other end.

7.2 Addresses

The Internet has an addressing scheme that every computer on the network understands. An IP (Internet Protocol) address is a 4-byte number separated by periods that identifies a specific computer connected to the Internet. Each byte is a number which can range from 0 to 255. When you connect to an ISP, depending on how your ISP assigns IP addresses, you may have one address all the time or a different address each time you connect.

Web servers have the same kind of addresses: if you type `http://204.162.80.180/` in your browser, you'll get the same result as if you had typed `http://www.cnet.com/`.

Internet domain names are the next level of Internet addressing, just as the street name is followed by the city and state. Domain names create a single identity for a set of locally connected computers used by a company or an institution. So while there may be 38 servers at a given company, each with its own IP address, they all share a common domain name, such as `wright.edu`.

The domain name identifies all the computers in a group. But if you want to get to a specific page stored on any of those computers, you'll need an even more precise address. That's why every Web page on the Internet, and even the objects you see displayed on Web pages, has its own unique address, known as a Uniform Resource Locator (URL), which tells your browser exactly where to go on the server to find a page.

7.3 Modems

Modem stands for MOdulator/DEModulator. A modem converts digital signals generated by the computer into analog signals which can be transmitted over a telephone line and transforms incoming analog signals into their digital equivalents.

The specific techniques used to encode the digital bits into analog signals are called modulation protocols. This encoding process puts the transmission into a mode that is compatible with the various transmission media used by the telephone company, such as copper wire, microwave, satellite, and fiber optics.

Modems transmit data to other computer and their speed is measured in bits

per second (bps), as in 36.6 Kbps. Most new modems can send and receive data at 33.6 kbps and faxes at 14.4 kbps. The raw speed (the speed without data compression) of a modem is determined by the modulation protocols.

ISDN (Integrated Services Digital Network) modems use digital telephone lines to achieve a transmission speed of 128 kbps.

7.4 NICs

7.5 Networks

A LAN (Local Area Network) is a network of interconnected computers within a relatively small geographic area. Typically, this might be within the area of an office building.

The main LAN technologies are: Ethernet, Token ring, ARCNET, and FDDI (Fiber Distributed Data Interface). Typically, a suite of application programs can be kept on the LAN server. Users who need an application frequently can download it once and then run it from their local hard disk. Users can order printing and other services as needed through applications run on the LAN server. A user can share files with others at the LAN server; read and write access is maintained by a LAN administrator.

A LAN server may also be used as a Web server, if safeguards are taken to secure internal applications and data from outside access.

A WAN (wide area network) is a geographically dispersed telecommunications network. A wide area network may be privately owned or rented, but the term usually connotes the inclusion of public (shared user) networks. An intermediate form of network in terms of geography is a metropolitan area network (MAN).

7.6 Routers, Gateways, Hubs

A router is a connector between Local Area Networks that uses identical protocols; packets are received and examined and then sent on. It is a device or, in some cases, software in a computer, that determines the next network point to which a packet should be forwarded toward its destination. The router is connected to at least two networks and decides which way to send each information packet based on its current understanding of the state of the networks it is connected to. A router creates or maintains a table of the available routes and their conditions and uses this information along with distance and cost algorithms to determine the best route for a given packet. Typically, a packet may travel through a number of network points with routers before arriving at its destination.

A gateway is a computer that lies at the intersection of two networks, and routes traffic correctly between them, while keeping traffic internal to the two networks separated.

A hub is the central part of a wheel where the spokes come together, as in an airport "hub" to make connecting flights from one point to another. In data communications, a hub is a place of convergence where data arrives from one or more directions and is forwarded out in one or more other directions. A hub usually includes a switch of some kind. (And a product that is called a "switch" could usually be considered a hub as well.) The distinction is that the hub is the place where data comes together and the switch is what determines how and where data is forwarded from the place where data comes together. Regarded in its switching aspects, a hub can also include a router.

Typically, devices with network interface cards (NICs) are connected to each hub with shielded twisted pair (STP) or unshielded twisted pair (UTP) cable.

7.7 Protocols

In information technology, a protocol is the set of rules that a telecommunication connection uses when the two end points send signals back and forth. Protocols exist at several levels in a telecommunication connection. There are hardware telephone protocols. There are protocols between the end points in communicating programs within the same computer or at different locations. Both end points must recognize and observe the protocol.

7.7.1 TCP/IP

On the Internet, there are the TCP/IP protocols, consisting of:

TCP (Transmission Control Protocol), which uses a set of rules to exchange messages with other Internet points at the information packet level.

IP (Internet Protocol), which uses a set of rules to send and receive messages at the Internet address level.

HTTP, FTP, and other protocols, each with defined sets of rules to use with other Internet points relative to a defined set of capabilities.

7.7.2 HTTP

The Web is based on a set of rules for exchanging text, images, sound, video, and other multimedia files, which is collectively known as HTTP, or hypertext transfer protocol. Web pages can be exchanged over the Net because browsers (which read the pages) and Web servers (which store the pages) both understand HTTP.

7.7.3 SSL

SSL (Secure Sockets Layer) is a program layer created by Netscape for managing the security of message transmissions in a network. SSL uses the public-and-private key encryption system from RSA, which also includes the use of a digital certificate. SSL is an integral part of a browser.

7.7.4 Telnet

Telnet is the way you can access someone else's computer, assuming they have given you permission. (Such a computer is frequently called a host computer.) More technically, Telnet is a user command and an underlying TCP/IP protocol for accessing remote computers. The Web or HTTP protocol and the FTP protocol allow you to request specific files from remote computers, but not to actually be logged on as a user of that computer. With Telnet, you log on as a regular user with whatever privileges you may have been granted to the specific applications and data on that computer.

A Telnet command request looks like this:

```
telnet paladin.wright.edu
```

The result of this request would be an invitation to log on with a userid and a prompt for a password. If accepted, you would be logged on like any user who used this computer every day. Telnet is most likely to be used by program developers and anyone who has a need to use specific applications or data located at a particular host computer.

7.7.5 FTP File Transfer Protocol

FTP is the simplest way to exchange files between computers on the Internet. FTP is an application protocol that uses the Internet's TCP/IP protocols. FTP is commonly used to transfer Web page files from their creator to the computer that acts as their server for everyone on the Internet. It's also commonly used to download programs and other files to your computer from other servers.

As a user, you can use FTP with a simple command line interface or with a commercial program that offers a graphical user interface. Your Web browser can also make FTP requests to download programs you select from a Web page. Using FTP, you can also update (delete, rename, move, and copy) files at a server. You need to log on to an FTP server. However, publicly available files are easily accessed using anonymous FTP.

7.7.6 SMTP Simple Mail Transfer Protocol

Sends your e-mail messages from your computer to an e-mail server.

7.8 Internet2

Internet2 is a U.S. governement project to create a powerful leading edge network for universities and the national research community. Internet2 is not a physical network that will replace the Internet. Rather, Internet2's goal is to develop new technologies that can be deployed in the Internet. Internet2 is working to enable applications, such as telemedicine, digital libraries and virtual laboratories that are not possible with the technology underlying today's Internet.

Read more about it at `http://www.internet2.edu/`.