Yukon College
CPSC100 Practical Computer Fluency
Fall 2009 Term Notes

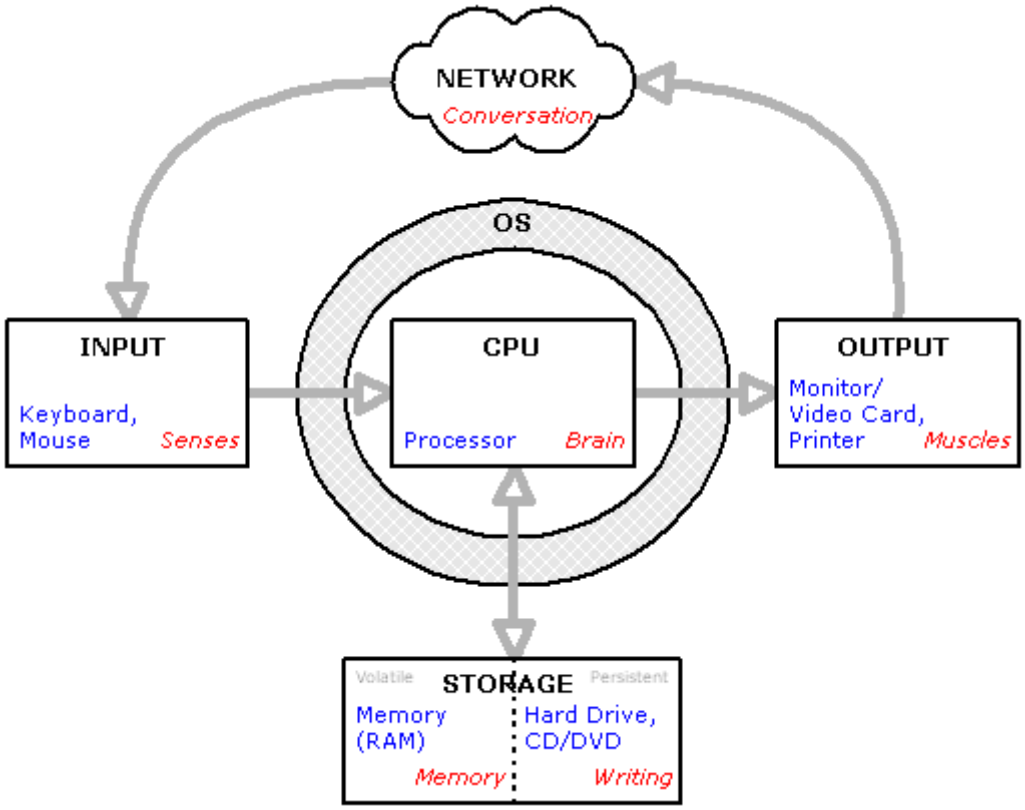
Contents

Computer Components 1
CPU, Boot, BIOS, and OS 3
CPU History and Comparison 6
Computer Data Storage 7
Data, Drives, Folders, and Files 9
File and Folder Tasks 13
Lab Account Notes 15
Editing Tips and Techniques 16
The Internet 21
Internet Security and Privacy 26
An HTML Primer 35
Creating an HTML File from Scratch 40
Computer Graphics and Multimedia 43
Computer Image Formats 46
Managing Computer Images 48
HTML Images and Colour 51
Introduction to Microsoft Word 53
A Field Guide to Word's Style Screens 56
Word Objects 58
Heading and Page Numbering in Word 60
Spreadsheets 62
Handout: Common Excel Functions 65
Spreadsheet Tips 67
Handout: Charts 69
The Can't-Fail Guide to Creating Excel Charts 71
Mail Merge 72
Exercise: Working with Files and Folders 74
Exercise: HTML File Practice 75
Exercise: Formatting with Styles 76
Exercise: Combining Text, Images, Tables, and Charts in Word 78
Exercise: Outline and Page Numbering 80
Exercise: Spreadsheet Formulas 82
Exercise: Spreadsheet Functions 83
Exercise: Date Calculations 84
Exercise: Charting 85
Exercise: Mail Merge 86

Computer Components

After our brief computer shopping excursion, we realized that there are a number of components making up the average computer. Some of these weren't all that familiar: Processor, Operating System (OS), Memory, Network, Hard Drive, CD/DVD, Monitor/Video Card, Printer, Keyboard, Mouse.

To help us make sense of these, we first organized them into four categories: Input, Output, Central Processing Unit (CPU), and Storage (Network and Operating System too, but they're treated a bit differently):



Information flows in the direction of the arrows. Sample devices or terms are shown in blue while their human analogy is shown in italicized red.

Input devices provide data from the outside world to the processor: keyboards, mice, joysticks, microphones, digital cameras, scanners, and so forth, including various sensors for embedded computers that measure temperature, speed, throttle position, or whatever.

Output devices allow the processor to communicate the results of its work to the outside world: monitors, printers, speakers, LED and LCD displays, and all sorts of different servos and actuators for embedded computers that change valve positions, angle rudders, pump the brakes--the list is endless.

Storage devices are used by the processor to temporarily or permanently store data so that it can be retrieved at a later time. "Volatile" storage loses its contents when the computer power is turned off (the human analogy would be your own memories). The most common example of this type of storage is Random Access Memory, or RAM, or just "memory". "Persistent", or "non-volatile", storage does not lose its contents when the power is turned off: hard disks, floppy disks, CD-ROMs, DVDs, and so forth (the human analogy would be books, cave drawings, and oral histories passed from generation to generation). We'll study the storage component in greater detail in the next class.

The **Central Processing Unit**, or just "processor" for short, is the engine or brain of a computer. The processor executes a series of instructions that gather data from the input devices, occasionally store intermediate results using the storage devices, and then produce final results suitable for the output devices.

The diagram above also shows a **Network** of some sort (perhaps the Internet, or maybe a cell phone network). The computer's interface with a network involves both input and output devices, although these may be physically incorporated into a single piece of hardware (e.g. a network interface card, or NIC). This is because the computer both sends messages out to a destination computer somewhere on the network, and also listens for messages that arrive as input to the computer.

The **Operating System (OS)** is also shown as an intermediary between the CPU and its attached input, output, and storage devices. The responsibilities of the operating system will be described later. As far as human analogies are concerned, while it's a bit of a stretch, you could say the operating system is like the autonomic brain functions that handle breathing, beating the heart, and all of the muscle coordination needed to catch a ball--in other words, things that we don't consciously think about.

All of the arrows on the diagram are typically realized by cables of some sort, or they may be wireless (radio or infrared) signals, or they may be wires built in to the computer's main circuit board.

If you're not satisfied with the human analogies, you could try a vehicle: the input devices are the steering wheel, gear shift, and pedals; the processor is the engine; the output device is the drivetrain; the storage is the fuel; and the network is the road.

CPU, Boot, BIOS, and OS

The Central Processing Unit (CPU)

The CPU is the brains of the computer. It's only about one inch square in size, but is encased in a large plastic enclosure with many electrical pin connectors sticking out. Recent CPUs also have large heat sinks and sometimes even little fans to keep them cool.

CPUs are purely digital devices. This means that they only understand discrete numerical values. In fact, the only numerical values a CPU understands is zero and one: a numbering system called "binary". Ultimately, everything we do with a computer must be translated into those binary values.

The instructions that CPUs understand are very simple: fetch a value from memory, perform simple calculations, store a value in memory, and compare two values and jump to another instruction depending upon the outcome. Other than these jumps, the CPU simply executes each instruction in turn, one by one.

The CPU is hooked to a crystal clock that ticks very quickly. Each time the clock ticks, the CPU performs one of these very simple operations (often it takes multiple ticks to perform one operation, but never less than one).

These ticks are produced at a certain frequency. Apple][+ era computers (The Mostek 6502 CPU) had a frequency of about 1.5MHz (1.5 million ticks per second). *If you look very closely at the internal video display used by the Terminator in the first movie, it appears that Cyberdyne Systems Model 101 is built using a Mostek 6502 CPU.*

The most recent CPU chips from Intel run at a frequency of between 3 and 4GHz (i.e. more than 2,000 times that of the first Apples). But frequency alone shouldn't be used as criteria for comparing the speed of CPUs, particularly if the chips are from different product lines or from different manufacturers.

In 1964, Gordon Moore (who co-founded Intel in '68) announced that the number of transistors on computer chips appeared to be doubling every 18 months. The trend has continued to this very day. It also loosely translates to a doubling of CPU speed every 18 months.

Computer Software

Software is that part of the computer that you cannot kick.

You may think of software as the recipe that a CPU must follow in order to perform a task.

Computers are very literal: they will only do exactly what you tell them to do--even if you really meant to tell them to do something else (this is what we call a bug).

Computers are also deterministic: given the same input, they will always produce the same output. This is a good thing, because otherwise it would be impossible to program computers to do the same thing twice (although quantum theory heralds the day when computers don't behave deterministically).

A program is just a list of instructions to the CPU, that when executed, turn into what we think of as software. A program implements an algorithm (a mathematical formula or set of instructions) in the "machine language" of the CPU (lots and lots of ones and zeros).

The Boot Process

Before any computer can perform useful work, it must first initialize itself using a process known as bootstrapping, or simply "booting". When power is first applied to a computer, it is in a random state: all of the volatile storage devices are filled with meaningless noise. To move from this chaotic state to an ordered state involves a bit of effort (entropy must be reversed). Most PCs follow a similar set of steps to bootstrap themselves, as outlined in the following points. Compare this to your own experience of first waking in a hotel room while on vacation: you need to do

some work to: remember where you are and why, get up, get dressed, shower, and so forth. You can't really perform useful work until all of that has been done.

A special chip in the computer holds what is called the BIOS (Basic Input/Output System), essentially just the instructions needed by the PC to get things started. The first thing the BIOS does is execute the POST (Power On Self Test) which checks the video card (so that any errors can be displayed to the user--if the video card doesn't work, then the BIOS beeps a few times to indicate the problem).

The BIOS then checks that the major input devices (e.g. keyboard, mouse) are connected and functional.

The next step is to verify that the RAM (volatile storage) in the computer works. The BIOS writes to selected locations in memory and then reads the same data back to make sure that the memory chips are functioning properly.

The BIOS also checks the main storage devices connected to the system, including floppy drive(s), CD-ROM drive(s), and hard drive(s).

The BIOS then starts looking for an operating system to load. It usually checks the floppy first, then the CD-ROM, and finally the hard drives. As soon as it finds something promising, it loads the first "chunk" of data from the storage device into memory and tells the processor to execute the instructions found in that chunk. The BIOS's work is now done.

The first chunk begins to execute. It should contain just enough instructions to load the next chunk stored on the disk into memory and execute those instructions. There may be many such steps, but the end result is to load the operating system into memory and start it running.

The operating system (OS) first initializes itself, and then loads any special device drivers (little programs that allow the OS to communicate with specific types of hardware).

The computer is now ready to perform useful work.

The Operating System (OS)

When sold as shrink-wrapped products, we know operating systems by such names as Microsoft Windows, Apple Macintosh, UNIX (and its variants Linux, Solaris, AIX, HP-UX, etc.), OS/2, and even ones you may never have heard of: OS/390, OS/400, Be, CP/M, VMS, TOPS, MVS, ITS, etc.

But those products include a whole lot more than just the operating system program: web browsers, disk utilities, graphical user interface shells, file managers, calendars, e-mail programs, and so forth.

The operating system is a program that is always running on the computer, and it has two primary responsibilities:

1. Manage hardware and software resources so that programs may use these resources without conflict.
2. Provide a consistent and hardware-independent programming interface for software development.

The first of these responsibilities is most important for operating systems that can run more than one program at a time (almost any OS with which you might be familiar). These programs are completely unaware that other programs are also running. Suppose that two programs that are currently running each wish to display something on the screen, save a file to the hard disk, and play a sound on the speakers. If they both tried those simultaneously, they would likely overwrite each other's efforts: producing a garbled screen display, a corrupted file containing interspersed bits written by both programs, and some random noise from the speakers. The operating system's job in such a scenario is to isolate each program from the hardware devices by playing a "traffic cop" role. The programs are still allowed to use the hardware, but they always do so through the intermediating OS, which prevents the programs from interfering with each other (perhaps by

isolating the two programs' display into separate windows, by placing the saved files in two different locations, and by letting only one program at a time use the speakers). The operating system does this by essentially restricting each running program to its own "virtual computer" (the term virtual means "fake" or "pretend" in the computer world). If a program ever tries to break out of this virtual environment (usually because of a programming defect, or "bug"), the operating system can detect this condition and halt the offending program. You may get an error message (perhaps something called a "General Protection Fault"), and you will lose any unsaved data in that application, but the OS should be able to restrict the damage to only that virtual computer; all of the other programs should continue running normally.

Because the OS must be an intermediary between programs and the computer hardware, this leads directly to the second responsibility: The OS hides the specifics of any particular piece of hardware, proving only a standard way of accessing hardware of that nature. This means that application programs (word processors, spreadsheets, games, etc.) do not have to be written for specific hardware, only for specific operating systems. In the days of yore, this was not the case, and WordPerfect (the champion word processor of its day) had to include hundreds of printer definition files for every conceivable printer make and model on the market. Today, the operating system loads just one such device driver for a particular printer (or any other device) and all programs simply ask the operating system to print something, none the wiser about the model of printer connected to the computer.

How does a single computer run many programs simultaneously? The operating system manages this by letting each program in turn use the processor for a short period of time (microseconds or nanoseconds). The switch is so fast that we don't notice that only one program is really running at any one time. CPUs with two or more built-in processors are now on the market, permitting more than one program to actually run simultaneously, but the OS still manages the scheduling.

CPU History and Comparison

The list below is far from complete, but it certainly illustrates an exponential growth in computing power over the past twenty years or so.

Manufacturer	**CPU Name**	**Date**	**Transistors**	**Trace Width (microns)**	***Clock Speed (MHz)***
Intel	8080	1974	6,000	6.0	*2*
Intel	8088	1979	29,000	3.0	*5*
Intel	80286	1982	134,000	1.5	*6*
Intel	80386	1985	275,000	1.5	*16*
Intel	80486	1989	1.2 million	1.0	*25*
Intel	Pentium	1993	3.1 million	0.8	*60*
Cyrix	6x86/M1	1995	?	?	*200*
AMD	K5	1996	4.3 million	0.35	*166*
Intel	Pentium II	1997	7.5 million	0.35	*233*
AMD	K6	1997	9.3 million	0.25	*300*
Intel	Celeron	1998	7.5 million	0.25	*300*
IBM & Motorola	PowerPC 750	1998	40 million	0.22	*400*
Intel	Pentium III	1999	9.5 million	0.25	*450*
AMD	Athlon	1999	67 million	0.13	*1,400*
Intel	Pentium 4	2000	42 million	0.13	*2,000*
IBM & Motorola	PowerPC 970	2003	52 million	0.13	*1,800*
AMD	Opteron	2003	106 million	0.13	*1,500*
Intel	Pentium 4 HT	2004	178 million	0.09	*3,500*
AMD	Athlon 64 X2	2004	233 million	0.09	*2,400*
Intel	Pentium D	2005	169 million	0.09	*3,000*
Intel	Xeon MP	2005	286 million	0.09	*3,600*
Intel	Core 2 Quad	2006	582 million	0.065	*3,300*
Intel	*Quad-Core Itanium*	*2008*	*2 billion*	*0.065*	*?*

"Trace width" is the width of the circuit "wires" on the chip. A micron is a millionth of a metre. A human hair is about 50 microns wide. The clock speeds quoted are for the most advanced version of each CPU series, except for the most recent chips which are continually increasing in speed.

Computer Data Storage

Units of Computer Storage

Unit	Abbr.	Bytes	Notes
Bit	b	$^1/_8$	The smallest piece of information possible: On/Off, True/False, Yes/No.
Byte	B	1	A byte is eight bits (a nybble is four bits). A single byte can store a number between 0 and 255 (2^8-1). A lowercase "b" usually means bit, while an uppercase "B" usually means byte.
Kilobyte	KB	1,024	Most files on your computer, or downloads from the Internet, are measured in KB. This page is about 8 KB in HTML format.
Megabyte	MB	1,048,576	Large computer files, especially images, music, or video, are often in this range. 1 MB = 1,024 KB.
Gigabyte	GB	1,073,741,824	Your computer's volatile memory and most hard disk capacities are measured in GB these days, but hard disk manufacturers measure their disk sizes in SI units rather than powers of 2 ("K" to them means 1,000 instead of 1,024). This means a 80GB hard disk is really only 74.5GB in size when comparing against other types of storage devices. Sneaky!
Terabyte	TB	1,099,511,627,776	You can now buy a 1TB hard disk for under $200. 83 of these, and you could store the entire Library of Congress.

And if you keep going, multiplying by 1,024 at each step, you'll reach Petabytes, Exabytes, Zettabytes, and even Yottabytes (1,208,925,819,614,629,174,706,176 Bytes). Downloading a Yottabyte-sized file over a 1Mbps connection would take 292 billion years.

Types of Computer Storage

As you move down the list, access times slow down, but the cost per unit of storage also decreases. Volatile storage types lose their data when the computer's power is turned off.

Type	Typical Size	Volatile?	Use
CPU Registers & Cache	1B - 1MB	Yes	Temporary and extremely fast storage of program data.
Random Access Memory (RAM)	256MB - 4GB	Yes	Data used by currently running programs. Often referred to as "main memory".
Virtual Memory	256MB - 4GB	Yes	RAM data that is temporarily moved to a hard disk to free up memory space. Virtual memory is thousands of times more slow to access than RAM, because it must first be reloaded from the hard disk when accessed.
Fixed Media	80GB - 1TB	No	Almost always consists of one or more hard disks installed in the computer.
Network Storage	*variable*	No	Hard disk storage on a remote computer that, by way of a connected network, appears to the user as a local hard disk.
Removable Media	variable	No	Usually a convenient format for transporting or archiving data.

Common Storage Media

Media	Typical Capacity	Notes
3-1/2" Floppy Disk	1.44MB	Handy for copying small files from machine to machine (pretty rare now).
Keychain, Flash, or USB Drive	128MB - 8GB	Solid-state (no moving parts) persistent memory devices that are very handy for copying files between computers. The same sort of memory is used in digital cameras.
Compact Disc (CD)	650MB - 700MB	CD-ROM media can only be read. CD-R (CD Recordable) media and drives allow one-time, or WORM (Write Once Read Many), recording (a.k.a. "burning"). CD-RW (CD Rewritable) media can be written, erased, and read repeatedly. CD-RW disks may not be readable in all players.
Digital Versatile Disc (DVD)	4.7GB - 8.5GB	Like CDs, DVDs also come in read-only, write-once, and read-write formats.
Hard Disk	80GB - 1TB	Hard disks (which spin inside Hard Drives) are the fastest, cheapest, and highest-capacity storage media available. It was only a couple of decades ago that these things were the size of washing machines. Now a 160GB hard disk fits inside Apple's iPod.

Data, Drives, Folders, and Files

Computer Data and Information

What is computer data? (Technically, "what are computer data?", since the word "data" is plural.) Think of it as the raw form of information. Data are really only understood by computers. It's not until we run a program, that the data can be interpreted as "information"--something useful to we humans.

Ultimately data are stored and processed as numbers; binary numbers to be precise.

Human-friendly information could be anything: your rÃ©sumÃ©, music, e-mail messages, a web page, etc.

We need a "memory" to be able to store this data for ourselves to use again at a later date, and also to share with others. In the distant past, we've used what's in our noggins to store this information, and told stories when we wanted to share it. In the more recent past, we've tended to use engraved stone, papyrus, and paper to accomplish the same thing.

Unsurprisingly, computers need the same capability. They use both volatile storage for quick access to data while the computer is running (volatile storage is lost when the power is turned off). They also use persistent, or "non-volatile", storage for long-term archiving of data. Persistent data can be retrieved at a later date, much as can a story written on paper, that might have otherwise been forgotten if left in our non-persistent (or "volatile") memories.

An Aside: Binary Numbers

Keep in mind that the computers we use base everything on numbers, even music and images are stored as numbers. Specifically, computers store everything using binary numbers (or base 2).

We're more used to base 10 numbers for our day-to-day lives, but we occasionally use other bases: base 60 for seconds or minutes, base 12 or 24 for measuring hours, and base 12 again for measuring distances in carpentry (1 foot = 12 inches).

Although there's no need in this course to worry too much about base 2 (binary) numbers and math, perhaps a quick example will demonstrate how they relate to our usual base 10, or decimal, numbers:

Base 10, Decimal:	$183_{10} = 1 \times 10^2 + 8 \times 10^1 + 3 \times 10^0$ $183_{10} = 1 \times 100 + 8 \times 10 + 3 \times 1$ $183_{10} = 100 + 80 + 3$ $183_{10} = 183$
Base 2, Binary:	$10110111_2 = 1 \times 2^7 + 0 \times 2^6 + 1 \times 2^5 + 1 \times 2^4 +$ $0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 1 \times 2^0$ $10110111_2 = 1 \times 128 + 0 \times 64 + 1 \times 32 + 1 \times 16 +$ $0 \times 8 + 1 \times 4 + 1 \times 2 + 1 \times 1$ $10110111_2 = 128 + 0 + 32 + 16 + 0 + 4 + 2 + 1$ $10110111_2 = 183$

In the world of computers, many things are measured in natural base 2 numbers. For example, this is why you see computer memory sold in blocks of 64MB, 128MB, 256MB, 512MB and so forth. Also why the current crop of processors are called 32-bit chips.

Persistent Storage

Understanding persistent storage on computers is important because, unless we keep our computers running 24x7 with every single document we've ever worked with still onscreen, we're going to need to store information from time to time, and expect it to retrieve it at a later date.

In Windows, to store a document, you "Save" it, and to retrieve a document later, you "Open" it.

But let's also ask ourselves why operating systems make us think this way. When you're working with paper documents there's no real concept of saving and opening them. That's all purely a computer-related concept. Wouldn't it be better if we didn't have to worry about such things?

But we're stuck with the way things work in today's operating systems, and that means we have to learn a bit how persistent storage media works, and how it is organized.

Of all the persistent storage media out there (listed on one of this week's handouts), hard disks are the one we deal with most frequently.

Drives and Partitions

In Windows, when you open "My Computer", you'll see a list of all of the "[disk] drives" that the OS knows about, helpfully named using a letter of the alphabet (e.g. C:, D:, E:). Windows actually calls these "logical drives". In computerese, "logical" means pretty much the same thing as "pretend". So a logical drive may not correspond exactly to a real piece of disk hardware, but we can pretend it does.

Each such logical drive may correspond to a physical hard disk, or it may in fact correspond to just part of a disk (a "partition") or even a hard disk that's not actually installed on your PC but is located somewhere on the network.

It used to be that organizing partitions was a big part of setting up a computer, but these days we can safely ignore the concept most of the time.

But do note that once a partition is created (perhaps by the company that sold you a computer), it is difficult (if not impossible) to change its size without losing all of the data on the partition, and perhaps on other partitions as well. There are tools that will let you dynamically change the size of the partitions without losing data, but these are sold separately from the operating system.

From now on, we'll just focus on the things Windows calls logical drives.

Filesystems

A filesystem is a low-level organization of a logical drive's recording surface. Only one filesystem may be installed on a logical drive. "Formatting" is the process used to install a filesystem on a logical drive. Formatting also "erases" all previous data on the partition, so be careful. *While formatting does make it difficult to recover data, it's not impossible, so before selling or giving away a computer, either remove the hard drive (and smash it), or use a special "nuke" program (see the link in this week's readings).*

Filesystems allow us to organize our data into files and folders. (Folders are commonly called "directories" on older versions of Windows and currently on UNIX-style operating systems. We'll use the terms interchangeably.)

Filesystems also provide other attributes that apply to files and folders: read-only, modification date, hidden files, whether the file is ready to be archived, etc.

Some filesystems give us additional features like security, encryption, compression.

Here are the more common filesystems used by various versions of Microsoft Windows:

Filesystem	Windows Version	Notes
FAT16	all versions, including MS-DOS	FAT16 (File Allocation Table) originated in MS-DOS and was famous for the limit on file and folder names of just 8 characters for the name, and 3 for the extension (known as 8.3). FAT16 can manage a maximum of 65,536 allocation units (pieces of hard disk space, or "blocks") and had a maximum logical drive size of 2GB.
VFAT	95, 98, ME, NT 4, 2000, XP	VFAT was a modification to FAT16 that allowed for long (256 character) file and folder names. It has the same limitations on allocation units and disk size as FAT16.
FAT32	95, 98, ME	FAT32 arrived in later versions of Windows 95 and allowed for both long file and folder names and logical drive sizes of up to 2TB in size.
NTFS 1.0	NT	NTFS 1.0 debuted in Windows NT and permitted both long file and folder names and logical drive sizes of up to 2TB in size. NTFS 1.0 also added access control to files and folders, and compression of files or whole folders.
NTFS 3.0	2000, XP, Vista	NTFS 3.0 enhanced the 1.0 version by adding encryption for files or whole folders.
WinFS	*Future?*	Not really a true filesystem, WinFS is supposed to allow Windows users to organize their folders and files in new and exciting ways (if you get excited about such things). It was originally going to be part of the next version of Windows (called "Vista") but seems to be on indefinite hold.

Other operating systems use different filesystems, but with similar features and limitations.

Filesystems store files by dividing the logical drive into a large number of equally-sized "blocks". The size of these blocks can vary dramatically from drive to drive, from 512B up to 64KB. Each file must occupy an integer number of blocks, and blocks can't be shared between files. Small files may only take up one block (and leave a lot of free space that can't be used for anything else), while larger files can be split across many blocks (usually with space left over in the last block). Organizing files into blocks has a couple of important ramifications for us:

- All of the Windows filesystems are optimized for fast writing of data. This means that as soon as the system finds an empty block on the logical drive, it starts writing. Over time, this behaviour means that the files become "fragmented". Windows does come with defragmenting tools to clean up this mess (and speed up file access) and third party vendors also sell more feature-full tools. To defrag a logical drive in Windows, right-click on the drive icon in My Computer, choose Properties from the menu, choose the Tools tab, and then click the Defragment Now button. You can now select any logical drive and click the Defragment button to start the process.
- When you delete a file (and don't just copy it to the Recycle Bin), the filesystem doesn't actually erase anything on the disk; it merely designates the blocks that formerly contained that file as empty. Until a new file is stored on the disk, the file's data is not truly gone. Special "undelete" utilities are available to recover these files, but use them quickly before new files overwrite those blocks and you lose the file forever.

Folders

A folder is a "logical container of files and other folders." It's a "logical" container because the physical disk structure of a folder has nothing whatsoever to do with the files and folders that are inside. As mentioned before, folders and directories are the same thing.

Folders are used to organize the files on a computer into a human-understandable hierarchy.

Folders form a tree structure (actually an upside-down tree), starting at the "root" folder.

To navigate the folder, we start at the root and work our way down from folder to folder until we reach the file in which we're interested. When writing out the location of a file, we write out the full "path" from the root folder to the file.

Depending on the operating system, paths are written using different characters to separate folders. For Windows, we use the backslash (\), on UNIX and the Internet a slash (/), and on Macintosh a colon (:). So, in Windows, for a file located on the D: logical drive, the full path might be: D:\aaa\bbb\ccc\file.txt. Where aaa, bbb, and ccc are folder names. The first backslash in the path (after "D:") is the root folder.

Because strict hierarchal organization of folders and files is sometimes a bit too constraining, certain filesystems allow you to circumvent the hierarchy to some extent by using "pointers" called shortcuts or links. In Windows, shortcuts are used to simulate copying a single file to many different directories, but in reality there's just the one file. Changes made to any of the shortcuts, or the original file, will be visible if any of the other shortcuts is opened in an editor later (since there's only just the one file that was changed).

Files

A file is a container of "organized data understandable to a computer program." Note that, to a human, a file is just a string of unintelligible 1s and 0s. And a file may not be understandable to all computer programs, just those programs that understand the format of the data. For instance, you can open a music file in Windows Notepad, but see nothing but gobbledygook because Notepad only knows how to read text files.

Windows files typically have a "file extension": the three or so letters after a dot at the end of the file name. The file extension is a hint to the OS about how to interpret the data in the file. "`.exe`" usually means that the file is a program. "`.txt`" usually means that the file contains simple text without fancy formatting. "`.xls`" usually means that the file is an Excel workbook. But again, these extensions are just suggestions and there's nothing to prevent you from naming a Word document "`My Resume.exe`" to make it look like it's a program (of course, it won't run like a program).

File and Folder Tasks

There are many, many different ways to accomplish certain tasks involving files and folders. This cheatsheet lists only the most common techniques. Experiment to find the way that works best for you.

All of these techniques require using the Windows Explorer program, which you open by clicking My Computer or My Documents on the desktop.

If you want to...	Try doing this...	Comments
View all logical drives.	Double-click the My Computer icon on the desktop.	On a home Windows XP computer, the My Computer icon may also appear on the Start menu near the top-right.
View your own files and folders.	Double-click the My Documents icon on the desktop.	On a home Windows XP computer, the My Documents icon may also appear on the Start menu near the top-right.
View the properties for a logical drive, folder, or file.	Right-click on the drive, folder, or file icon and choose Properties from the pop-up menu. You may then need to chose the General tab.	The logical drive property page shows the filesystem and a pie chart of the used and free space. The folder property page shows the full path to the folder, the size of its contents, and the number of contained folders and files. The file property page shows the full path to the file, its size (the real size as well as the size required by the filesystem's blocks to store it), and the date and time it was created, last modified, and last accessed. Both the folder and file property pages also let you change some attributes: read-only and hidden, for example.
Change the way files and folders are displayed.	With a folder open, choose one of the following from the View menu: Filmstrip, Thumbnails, Tiles, Icons, List, or Details (not all choices may appear for all folders).	Filmstrip and Thumbnails are best for folders containing images. Tiles displays each file and folder with extra information. Icons and List are just two different sizes of icon. Details formats the contents of the folder as a table and usually includes the file size, type, and last modified date.
View the hierarchy of folders within a logical drive.	From the View menu, choose Explorer Bar, and then choose Folders.	The hierarchy will be shown on the left-hand size of the window.
Create a new folder.	From the File menu, choose New, and then choose Folder. Right-click in an empty space in the window and choose New and then Folder from the pop-up menus.	Once the new folder is created, it will be selected so that you can change its name.
Create a new file.	From the File menu, choose New, and then choose the type of new file you would like to create. Right-click in an empty space in the window and choose New and then the type of file you would like to create.	The list of file types depends on the programs installed on the computer. Once the new file is created, it will be selected so that you can change its name.
Rename a folder or file.	Right-click the folder or file and choose Rename from the pop-up menu. Select the folder or file and press F2.	If you try to change a file's extension, you will be asked to confirm your intentions.

Copy a folder or file.	Right-click the folder or file and choose Copy from the pop-up menu. Right-click the destination folder and choose Paste from the pop-up menu. Select the folder or file and press Ctrl+C. Select the destination folder and press Ctrl+V. Right-click the folder or file and drag it to the destination folder. Choose Copy Here from the pop-up menu that appears.	Copying a folder also copies all of its contents. If you change you mind following the copy, either chose Undo from the Edit menu or press Ctrl+Z to remove the copied version. You will first be asked to confirm that you want to delete the copied version.
Move a folder or file.	Left-click the folder or file and drag it to the destination folder. *If the destination folder is on a different logical drive, this will copy the folder or file rather than move it.* Right-click the folder or file and choose Cut from the pop-up menu. Right-click the destination folder and choose Paste from the pop-up menu. Select the folder or file and press Ctrl+X. Select the destination folder and press Ctrl+V. Right-click the folder or file and drag it to the destination folder. Choose Move Here from the pop-up menu that appears.	Moving a folder also moves all of its contents. Note the distinction when dragging a folder or file to a different logical drive: it is copied rather than moved. If you change you mind following the move, either chose Undo from the Edit menu or press Ctrl+Z to return the moved item to its original location.
Create a shortcut to a folder or file.	Right-click the folder or file and choose Copy from the pop-up menu. Select the destination folder for the shortcut and choose Paste Shortcut from the View menu. Right-click the folder or file and drag it to the destination folder for the shortcut. Choose Create Shortcuts Here from the pop-up menu that appears.	A shortcut is a link to a folder or file, and double-clicking it produces the same result as if you had double-clicked the original folder or file. You can rename a shortcut to anything you like. Shortcuts are identified by a little curved arrow at the bottom-left of the icon.
Delete a folder or file using the Recycle Bin.	Right-click the folder or file and choose Delete from the pop-up menu. Select the folder or file and press the Delete key.	Deleting a folder also deletes all of its contents.
Recover a deleted folder or file from the Recycle Bin.	Double-click the Recycle Bin icon. Select the folder or file to recover and then choose Restore from the File menu.	Older items will eventually disappear from the Recycle Bin, so don't wait too long before recovering.
Delete a folder or file without using the Recycle Bin.	Select the folder or file, hold the Shift key down, and press the Delete key. Keep holding the Shift key down until the confirmation box is shown.	While it's not easy to recover files deleted in this manner, it is possible with special undelete programs, so don't rely on this method for security reasons.

Lab Account Notes

To accommodate all of the courses and students at the College, the computers in the labs are configured a little differently from what you might be used to at home or at the office.

Windows Explorer versus Internet Explorer

The term "explorer" is overused by Microsoft. The program called **Windows Explorer** allows you to navigate through folders and files stored on the lab computer and over the network. Windows Explorer rarely identifies itself using that name though; it's more commonly known as "that program that shows me the files in My Documents or My Computer." To open Windows Explorer, double-click either the My Documents or My Computer icons on the desktop.

Internet Explorer, on the other hand, is a web browser program intended for surfing the Internet and visiting websites. The lab computers also have the Mozilla browser installed if you prefer that one.

In class, we will use Windows Explorer to deal with files and folders, but you can choose any web browser you like for surfing.

Your Files

The lab network of computers is configured so that as you move from machine to machine, your files and personal preferences follow along. But there are a couple of rules to keep in mind for this to work smoothly.

Files that DO follow you

Any folders or files that you store in the **My Documents** folder will follow you to any lab computer in the College. You can access this folder by double-clicking the My Documents icon on your Windows desktop. This will open Windows Explorer and display the contents of the My Documents folder.

It turns out that the My Documents folder is just another name for the My Documents folder under the **`P:`** drive on the lab computers (`P:\My Documents`).

The two are just different names for the same folder, so both will contain the exact same folders and files at all times. If you make changes to one, you may need to press the refresh (F5) button to see the results through the other.

Any icons or files you place on your desktop will also follow you from lab computer to computer.

Files that DON'T follow you

Any folders or files that you store on the **`C:`** or **`D:`** drives will **NOT** be available if you move computers, and may not even be available if you return to the same computer at a later time. Therefore, do not save important files that you want to keep to either the C: or D: drives.

Space on the P: Drive/My Documents

Each student lab account is allotted a limited amount of space for file storage. Be careful that you don't reach that limit or you won't be able to store any more files. Any e-mail messages that you receive will also count towards that limit.

CPSC 100 Course Folder

Course files that we'll use for lab exercises and assignments can be found in the **`S:\cpsc100`** folder (You'll find the `S:\` drive if you first open My Computer). You cannot save your own files in that folder, or change the ones that are already there. Instead, whenever you need to work with a file in the course folder, first copy it to your My Documents folder.

Editing Tips and Techniques

The tips and techniques described on this page are widely used in Windows (and usually apply to other graphical operating systems as well). However, despite these conventions, you will run across the occasional program that behaves differently. In those cases, you may have to resort to looking through the program's menu items until you recognize a similar feature, perhaps under a different name.

Moving Around in Documents

While editing documents, you'll often want to move the position of the cursor so that you can change different parts of the document. (Technically, the flashing vertical bar in an open document is called a *caret* and the arrow--or another pointing shape--that moves about the screen as you move the mouse is called a *cursor*. The caret is also known as the "insertion point" or the "I-beam" because of its shape. I'll use the terms interchangeably unless there's a reason to distinguish between the two.) Aside from simply using the mouse to change the position of the caret, there are many keyboard shortcuts that are often more convenient:

Keystroke	Meaning
arrow keys	As expected, the keyboard arrow keys move the caret one character to the left or right, or one line up or down.
Home	Move the caret to the beginning of the current line.
End	Move the caret to the end of the current line.
PageUp	Move the caret one screen (not necessarily one page) upwards.
PageDown	Move the caret one screen (not necessarily one page) downwards.
Ctrl+<left arrow>	Move the caret one word to the the left. A "word" in this context generally means any series of non-whitespace characters.
Ctrl+<right arrow>	Move the caret one word to the the right. A "word" in this context generally means any series of non-whitespace characters.
Ctrl+Home	Move the caret to the first line of the document.
Ctrl+End	Move the caret to the last line of the document.

Selecting Text

From time to time you will want to select portions of a document to work with. The simplest way to do this is often to click the left button of the mouse where you wish to begin the selection and, while holding the button down, drag the mouse to the end of the selection and then release the mouse button (I often refer to this action as "swiping"). The selected portion will be displayed in inverse.

Although this section is directed towards selecting text, the same techniques often apply for different types of documents (e.g. cells in a spreadsheet, picture elements in an image, records in a database).

Mouse Selection Techniques

Mouse Action	Meaning
double-left-click	Select the entire word under the cursor. A "word" in this context generally means any series of non-whitespace characters.
triple-left-click	Select the entire paragraph under the cursor, or the entire contents of a text-entry box. A "paragraph" in this context generally means all of the text between two carriage returns. Not all programs or controls understand triple-left-click.
single-right-click	Doesn't perform any selection on its own, but does display a pop-up menu containing commands appropriate for the selection under the cursor.

Keyboard Selection Techniques

Keystroke	Meaning
Shift+*arrow keys*	Select the text from the caret to one character to the left or right, or one line up or down.
Shift+Home	Select the text from the caret to the beginning of the line.
Shift+End	Select the text from the caret to the end of the line.
Shift+PageUp	Select the text from the caret to the line one screen (not necessarily one page) upwards.
Shift+PageDown	Select the text from the caret to the line one screen (not necessarily one page) downwards.
Shift+Ctrl+<left arrow>	Select the text from the caret to the beginning of the word to the left. A "word" in this context generally means any series of non-whitespace characters.
Shift+Ctrl+<right arrow>	Select the text from the caret to the end of the word to the right. A "word" in this context generally means any series of non-whitespace characters.
Shift+Ctrl+Home	Select the text from the caret to the first line of the document.
Shift+Ctrl+End	Select the text from the caret to the last line of the document.
Ctrl+A	Select the entire document.
F2	This key will often let you rename the selected item, for example: desktop icons, folders and files in Windows Explorer, and the contents of a cell in a spreadsheet.

Find & Replace

Find & Replace (a.k.a. Search & Replace) is an oft-neglected tool for making multiple similar changes to a document. Find & Replace usually has a larger role to play when editing structured documents such as program source code or HTML, but can often accomplish surprising tasks for prose documents, especially when one takes advantage of the additional features in a word processor such as Microsoft Word. For instance, formatting codes throughout a Word document can be changed all at once, from italics to boldface, from headings to a sub-headings, from blue to red, and so forth. In advanced editors and word processors, even "invisible" characters can be found and replaced (e.g. tabs, paragraph marks/carriage returns, page breaks).

Keystroke	Meaning
Ctrl+F	Find a specified item.
F3	Find the next occurrence of the same search term as specified during the previous Find.
Ctrl+H	Find a specified item and Replace it with another item.

As the course progresses, we will take advantage of Find & Replace at every suitable opportunity.

The Clipboard

The clipboard is a system-wide data storage area that can be used to temporarily hold **one** "item" at a time. This "item" may be a selection of text, an image, a range of cells in a spreadsheet, or what have you. Each time a new item is added to the clipboard (through the Cut or Copy commands as described below), the previous item is lost. The items are stored along with a description of the type of data the item represents (e.g. plain text, formatted text, spreadsheet cells) and some programs store the item in multiple data formats. For instance, if a piece of formatted text (complete with bolded and italicized words) is copied to the clipboard from Microsoft Word, it can then be pasted into a Word document with all of its formatting intact, or into the Notepad editor as just the raw text.

Be forewarned that some applications are a little fast 'n' loose in their treatment of the clipboard metaphor. Notably, Microsoft Excel does not actually "cut" cells until they are pasted somewhere else; instead, a flashing "marquee" is displayed around the cut selection, leading new users to believe they may have selected Copy rather than Cut. Windows Explorer permits users to cut and paste files or whole directories from one location to another, but this data never passes through the system clipboard. Microsoft Office also introduces a multi-item clipboard that, while useful, can be annoying if you want nothing more than simple cut 'n' paste.

Aside from right-clicking the mouse and choosing Cut, Copy, or Paste from the pop-up menu, the clipboard commands are accessible through keyboard shortcuts. There are two styles of these keyboard shortcuts, but Microsoft now urges users to use the shortcuts listed in the New-Style column, below.

New-Style Keystroke	Old-Style Keystroke	Clipboard Command	Meaning
Ctrl+X	Shift+Delete	Cut	Copy the selection to the clipboard and delete it from the document.
Ctrl+C	Ctrl+Insert	Copy	Copy the selection to the clipboard and leave the document unaltered.
Ctrl+V	Shift+Insert	Paste	Paste the contents of the clipboard into the document at the insertion point.
PrintScreen (PrtScn)		Copy Screen	Copy the entire display screen into the clipboard as an image. Once in the clipboard, the image can be pasted using the regular commands.
Alt+PrintScreen		Copy Window	Copy the foreground window into the clipboard as an image. Once in the clipboard, the image can be pasted using the regular commands.

Representation of Text

Character Sets

When computers were still in their infancy, it was realized that there would have to be some method to translate the ones and zeros that computers use into the letters, numbers, and symbols with which humans are more familiar. The 7-bit ASCII (American Standard Code for Information Interchange) character set was invented to number all of these characters, at least the ones used by languages with Roman alphabets. Since a 7-bit wide character can only store 128 different values, that left a lot of other languages out in the cold, even those that use Roman alphabets but with accents or non-English letters or punctuation: French (Ã©), Swedish (Ã¸), German (ÃŸ), or Spanish (Â¿), to name a few.

A byproduct of the ASCII character set design is that capital letters occur before lowercase letters in the sequence. Therefore, some programs may sort items first by capitals and then by lowercase letters. Most modern programs sort using a case-insensitive algorithm; the same way that humans naturally do.

Most operating systems settled on an 8-bit (128 more characters) extension of ASCII called ANSI (American National Standards Institute) which added a number of accented characters, or diacritics, to the Roman alphabet. However, there are clearly many languages that do not use the Roman alphabet, or alphabets at all. Currently, the Unicode character set is the international champ, and can manage potentially billions of different characters which is more than adequate for all of the world's languages (including the Asian ideographic languages).

Special Characters

To access the special characters provided by ANSI or Unicode, you can use the **Alt** key and the numbers from the numeric keypad on the right of the keyboard (not the numbers across the top row). To insert a special character (provided you know its ANSI or Unicode code), press and hold the **Alt** key and then press the **0** key on the numeric keypad followed by the three-digit code of the special character, then release the **Alt** key.

For instance, pressing **Alt+0234** produces: Ãª.

To find out the codes for these special characters, open the Character Map tool (usually found under **Start** | **Programs** | **Accessories** | **System Tools**). When you click one of the displayed characters, the code will appear at the bottom-left corner of the window. You may also simply copy the selected character to the clipboard and then paste it into your document.

Line Terminators

In Windows, when you press the **Enter** (or Return) key while editing a text file, two characters are actually recorded: Carriage Return and Line Feed. These two characters together are called Line Terminators; they are used by Windows to tell when one line ends and another begins. This is reminiscent of typewriter days when to advance to the next line you had to return the carriage (the assembly with the platen roller) all the way to the right, and roll the page one line up, so that you could begin typing again on the first column of the next line. The carriage return arm usually performed both operations at once, but Microsoft chose to preserve the two actions.

However, other operating systems do not follow this convention. The Apple Macintosh uses a single Carriage Return character as its line terminator. And, wouldn't you know it, UNIX uses a single Line Feed character as its line terminator. This raises obvious problems when transferring files from one operating system to another.

Miscellaneous Editing Tidbits

Insert vs. Overwrite

Most of the time that you are editing text, you will be in "Insert Mode". This means that if the caret is positioned within a line of existing text, every key you press will position the corresponding character at the caret, pushing the rest of the line to the right of the caret one position to the right. You can also edit in "Overwrite Mode" in which case every character entered will take the place of the character immediately below the caret. In most programs, press the **Insert** key to toggle between the two modes. Usually an indicator at the bottom-right of the program window will display "INS" for Insert Mode (or nothing at all), and "OVR" for Overwrite Mode. Some programs, usually text editors, may also change the caret from a blinking line in Insert Mode, to a blinking character-wide block in Overwrite mode. Windows Notepad, however, does not have an Overwrite Mode.

Delete vs. Backspace

In Windows, the **Backspace** key erases the character immediately to the left of the caret while the **Delete** key erases the character immediately to the right of the caret. In other operating systems, both keys often behave identically, in some cases like the Windows Backspace and in others like Delete.

Word Wrap

Modern text editors and word processors automatically manage "Word Wrap": continuing a line too long to display on the following line without breaking in the middle of a word. Therefore, do not press the Enter key at the end of each line of text. Instead, just press the Enter key at the end of each paragraph.

Tabs vs. Spaces

Tabs are different from spaces. A space is always one character in width, but different programs evaluate the width of tab characters differently (e.g. 8 spaces, 4 spaces, 0.5"). The width of a space character is only consistent for fixed-width, or "non-proportional", fonts such as Courier New. In proportional fonts, such as Times New Roman or Arial, a space may be an arbitrary width, but in any case will usually be much more narrow than the other letter or number characters. Therefore, don't use spaces to horizontally align bits of text when using proportional fonts. Tabs are a better choice for alignment, but word processors afford an much more elegant method of aligning text, which we'll cover in the weeks to come.

Spaces following Periods

In typing or keyboarding class you may have learned to type two space characters after the period ending a sentence. You may also do the same after commas.

Stop it.

Two spaces after periods made sense for fixed-width typewriter fonts (Courier) but are obsolete now that we can use proportional fonts. The word processor will automatically adjust the width of the space following a period to fit the line, and two spaces make the text look odd.

The Internet

The Internet is a collection of interconnected networks. Each network is independent from its neighbours, and so ultimately, no one country or organization "owns" the Internet.

For instance, the computer network in the lab is connected to the overall network at the College, which connects to the YTG network, which connects to the NorthwesTel network for the North, which connects to networks in the rest of Canada, North America, and the World.

As do most technological advances, the Internet began as a military project. The story goes that its initial design requirement was to be nuclear blast-proof. It probably wasn't designed that way, but the Internet is very reliable even in the face of widespread outage. This is because it has no single point of failure: whole countries can go offline while outside the borders the Internet hums along (perhaps with degraded performance, but messages can still find a way to get through).

The Internet's design also mandates that all hosts (computers) connected through the network are equal, or peers. This facet of the design permits the "no single point of failure" goal since there are no easily-targeted "master" servers required to keep traffic moving.

An illustration of the growth of the Internet: in 1969, only 4 computers made up the then ARPANet; by 1971, the Internet had grown to 15 computers (the first e-mail message was sent in 1972); in 1973 the number was 37; skipping ahead to January 2002, the estimate ballooned to 147 million computers; by January 2004, 233 million; and by July of 2006, 439 million. Of course, like all statistics, these figures should be taken with a grain of salt: some estimates are twice as high.

Internet Layers

In order to understand a little better how the Internet actually works, we can divide it into four layers of responsibility. Starting with the hardware and moving upwards, the layers are: Physical, Address, Transport, and Application. (Technical references often divide networks into 7 layers, and use different names, but these 4 layers should be sufficient for our level of inquiry.)

Each of the layers on a single computer (or host) connected to the Internet communicates with the corresponding layer on other hosts. This communications is carried out through a number of **protocols**. A protocol is just the set of rules for such a conversation (a quick example that we're familiar with is that people are expected to say "Hello" or something similar, when answering the telephone).

Physical Layer

The physical layer is just that: physical. It contains all of the stuff we can see and touch, together with the protocols that allow this hardware to communicate with other hardware.

Examples of physical layer hardware: modems, network cables, network interface cards (NIC), wireless cards and antennae (for Bluetooth or Wi-Fi), Ethernet hubs (Ethernet is the most popular physical layer protocol, and is the one used in the College labs), and now even G3 cell phones can connect to data networks.

The protocols at this layer just make sure data can pass between these pieces of hardware; they have no concept of how this hardware is strung together to form the larger Internet.

When speaking of the physical layer, the most often bandied-about term is **bandwidth**, meaning the amount of raw data that can travel through the wires in a given measure of time, usually bits per second, or "bps". A slow telephone modem runs at about 28Kbps (thousand bits per second), faster modems at up to 56Kbps. A cable or DSL modem can run between 500Kbps and 5Mbps (million bits per second). Wireless (Wi-Fi) networks usually runs at 54Mbps. And an internal network like in the College labs can run between 10Mbps and 100Mbps.

Address Layer

The address layer is responsible for figuring out where a specific host is on the Internet, just given its address. A reasonable analogy would be the post office, who somehow figure out how to deliver mail around the world using only the address printed on an envelope.

The address layer also breaks single messages into bite-sized chunks. These chunks, or **packets** as they're often called (also streams, segments, and frames), are sent across the network(s) independently. This means that the many packets that make up a single message may take completely different routes through the Internet from source to destination. This design better ensures reliable delivery of information in the face of Internet outages, but requires the transport layer (see next) to perform some extra work to guarantee that the multi-packet message arrives intact.

The address layer doesn't pre-calculate the entire route for a packet to take on the Internet. Instead it simply figures out what the next part of the route should be as the packet bounces from one server to the next as it moves from its source to its intended destination. In this way, whole sections of the Internet can go down (perhaps following a nuclear attack, but more likely due to blackouts) but messages can still get through as the address layer figures out alternative routes through Internet nodes that are still functioning.

Computers being computers, the addresses for hosts are all numbers. Every host connected to the Internet has a unique address number, which allows any computer anywhere in the world to communicate with precisely one other computer using just this address. This is no different from North American 10-digit telephone numbers: no two are the same. If two people did have the same telephone number, how could you contact just one of them?

The protocol that carries out this addressing is called the **Internet Protocol (IP)**, and so these numeric addresses are called **IP addresses**. You most likely have run into these addresses before; they are composed of a sequence of four numbers that range between 0 and 255. For example, the IP address for the computer you are connected to as you read this is: 199.247.245.153. These numbers are handed out by ISPs and telephone companies. Most Yukon IP addresses begin with 199.247. The IP address for the Yukon College's webserver is 199.247.245.8.

These numerical IP addresses are terribly inconvenient for humans. Fortunately, we can represent these addresses in a text form called a Uniform Resource Locator (URL). See the section below for more about URLs.

Transport Layer

The transport layer's responsibility is to ensure that messages arrive intact at the destination host.

The most common protocol running in the transport layer is the **Transport Control Protocol (TCP)**. The combination name **TCP/IP** is often used as the general name for the protocols underlying the Internet.

There is also the **User Datagram Protocol (UDP)**, but it is not used as often and doesn't make the same reliability guarantees as does TCP.

TCP carries out the transport layer's tasks by guaranteeing that all packets in a message arrive at the destination, in the proper order, and without any data corruption. The order of the packets is important if you consider an Internet message that transmits video, for example. If the packets' order is shuffled, bits of the video would appear in the wrong order and the overall effect would be unwatchable (unlike, say, Memento or Pulp Fiction, which both reorder the normal sequence of events, but in an eminently watchable fashion).

TCP can define multiple Internet services on a single computer. The Internet Protocol will get messages to a specific computer, but these transport layer protocols are required to direct the message to the proper program running on that computer. They do this by assigning numeric **ports** (0 through 65535) to certain services. So, if a single computer is running both a webserver program and an e-mail server program, web requests will be directed by TCP to the port used by

the webserver program, and e-mail messages will be directed by TCP to the port used by the e-mail server program.

Compare the transport layer's idea of ports to telephone extension numbers in a business office. There may just be one main telephone number (the IP address) but many extension numbers (ports) for the various departments in the business.

Application Layer

At the very top of the Internet layer "cake" is the application layer. The protocols in this layer deliver all of the Internet services we're familiar with:

- HTTP (HyperText Transfer Protocol, TCP port 80): transmits web pages across the Internet
- FTP (File Transfer Protocol, TCP ports 20 & 21): used for moving files from one machine to another
- SMTP (Simple Mail Transfer Protocol, TCP port 25): for sending e-mail messages
- POP3 (Post Office Protocol, version 3, TCP port 110): for retrieving e-mail messages from a e-mail server
- NNTP (Network News Transfer Protocol, TCP port 119): network newsgroups
- Telnet (TCP port 23): allows users to log on to remote computers just as if they were sitting in front of the remote computer
- DNS (Domain Name Service, UDP port 53): translates the machine name part of the URL into IP addresses (DNS is the only one of the protocols listed here that used UDP as its transport layer protocol)

There are hundreds, if not thousands, of such protocols, but those are some of the most common ones.

Each protocol is designed specifically for its task: HTTP retrieves web pages immediately; SMTP eventually gets the e-mail to its destination but doesn't require that it get there right away, FTP is tuned for transferring large files, but isn't as fast as HTTP when it comes to smaller files.

Uniform Resource Locators (URLs)

Humans being humans, prefer not to have to remember IP addresses like "17.112.152.32". Instead, we'd rather use a little piece of text like "www.apple.com", which is called a **Uniform Resource Locator**, or **URL** (pronounced "You-Are-Elle", or sometimes "Earl"). In fact, both 17.112.152.32 and www.apple.com refer to the same Internet host: Apple Computer's webserver. Don't believe me? Try just entering 17.112.152.32 into a browser and see where you end up.

URL's are used for more than just websites; they're also used for e-mail addresses, FTP servers, network newsgroups, or any other service running on a computer connected to the Internet.

When we type in a URL, an application-layer service called the **Domain Name Service (DNS)** translates the human-friendly URL into the computer-friendly IP address for us. DNS performs this translation by consulting the databases maintained by the Domain Name Registrars. DNS is responsible for translating "www.apple.com" into "17.112.152.32" whenever we type that into a browser address bar. DNS is the "phone book" of the Internet: look up a name, and find its number.

A URL is made up of many parts. For example, consider the following:

```
http://www.apple.com/ipod/red/index.html
```

- `http` is the application-layer protocol. In this case HTTP, the protocol used by the World Wide Web. If you omit the protocol when typing a URL into a browser address box, the browser assumes you meant "`http://`". (Technically, the whole string "`http://`" is called the **scheme**.)

- `www.apple.com` is the **Fully Qualified Domain Name (FQDN)**. In turn, the FQDN is made up of pieces as well (from right to left, or most general to most specific):
    - `com` is the **Top Level Domain (TLD)**, in this case meaning "company" or "commercial".
    - `apple` is the specific company name, chosen by Apple Computer.
    - `www` is the name of the webserver hosting Apple Computer's website. Using "www" for the name of the webserver is merely a convention, and you will run into cases where "www" is not used (e.g. the course website at "cpsc100.yukoncollege.yk.ca").
- `/ipod/red/index.html` is the path to the requested HTML page on the webserver. This path and file is used by the webserver to figure out which file to return to a browser. Other application-layer protocols might not need a path or file name.

And now a more complex URL:

`http://www.tc.gov.yk.ca/digitization/public/index.php`

- `http` is the HTTP protocol.
- `www.tc.gov.yk.ca` is the Fully Qualified Domain Name:
    - `ca` is the Top Level Domain (TLD), meaning "Canada" in this context. There are only a few designated Top Level Domains. See the section below for a description of these TLDs.
    - `yk` indicates a sub-federal geographical region, in this case the Yukon (why they used "yk" instead of "yt" is anybody's guess).
    - `gov` was chosen by YTG to mean "government".
    - `tc` is a department within YTG (Tourism & Culture in this case).
    - `www` is the host computer.
- `/digitization/public/index.php` is the path to the requested resource (page) on the webserver (PHP pages are just like HTML but with a bit of programming smarts).

A simple URL: `http://slashdot.org`

- `http` is the HTTP protocol.
- `slashdot.org` is the Fully Qualified Domain Name:
    - `org` is the Top Level Domain (TLD), meaning an "organization".
    - `slashdot` is the organization, and it was chosen by the fine folks at Slashdot:
    - There is no host name. If no host name is specified, a default server at that domain is used. This just happens to be Slashdot's webserver, so this works fine.
- There is no path to a file. In this case, the default HTML document is selected from the root directory of the webserver.

Top Level Domains

There are only a few top level domains, and these are controlled by one of the Internet self-regulating authorities (ICANN[1]).

The most common TLDs are as follows:

- .com: companies or commercial ventures
- .net: ISPs or other companies that work with the Internet
- .org: non-profit organizations and institutions

[1] www.icann.org

- .mil: US military
- .edu: US education institutions
- .gov: US government

Recently, ICANN specified some new TLDs that haven't really caught on much yet: .info, .biz, .pro, .museum, .aero, etc.

TLDs are also assigned to countries:

- .ca: Canada
- .us: United States
- .uk: United Kingdom
- .tv: the tiny South Pacific island nation of Tuvalu which bases most of its economy on the sale of domains using this TLD

Note that some of these TLDs can be split up into sub-federal regions: .yk.ca is the Yukon, .on.ca is Ontario.

The domain name part of the URL can be chosen by anyone (but you can't pick something that's already owned) and is registered with one of many companies called **Domain Name Registrars**. When you register a domain name, you really register the combination of domain name and TLD, so if someone's already registered *yourname*.com, you can always try for *yourname*.org or *yourname*.net or *yourname*.ca. Even though you're not really a Internet company nor an organization, you can still use those TLDs.

Internet Security and Privacy

Why is Security Needed?

The Internet was originally designed for sharing information between collaborating scientists. And at the time (early 70s) personal computers didn't yet exist. Nor, for that matter, were there widespread data network lines. So, security didn't much enter into the thinking of the Internet's founders; they were far more concerned with robust message delivery in the face of an anticipated World War III.

In fact, the very open nature of the Internet protocols likely contributed to its enormous success through the 1990s and today.

But when so many people can connect so easily, there are bound to be a few bad apples trying to ruin the party. In earlier days, a tremendous amount of technical know-how was required to crack into computers through the network. Gradually the software used by these `l33t hax0rs` ("elite hackers" in the lingo) made it onto the Internet and could be wielded by any old pimply-faced high-school punk, the aptly-named "script kiddies". (The self-titled MafiaBoy who cracked CNN, eBay, Yahoo!, and others in 2000, was a 15-year old Montreal brat who fit nicely into this category.)

These software tools are now run automatically by programs that search the Internet for weak spots. It is said that an unprotected computer will be targeted within 15 minutes of its initial network connection. Some security companies set up such computers (called "honeypots") from time to time to attract just this sort of malicious attention in order to study the means and methods of the attackers.

Writing viruses and breaking into computer systems was once an ego-driven activity. Now it's a business. Taking over personal computers (called zombies) can be profitable. Control of these zombies is traded and sold in blocks of thousands. The victimized computers are then used to send out spam e-mail messages, or attack a certain computer in coordination with one another. The latest nefarious idea is to break into a computer, encrypt all of its files, and then demand a ransom payment from the computer owner: no money, no files.

Even with all of this technical acumen so widely available, the number one way to gain illegal entry to a compute system is through "social engineering". That is, calling up the target organization and weaseling your way into their confidence, perhaps enough to get a naive desk jockey to cough up a password. Walking purposefully through a building with a clipboard is an awfully good way to get into restricted areas.

Disgruntled employees are also a great source of trouble to security-conscious organizations. But these unhappy folks are unlikely to cause the Internet population as a whole much harm.

Securing Your Computer

Securing your computer really means two things: preserving the integrity of your files and keeping them out of the hands of others, and preventing your computer from being used as a launchpad for security attacks on other computers. One of the best defenses is using your common sense:

- Never open e-mail attachments, even if sent by people you know, unless you have already requested a specific file from them, or can verify directly that the person meant to send you a file. Email attachments are the number one infection vector of today's viruses, and most come masquerading as something innocuous, and often appear to have been sent by an acquaintance.

- When surfing the web, do not download software unless you are sure you understand what the software does, and who created it. When using Microsoft's Internet Explorer browser, you will see a pop-up window like the following when a website wants you to download a program:

  Click the **Don't Install** button unless you're absolutely sure. (This example shows Macromedia's Shockwave Player which is probably safe, but if you don't really need it, don't download it.)
- Consider changing your web browser program from Microsoft Internet Explorer to Mozilla Firefox[2], Google Chrome[3], Opera[4], or Safari[5]. All four are free and are currently believed to be much less susceptible to browser-based security intrusions than IE. Firefox is the most popular of these alternatives at the moment and is easy to customize with downloadable "extensions" that, for example, block advertisements, show you the current weather forecast, look up the definition of any word on a page, and synchronize your bookmarked sites across multiple computers, to name just a very few. (Of course, extensions may themselves be security issues, but so far there haven't been any major problems.)
- Regularly check that your computer has all of the latest updates (a.k.a. "patches") from the operating system manufacturer. For Microsoft Windows, visit Microsoft Windows Update[6]. If you're using Windows XP or Vista, you should already have the Automatic Updates feature turned on which will do the checking and downloading for you.

If you ever do suspect that your computer has been compromised in some way, immediately disconnect it from the network (physically unplug the cable) and turn off the computer's power. Use a different computer--one that you know is clean--if you want to search the Internet for a solution to your predicament.

Computer Viruses

Computer viruses existed long before the popularization of the Internet, but they can now proliferate far more rapidly in the connected world.

---

[2] www.mozilla.com/firefox

[3] www.google.com/chrome

[4] www.opera.com

[5] www.apple.com/safari

[6] windowsupdate.microsoft.com

There are actually three types of infectious programs: viruses, worms, and trojan horses.

A computer virus is a remarkable analogy of its biological counterpart. Like a bio-virus, computer viruses are not "alive", in the sense that a computer virus must first infect a normal program before its code can be run. A bio-virus must infect a cell before it can hijack the DNA replication in the cell to create copies of itself. Once a computer virus has infected a program, the next time the program (possibly even part of the operating system) runs, the virus's payload will be activated. Typically a virus will try to copy itself to other programs, or across the network, but most also contain some form of harmful instructions: delete files, format drives, send personal files across the Internet, and so forth.

A computer worm, on the other hand, is usually a fully-functional program that depends exclusively on a network for its reproduction. The MS Blaster worm that attacked, and continues to attack, Microsoft Windows installations is an example of such a worm. Once a computer is infected, the worm will create copies of itself with instructions for each to seek out another nearby Windows computer and infect it also. In this way, the MS Blaster worm (and variants) circled the globe in less than 24 hours. (The "Slammer" worm is estimated to have reached every Internet-connected computer in the world within 10 minutes.)

Finally, a trojan horse is a program that looks normal (even inviting, like the original Trojan Horse), but that carries a terrible payload. Some trojan horses may masquerade as cute little dancing frogs on the screen, or whatever. Others may set up a screen that looks exactly like the normal Windows login screen and as soon as you type in your user name and password, the trojan horse sends that information off to someone on the Internet, and then pops up the regular login screen (just as if you'd accidentally typed an incorrect password).

Anti-Virus Tools

An anti-virus utility is essential for any 'net-connected computer. These utilities scan each file that enters or leaves (through the web, e-mail, keychain drive, floppy disk) your computer for known viruses and deletes, or at least "quarantines", them when found. Be sure to subscribe to the maker's update service (usually a year or so is included in the purchase price) so that your computer will be protected from the most recent viruses. It's also a good practice to run a complete check of your computer files at least every month, even though this can sometimes take hours. The following companies offer the most popular anti-virus tools at the moment.

- Norton AntiVirus[7] (Norton also bundles their anti-virus software together with a firewall and other tools in the Norton Internet Security[8] suite.)
- McAfee virusscan[9] (Also available in suite form[10].)
- Trend Micro's PC-cillin Internet Security[11] (Only available in a bundled suite form.)
- F-Secure Anti-Virus[12] (Also available in suite form[13].)

---

[7] www.symantec.com/nav/nav_9xnt

[8] www.symantec.com/sabu/nis/nis_pe

[9] us.mcafee.com/root/package.asp?pkgid=100&cid=9052

[10] us.mcafee.com/common/en-us/promos/mis6taxcut/default.asp?cid=9395

[11] www.trendmicro.com/en/products/desktop/pc-cillin/evaluate/overview.htm

[12] www.f-secure.com/home_user/antivirus.html

[13] www.f-secure.com/home_user/complete_protection.html

There are also some reasonably good free (for personal use) anti-virus packages:

- avast! 4 Home[14]
- AVG Anti-Virus[15]
- Comodo Internet Security[16] (Also includes a firewall.)
- Avira Antivir [17]
- ClamWin (not for beginners)[18]

If you receive an e-mail message warning you about a new virus, don't believe it, and don't delete any system files without assurance from a recognized technical support person. A good rule of thumb is to count the number of exclamation points in a virus alert. If there are more than zero, it's probably a hoax.

Anti-Adware/Malware/Spyware Tools

There is another category of virus-like software variously called adware, malware, or spyware. These programs usually display extra advertisements, sometimes in your browser, sometimes independently of your browser. While generally not as destructive as viruses, these programs are certainly annoying, may prevent you from being able to surf the web, and may release private information about your surfing of online shopping habits. If you're a fan of online music sharing, chances are your computer is already infected with these. Fortunately, there are a couple of free tools that can help eradicate these gremlins:

- Spybot - Search & Destroy[19]
- Ad-Aware[20]

Unlike anti-virus programs, you can install more than one of these tools at the same time, and that's not a bad idea.

Firewalls

A "network firewall" is like a filter for Internet messages: it lets some pass through, but blocks others.

A firewall selectively allows certain messages to pass through based on its type (e-mail, web, file sharing). All other messages are stopped at the gates. In this manner, you can protect yourself to a certain degree by simply not allowing Internet traffic to contaminate your machine unless it arrives through a recognized port (e.g. the one used for the web, or for e-mail, or for file transfers).

Obviously, a firewall can't protect you from malicious messages that arrive on ports that you need to remain open, so a firewall is never a complete security solution.

However, firewalls are pretty good at fending off network worms.

---

[14] www.avast.com/eng/download-avast-home.html

[15] free.avg.com/download-avg-anti-virus-free-edition

[16] www.personalfirewall.comodo.com

[17] www.free-av.com

[18] www.clamwin.com

[19] www.safer-networking.org/index.php?page=download

[20] www.lavasoftusa.com/software/adaware

Internet security companies, journalists, and even Microsoft, are now saying that every computer should have not only an anti-virus scanner, but an installed and maintained firewall as well.

Firewall Tools

Together with anti-virus protection, firewall utilities are also essential. Firewalls filter the network traffic that enters or leaves your computer, throwing out anything that you haven't explicitly allowed.

Windows includes a firewall, although it's not as flexible as those from other sources.

Zone Labs' ZoneAlarm[21] is a free firewall for non-commercial use, although you get the "dumbed-down" version. If you're paying, it's often cheaper to buy a security suite product that includes both anti-virus and firewall--among other utilities--together.

If you want to test your firewall, check out Gibson Research's Shields Up![22] (you'll have to search through the site to find it as there's no direct link to Shields Up!.) Once there, run the File Sharing and Common Ports tests.

Protecting Your Privacy

Protecting your online privacy is the best step you can take in preventing the new but popular crime of Identity Theft: stealing someone's "identity" in order to fraudulently carry out monetary transactions in the name of the stolen identity.

The more pieces of information such a thief can piece together, the easier it is for him or her to request credit cards, a SIN number, driver's licenses, and so forth, all in your name. The thief then usually max-es out all of your lines of credit and skedaddles, leaving you responsible for the charges, and facing a very difficult process to re-establish your credit rating.

Of the 10,000 or so arrests made by the US Secret Service in 2002 (the agency responsible for such financial crimes), 94% of them were related to identity theft.

- Never give out your Social Insurance Number. By law, only certain government departments can ask for it. There is a great deal of information about protecting your SIN at the Privacy Commissioner of Canada's website[23].
- Avoid giving out any more personal information that absolutely possible. If a company's website requires a lot of information before they'll offer some service, either don't subscribe at all, or fill in fake information (serves them right for asking in the first place).
- Shred or destroy any mail or personal documents that you plan to throw in the garbage. You'd be surprised at what someone can find out about you by going through your trash (a.k.a. "dumpster diving").
- Check your credit rating periodically to make sure something hasn't injured it without your knowledge. The two biggest bureaus in Canada are Equifax[24] and TransUnion[25]. However, be warned that it is possible to lower your credit rating simply by checking it--credit bureaus are just that evil.

If you suspect you've been the target of identity theft, follow the list of instructions found on the Privacy Commissioner of Canada's website.

---

[21] www.zonelabs.com/store/content/catalog/products/sku_list_za.jsp?lid=nav_za

[22] www.grc.com

[23] www.privcom.gc.ca/fs-fi/02_05_d_02_e.asp

[24] www.equifax.com/EFX_Canada

[25] www.tuc.ca

Spam

Spam is the e-mail equivalent of the junk mail (pamphlets, menus, shopping flyers) that fill your non-Internet mailbox every day. Spam is simply any piece of unsolicited e-mail: it may be sent to just you, or it may be sent to a million others--either way it's spam if you didn't ask for it.

The term "spam" originated, as do many geeky concepts, from a Monty Python skit[26].

Lists of 100,000s or even millions of e-mail addresses are available for purchase which can then be used to send out that many spam messages in a few minutes. Even if only a minute fraction of people respond, spam is still the most cost-efficient marketing method in existence.

Preventing Spam

Spam is not usually dangerous (unless it comes with a virus attachment) but is more than a little annoying. There is no sure-fire way of preventing spam, other than refusing to use e-mail, but there are a few techniques that can help mitigate the onslaught.

- Don't reply to spam or click the "unsubscribe" links that are often part of the message. Doing so rarely does anything, but may alert the spammer that your e-mail address works: maybe you didn't want herbal Viagra, but maybe you will want 1,001 ideas for working from home.
- Use two e-mail accounts: one just for friends and family, and the other for any kind of online registration. You may want to use a free online e-mail account (e.g., Gmail, Hotmail, Yahoo!) for this "public" purpose but be warned that some online registrations won't accept those e-mail accounts. You can configure your e-mail client program to automatically dump the public account's messages into a separate folder so that they don't bother you. Once a week or so, you can go through that folder looking for legitimate messages.
- Never display your e-mail account on a website. Automated programs--called spambots or e-mail harvesters--search through the web gathering e-mail addresses from websites. If you really need people to know your e-mail address, type it out in an odd way: "drogers STRUDEL yukoncollege CIRCLE yk FULLSTOP ca". Visitors to the site won't be able to click it to send you a message, but they can still figure it out on their own, and spambots can't pick it up (although they're getting smarter and are looking for patterns like the one shown).
- Use a spam filter program (which may already be included in your e-mail client program) to automatically throw out obvious spam. The junk mail controls in Mozilla Thunderbird[27] learn over time what is spam and what isn't and do a very good job of cleaning up your inbox. The free online e-mail services are also quite good at detecting and eliminating spam.
- If offered, sign up to your ISP's anti-spam service. These filter out the spam before they ever get to your inbox, and catch most viruses at the same time. They're not foolproof though, and may discard legitimate messages by accident (i.e. false positives).
- Duh, never purchase any product or service sold through spam.

Authenticating Sites and Using Encryption

There are two main problems with the out-of-the-box Internet: identifying people or computers at the other end of the network, and guarding against eavesdroppers. Authentication solves the first problem, and encryption solves the second.

---

[26] www.detritus.org/spam/skit.html

[27] www.mozilla.com/thunderbird

Authentication allows us to identify people or organizations, and their computers. Normally, when meeting people face-to-face, we can identify them by their appearance, voice, or some form of recognized document, such as a passport or driver's licence. Similarly, we identify organizations by address, or perhaps business licence.

Across the Internet, authentication is much more difficult, and so mechanisms such as user name and password were invented to allow computer systems to recognize (authenticate) valid users. Passwords have a number of weaknesses, so the latest authentication rage is biometric authentication: fingerprint scans, retinal scans, face recognition.

All such authentication mechanisms have a common weakness: before any passwords or biometric information is traded, how can you know if the person you are talking to *really is* the person you think you're talking to?

Web shoppers have another authentication problem: how can you tell for certain that a company's e-commerce website truly is run by that company, and not some con artist who set up a fraudulent site to appear exactly like the legitimate company's website?

The solution to both these problems is something called a "digital certificate". A certificate is issued by a recognized Trust Authority (Verisign being the most well-known TA) that promises to only issue such certificates once the recipient--individual or organization--has credibly proved their identity to the authority. The actual process for applying for these certificates requires faxing all sort of incorporation documents (if one wants a corporate certificate), the names of the organization's president and directors, and a final voice confirmation from the trust authority to the president of the organization to ensure everything is on the level. The application process for individuals is a little less rigorous.

Once these certificates have been granted, the person or organization possessing the certificate can be reliably authenticated (presuming one trusts the Trust Authority). The technical makeup of these certificates (text files filled with a long string of mumbo-jumbo characters--really just a very, very long number) more or less guarantees that they can't be forged, at least not without an overwhelming investment in computer processing power and time, or a revolution in mathematical number theory.

Even once the two parties in an Internet conversation have been authenticated, and the operations that they may perform have been authorized, the messages that are sent back and forth are still readable by anyone able to listen in (at any of the points along the route between the two parties, or even on the local network of one of the parties if a simple broadcast protocol like Ethernet is used). These messages might contain private information such as: passwords, credit card numbers, or sensitive records about a person (e.g. legal, tax, medical).

The Internet was originally designed to allow the free exchange of information. No one at the time realized the potential of the Internet, and its consequent commercialization. Most of the protocols in use hearken back to those early days, and so do nothing to hide the contents of their messages.

In response to this problem, a series of encryption technologies was developed to scramble these messages, and then unscramble them once they reached their proper destination. The most common of these technologies is called SSL (Secure Sockets Layer, although the technology as a whole is now known as Transport Layer Security--TLS). SSL is usually added to an existing protocol to better secure its messages: HTTPS encrypts HTTP web traffic, SSH encrypts both telnet and FTP traffic, and PGP (Pretty Good Privacy) is one way to encrypt e-mail messages.

SSL is a certificate-based system, meaning that using SSL implies both authentication and encryption.

Whenever you must submit personal information to a website (e.g. a credit card number or password) you'll want to ensure two things: the website you're visiting is the one you think it is, and your personal information can't be read by eavesdroppers anywhere along the network. To address these two concerns, a website should implement an SSL certificate and use the encrypted HTTPS protocol. Browsers vary slightly in how they let you verify that SSL is in use, but not by much.

- The first clue that SSL encryption is in operation is that the browser's address bar will show the URL's protocol as `http`**`s`**`://` rather than the usual `http://`. Most browsers also display a padlock icon somewhere on the screen, usually at the right-hand side of the address bar, or along the bottom status bar.
- Click or double-click the padlock to view the details of the certificate. The information shown is hard to read, but it should identify both the website's domain name and the owner's company name in the Subject field of the certificate's details. If those don't appear to be what you were expecting, it's time to move on. Sometimes legitimate companies use third-party online stores to sell their services and the certificate will have been issued to that third-party company. If you're at all unsure, don't continue.

Some sites will only use the encryption feature of SSL and skip the costs associated with the authentication certificate (e.g. some Yukon College sites work this way).

Browser Footprints

When you surf the web you leave behind a number of "footprints" that can be used by others to determine which sites you visited, your geographical location, your browser settings, your operating system preferences, and in some cases even your e-mail address. This is more of a concern when the computer you use is public, or shared between co-workers or family.

Internet Explorer version 7 has conveniently organized the method for deleting all of the common types of these footprints in one location: choose **Internet Options** from the **Tools** menu, choose the **General** tab and then click the button labelled **Delete...** under the Browsing history section.

- As a performance measure, browsers temporarily store the HTML and image files of the sites you've visited. This temporary storage is often called a "cache". A knowledgeable user can find this cache on the computer's hard disk and see what sites you've visited. To delete these files, click the **Delete files** button.
- Many websites also leave small files on your computer called "cookies". Cookies are not really a security concern, but they are a privacy concern. Modern browsers can be configured to block out some or all cookies. To delete all cookies, click the **Delete cookies** button.
- Browsers also keep track of the pages you visit (so that the "back" button works, and also to change the link colour of sites you've already visited). The next person to use the browser can easily see all of the sites you've visited, called the "history" (in IE, press **Ctrl+H** to view the browsing history). To delete this history, click the **Delete history** button.
- As a convenience feature, browsers will automatically fill out forms and passwords on your behalf. This means that the next person to use the browser could visit the same pages that you did and recover passwords, credit card numbers, or other sensitive information. To delete these form entries and passwords, click both the **Delete forms** and **Delete passwords** buttons.
- Of course, you could just delete all of the above by clicking the **Delete all** button at the bottom of the screen.
- Finally, once you've finished using the computer, close down the browser program so that nothing is left in the computer's volatile memory. Rebooting the computer is also a fine idea.

Even if you're not using a public machine, at some point you will likely submit some information to websites to make online purchases or to register for some online service. Whenever you do this, you should first check if the company that runs the site has any sort of privacy policy in place to guarantee that your personal information won't be misused or sold (possibly to a spammer).

There are organizations that supposedly monitor these privacy policies such as TRUSTe[28] but they have a spotty record of actually enforcing any kind of rules on misbehaving companies. Some companies have even been known to modify their privacy policies without telling anyone. eBay, Amazon, and Yahoo! have all been caught changing people's personal preferences to allow for automated e-mail messages (spam, in so many words), even if the person didn't select any such "service" when they originally set up their profile.

There are tools and services such as Anonymizer[29] that will let you surf the web with a reasonably high level of anonymity. The same tools often delete any history, cached files, or cookies from your machine as a further precaution.

[28] www.truste.org

[29] www.anonymizer.com

An HTML Primer

The Hypertext Markup Language is "a document-layout and hyperlink-specification language. It defines the syntax and placement of special, embedded directions that aren't displayed by the browser, but tell it how to display the contents of the document, including text, images, and other support media. The language also tells you how to make a document interactive through special hypertext links, which connect your document with other documents--on either your computer or someone else's, as well as with other Internet resources."

from *HTML: The Definitive Guide*, 3rd Edition, by Chuck Musciano & Bill Kennedy, O'Reilly & Associates Inc.

Online HTML Tutorials

This primer only covers the basics of HTML. In order to learn more of the language, visit one or more of the following online tutorials:

- W3C's Getting started with HTML[30]
- W3Schools' HTML Tutorial[31]
- DeveloperZone's HTML Tag Quick Reference Guide[32]
- W3C's Structure of an HTML 4.0 Document[33]

And don't forget to use your browser's View Source feature to learn the tricks used by well-designed websites.

The HTML Format

HTML is a text-based format: a simple text editor like Windows Notepad can be used to create even the most complex web pages (although any embedded images, music, or video would have to be created with specialized tools). HTML uses **tags** both to control the layout of a document and insert hyperlinks or references to embedded multimedia items. These tags are surrounded by angle brackets (< and >) to distinguish them from the content of the document.

There are two main types of tags: **container tags** and **standalone tags**. Container tags surround a discrete region of the document with begin and end tags to specify a format or behaviour for that region. Container tags may be embedded within one another. A standalone tag simply inserts a non-text item into the document at that position (e.g. line break, horizontal line, image). Most HTML tags are of the container variety. Some container tags do not require end tags if the end of the contained region is obvious, but it's still a good idea to explicitly write end tags.

Each tag has a type and, optionally, one or more **attributes**. The end tag for a container tag always consists of a slash followed by the tag type within the angle brackets, and end tags never have attributes.

For example, to italicize a portion of text, we use a container tag of type `i`:

```
<i>This is italicized.</i>
```

Which, when viewed in a browser appears like this:

This is italicized.

---

[30] www.w3.org/MarkUp/Guide

[31] www.w3schools.com/html/default.asp

[32] www.devx.com/projectcool/developer/reference/tag-table.html

[33] www.htmlhelp.com/reference/html40/structure.html

We can also embed another container tag to display just the word "italicized" in boldface by using the `b` type tag:

```
<i>This is <b>italicized</b>.</i>
```
*This is **italicized**.*

Note that both the begin and end tags for the boldface format are completely inside the italics container tags. While most browsers aren't confused if the container regions overlap, it's not a good idea.

Standalone tags have no end tags. For example, to insert a line into a document we use the `hr` (horizontal rule) type tag. Optionally, we can add a `width` attribute to limit its width to 60% of the display space:

```
<hr width="60%">
```
Which, when viewed in a browser appears like this:

In this case, the `width` attribute is assigned a value of 60%. Note that the HTML standard requires that an attribute value appear inside double-quotes.

HTML ignores all but one contiguous whitespace character, including spaces, tabs, and carriage returns, treating them all like a single space. For one thing, this means that you cannot position items horizontally using spaces or tabs. Special tags and HTML tables can be used for that purpose.

HTML tags and attributes are not case-sensitive. That is, `<b>` and `<B>` mean the same thing. It's a common convention to always use the lowercase form.

The HTML Document

In order to indicate to a browser that a given file should be interpreted as an **HTML document**, we use the `html` container tag. Each HTML document includes a header section (`head` tag) that contains information about the document such as its title and character set. HTML documents also include a body section (`body` tag) that contains the actual content of our HTML document.

Therefore, a very simple HTML document might be as follows:

```
<html>
 <head>
  <title>Simple HTML Document</title>
 </head>
 <body>
  This is a simple HTML document.
 </body>
</html>
```

When displayed in a browser, all that would appear is the line "This is a simple HTML document." and the title bar of the browser would display "Simple HTML Document".

Indenting the HTML tags, as in the above example, to indicate the overall structure of the container tags is not necessary, but it often makes reading the raw HTML code easier.

Hyperlinks

Hyperlinks connect HTML documents together so that someone using a web browser can navigate from document to document by following the hyperlinks. When displayed, a hyperlink usually appears as a differently-coloured section of text (or an image) that is underlined. However, the appearance of these links is easily changed, so there is no standard from website to website.

Hyperlinks are embedded in HTML documents using the `a` tag. Clicking a hyperlink instructs the browser to open the document pointed to by the `a` tag's `href` attribute. The value of the `href` attribute is the URL of the target document.

For example, the following HTML creates a link to the Yukon College site:

```
<a href="http://www.yukoncollege.yk.ca">Jump to Yukon College</a>
```

Which, when viewed in a browser appears like this:

Jump to Yukon College

Sample Container Tags

Tag	Meaning
`<html>…</html>`	The html tag defines the beginning and ending of an HTML document. Usually this is the entire text file that contains the document.
`<head>…</head>`	This tag contains other tags that describe the HTML document in some way, but are not part of the document's visible content.
`<title>…</title>`	The title tag specifies the formal title of the document, and is usually displayed in the title bar of the browser. It is also the name used when a page is bookmarked. The title tag must appear within <head> section of the document.
`<body>…</body>`	The body tag encloses all of the content that will be displayed in the browser.
`<p>…</p>`	Defines the region as a paragraph, inserting an extra empty line after the end of the region.
`<i>…</i>`	Displays the region in *italics* (… does the same).
`<b>…</b>`	Displays the region in **boldface** (… does the same).
`<u>…</u>`	Underlines the region.
`<h`*x*`>…</h`*x*`>`	Formats the region as a heading (or title). The values for *x* can range from 1 through 6, with 1 representing the largest font.
`<font>…</font>`	Controls the specific font for the region using optional attributes such as face, color, and size.
`<a>…</a>`	Defines a hyperlink with the target URL specified in the href attribute.

Sample Standalone Tags

Tag	Meaning
` `	Inserts a line break into the document. This is often more convenient than the <p> container tag for formatting text.
`<img>`	Inserts an image into the document. The source of the image is defined using the src attribute which specifies the URL for the image.
`<hr>`	Inserts a horizontal line (or rule) into the document.

HTML Character Entities

Because HTML uses the < and > characters to denote tags, you may be wondering how to write a literal "<" character in a document (as I just did in this line). HTML provides a selection of **character entities** that can be used to embed any special characters that are either used by

HTML for its own purposes, or that are difficult to produce using a western keyboard. The w3schools.com site includes a long list of character entities[34].

[34] www.w3schools.com/tags/ref_entities.asp

Note that characters from the ANSI character set can also be displayed using the syntax: `&#x;`, where *x* is the one- to three-digit ANSI code of the required character.

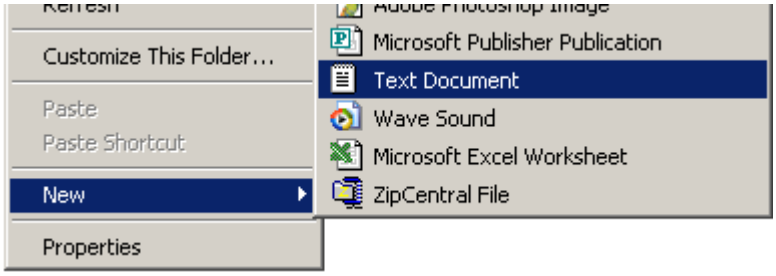
Entity	Visible As	Meaning
`<`	<	Less-than sign.
`>`	>	Greater-than sign.
`&`	&	Ampersand. Since the & character is used to introduce character entities, ampersands must always appear in HTML documents as: &.
` `		Non-breaking space. Just like a regular space, although multiple contiguous sequences are not compressed into a single space. This is particularly useful to prevent line breaks between two words that should always appear together.
`&quot;`	"	Double-quote. Most browsers permit just " to appear in HTML documents, but technically these are reserved for tag attribute values.
`&copy;`	©	The copyright symbol. Just one example of many of the symbols available.
`&eacute;`	é	The French "e accent égu." Just one example of many of the diacritic characters available.

Creating an HTML File from Scratch

Creating a brand new HTML file from scratch using the College's lab computers is a bit of a chore. These instructions will walk you through the entire process. The same steps should also work on your home computer, although things may look a little different.

1. To begin, open your My Documents folder (*your-account*'s Documents folder in the lab) by double-clicking the desktop icon:

2. In the Windows Explorer window that opens, right-click somewhere in the white space on the right-hand side of the window. From the pop-up menu that appears, choose **New** and then **Text Document** from the sub-menu.



3. A new file will appear with the name "`New Text Document.txt`". The file's name is selected (highlighted) so you can change it directly. Change the name of the file to something appropriate ("`page.html`" in this example), and replace the "`.txt`" extension with "`.html`".

4. If you clicked the screen somewhere else, the file name will no longer be selected. Select the file icon by clicking it once and then press the **F2** key to change its name.

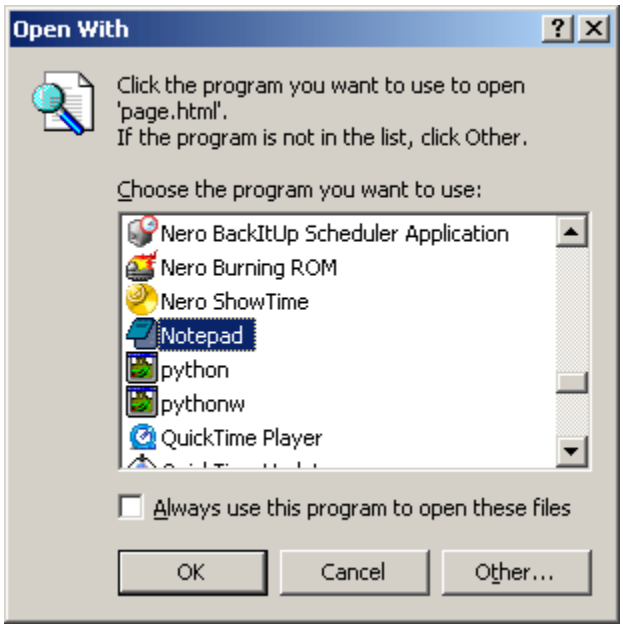
5. Press the **Enter** key to save the new file name. When you do, Windows will confirm the new file extension with a message box that reads: "If you change a filename extension, the file may become unusable. Are you sure you want to change it?" Click **Yes** because you definitely want to change the extension to "`.html`". The file icon now changes to show that it is an HTML document:

6. Note that if the default web browser is not Microsoft Internet Explorer, the icon may look different (e.g. an orange Godzilla-looking monster for Mozilla, or a whirl of flame for Firefox).

7. Double-click the new file icon to open the empty HTML document in the default web browser. The results are less than spectacular.

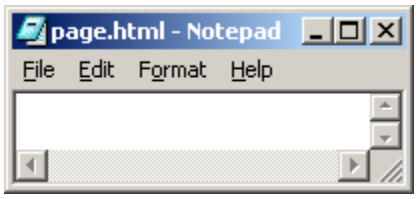
8. Return to Windows Explorer (where the HTML file icon was) and this time right-click the HTML file icon. If you are in one of the College labs, choose **Send To** from the pop-up menu that appears and then choose **Notepad** to directly open the HTML document in the Notepad editor, then skip ahead to step #8. Otherwise, choose **Open With...** from the pop-up menu that appears.

9. A dialog box called "Open With" will appear (it may take a couple of seconds). Scroll down the list of programs until you find Notepad and select it. Make sure that the **Always use this program to open these files** checkbox is *not checked*, and click the **OK** button.

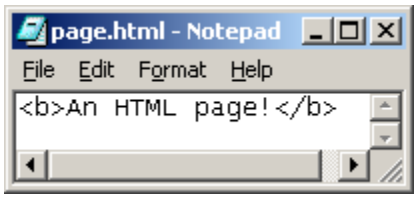


10. Until you log out of this lab computer, if you want to open an HTML document file from Windows Explorer, just right-click the file icon and choose **Open With** and then **Notepad** from the pop-up menu.

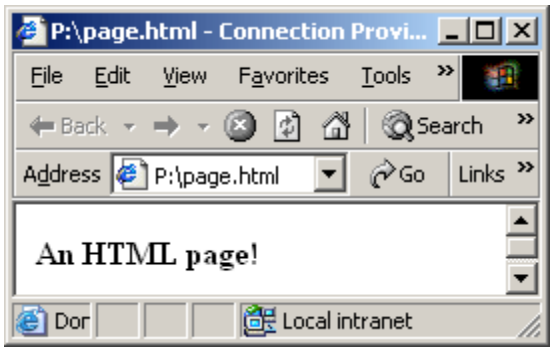
11. Now the empty HTML document file is open in the Notepad program.



12. You can now edit the HTML document in Notepad. For example:



13. When you switch to the browser program (perhaps by pressing **Alt+Tab**), you can view the changes to the HTML file by refreshing the browser's display (choose **Refresh** from the **View** menu, click the "recycle" icon on the toolbar, or press **F5**).



If you're planning on working with an existing HTML document, you can jump straight to step #5.

Computer Graphics and Multimedia

Video Display Hardware

Computers are almost always equipped with some form of **video adapter**. This may be a card that is plugged into the computer's motherboard, or it may be physically built in to the motherboard. Either way it will expose a socket that can be attached to a video monitor with the appropriate cable (usually attached to the monitor).

The video adapter's job is to store an in-memory representation of the currently displayed image on the video monitor. The adapter converts this representation to a signal understood by the monitor.

This in-memory representation of the display consists of a 2-dimensional matrix of dots, called **pixels** (for Picture Elements). The dimensions of this matrix can vary depending upon the characteristics of the video adapter and the monitor; the larger the matrix, the more detail can be displayed onscreen at any time.

The dimensions of this matrix is called the **display resolution**. Several resolutions are common: 640 pixels wide by 480 pixels high (once called VGA), 800 x 600, 1024 x 768, 1152 x 870, 1280 x 1024, 1600 x 1200.

There is one more "dimension" to this matrix: **colour depth**. Each pixel may be represented in memory as a number of bits. One bit can accommodate two colours: black and white. More colours require greater bit depths. Several colour depths are common: 1 (black and white), 4 (16 colours), 8 (256 colours), 16 (65,536 colours), 24 (~17 million colours, a.k.a. High Colour), and 32 (~4 billion colours, a.k.a. True Colour).

The amount of memory (in bits) required to store this matrix is the product of the height in pixels, the width in pixels, and the colour depth in bits. For example, an 800 x 600 black and white image will require 800 x 600 x 1 = 480,000 bits, or about 60,000 Bytes. That same image represented in True Colour will require 800 x 600 x 32 = 15,360,000 bits, or about 1,920,000 Bytes (almost 2 MegaBytes).

Video adapters are equipped with their own RAM chips to store these matrices. Adapters with more memory can represent higher resolutions and/or numbers of colours, than can comparable adapters with less memory.

Of course, just because an adapter supports higher resolutions, the video monitor may not (this is particularly true of thin-screen monitors). Higher resolutions and colour depth require more data to be transmitted from the adapter to the monitor. The adapter must also send this data to the monitor between 60 and 120 times per second (60-120Hz) to match the monitor's scan rate (how fast the electron beam zigzags its way down the phosphor screen). A higher frequency usually means less flicker (and less eyestrain) but not all monitors can support these high frequencies for all resolutions and colour depths (setting the adapter to too high a frequency can damage the monitor!). Thin-screen monitors, like the kind you find in laptop computers, function very differently than CRT (Cathode Ray Tube) monitors, and so frequency isn't as much of a concern for them.

Image Formats

The accompanying handout (Computer Image Formats) describes a few common bitmap and vector image formats.

A **bitmap** image--sometimes called a **raster** image--is made up of nothing but pixels. It is the format most commonly used for photographs or hand drawings. Bitmap images typically require lots of storage space (height x width x colour depth). Bitmap images also do not expand well: each pixel is stretched into larger and larger blocks, giving the overall image a "chunky" look (this is called **aliasing**).

Bitmap image formats often include **compression**. The details of this compression are complex, but the simple idea is just to record any repeating pattern once, along with the number of times it repeats. For example, a line of 57 blue pixels could be compressed by just storing the number 57 and 1 blue pixel.

Some compression techniques are **lossless**: the uncompressed image is identical to the original image down to the very last bit. **Lossy** compression can often produce more compact images, but at the expense of some visual detail. This is usually manifested as discoloured patches on the image.

Unlike Bitmap images, **vector** images are composed of simple instructions using drawing primitives: lines, arcs, polygons, text, etc. All that need be stored is the type of primitive, and a few extra characteristics: width, starting point, ending point, colour, texture, etc.

Vector image formats are therefore much more compact than bitmap formats, although they can't be used to express complex images such as photographs. However, vector images scale in size very well because they are not converted to pixels until immediately before they are drawn onscreen (which is always done using pixels).

Here is how the two format types compare:

	Bitmap	Vector
Scales (reaction to changes in size):	Appears "chunky" when enlarged.	Smooth.
File Size:	Large, often requiring compression.	Small.
Suited for:	Photographs, Paintings.	Drawings, Text.

Image Formats for the Web

Currently, the most common image formats used for the web are JPEG and GIF. PNG is the anticipated successor to both formats, but it's still not as popular as the others.

JPEG is best suited for photographs or drawings containing many colours.

GIF is best suited for icons or simple pictures that contain no more than 256 colours.

GIF images can also include a transparent colour, that lets whatever colour is behind the GIF show through.

Because JPEG uses a lossy compression algorithm (the amount of compression can be varied), some images may display patchy colours (JPEG artifacts). However, the payoff is that the images can usually be compressed much more than the comparable GIF.

Digital Audio

CD-quality audio is digitized by sampling the analog audio wave 44.1 thousand times per second and each sample is assigned an amplitude encoded into a 16-bit number. Stereo audio requires double the storage, so 1 second of CD-quality stereo audio takes: 44,100 * 16 * 2 = 172KB. That comes out to about 10MB per minute, meaning that most songs require between 20MB (The Ramones) and 170MB (Iron Butterfly).

Various compression techniques have arisen that are able to significantly reduce the size of audio files without seriously reducing the quality. One of the most popular at the moment is the MP3 format (MPEG III) that can achieve a variable compression ratio of between 5 and 20 (the most common ratio is 7, so a 4 minute song would require only 5.7MB rather than the 40MB needed by an uncompressed format.

Music compression is more complicated than the algorithms used in images. Typically, the compression algorithm removes parts of the audio signal that aren't audible to the human ear, or that don't drastically change the shape (timbre) of the sound. In this way, audio compression is a lossy type of compression since the original can't be perfectly reconstituted from the compressed form.

Digital Video

Digital video is essentially a sequence of bitmap images, displayed at a certain frequency. Most film is shot at 24 frames per second, so that's a common frequency.

Obviously, digital video requires a huge number of bits. For example, a 640 by 480 video display (the approximate resolution of a television) using 32-bit True Colour at 30 frames per second (TV's frequency) would require: 640 x 480 x 32 x 30 = 35MB. A 30-minute television program at that rate would consume about 62GB of storage.

By the way, that figure does not include the audio portion of the program.

Clearly, 35MB per second is far greater than the bandwidth of just about any consumer network technology in existence. Therefore, like images and audio, digital video is highly compressed using advanced lossy-type algorithms.

One of the things that can be done with video is to only store the parts of a single frame that are different from the previous frame. This so-called **psychovisual compression** closely matches the way humans watch motion pictures: most of our focus is on things that move. The other parts of the display can be much more compressed, but the loss of quality is not noticed as much.

Computer Image Formats

Bitmap Images

Bitmap images are made up of a fixed-size matrix of pixels (picture elements), of varying colour depths. The amount of video memory (bits) required to display a bitmap image is easy to calculate: width in pixels x height in pixels x colour depth. Many bitmap image formats compress this data by removing redundancy, although some formats lose information by doing so. Because the matrix of pixels is a fixed size, expanding a bitmap image will result in a "chunky" appearance as each square pixel grows in size. The following are some of the more common bitmap formats:

Name	Extension	Web?	Compression?	Notes
Graphics Interchange Format	.GIF	Yes	Lossless	GIF images can display a maximum of 256 colours, but can include a transparent colour. It is best used for icons and simple drawings.
Joint Photographic Experts Group (JPEG) File Interchange	.JPG .JPEG	Yes	Lossy	The JPEG format can support 24-bit colour, making it well suited for photographs. The format does use a variable but lossy compression technique, meaning that highly-compressed images are displayed with patchy discolourations.
Portable Network Graphics	.PNG	Yes	Lossless	This format is the intended successor for both GIF and JPEG and is gradually gaining in popularity. It uses a better lossless compression algorithm than GIF, and can display 32-bit colour, including transparencies.
Microsoft Windows Bitmap	.BMP	No	No	The Microsoft BMP format is commonly used for full-colour drawings or photographs on the Windows OS. It can't be used for the web, and since it is not a compressed format, tends to result in very large file sizes.
Tag Image File Format (TIFF)	.TIF .TIFF	No	Both	The TIFF format is commonly used by digital cameras and scanners. It is notable in that text "tags" can be added directly to the image file describing various aspects of the image.
RAW Image Format	*varies*	No	Lossless	The RAW format is actually a family of similar formats used primarily by digital cameras. It corresponds well to the way digital camera sensors capture light and colour and the images are usually of very high quality.

Vector Images

Vector images are made up of a set of drawing primitives: lines, arcs, text, polygons, etc. These primitives can be given colours and textures to produce very complex images. The size of a vector image file is typically much smaller than a bitmap file because only the attributes of the primitives are recorded: length, width, begin and end positions, colour, etc. Vector images also scale smoothly and do not have the same "chunky" behaviour of bitmap images. The following are some of the more common vector formats:

Format	File Extension	Web?	Notes
Encapsulated Postscript	.EPS	No	Postscript is primarily a printer typesetting language, but it is also used as a vector image format on UNIX systems.
Portable Document Format	.PDF	Yes (Acrobat Reader)	PDF (a.k.a. Adobe Acrobat) is really a page layout format, but under the covers it uses vector primitives.
Scalable Vector Graphics	.SVG	Someday	The SVG format is intended to introduce the flexibility and control of modern typesetting to the web in a browser-neutral fashion.

Managing Computer Images

There are whole professions built around creating, editing, and organizing computer-based images. An image may be a digitized photograph, a computer drawing, an icon, or what have you; most of the time these are bitmap, or raster, type formats. Yukon College offers courses that cover one of the popular image-manipulation tools, Adobe PhotoShop. For our purposes, we just want to be able to add images to our documents, whether they be HTML or Microsoft Word files, and also perform a few simple editing tasks.

One of the more burdensome tasks when dealing with images is translating them from one format to another.

Scanning Images

Before we can manipulate a printed image or a photograph, it must first be "digitized" (i.e. converted to a raster format computer file with height, width, and colour dimensions). A scanner is the device that can do this for us. Each of the labs in the College, except for A2301, has one scanner connected to the nearest computer.

Each model of scanner works slightly differently. The lab has handouts for the different types. Regardless of the model, the general sequence of events is as follows:

1. Turn on the scanner and log into the computer next to it.
2. Place the image or photograph you wish to scan face-down on the glass surface under the scanner's lid. Try not to touch the glass plate, as fingerprints and smudges will show up in the finished product.
3. Start a preview scan. The instructions for this vary depending on the brand and model of scanner, but at some point a program that controls the scanner will be run on the computer. The idea is that the scanner will quickly do a low-quality scan of the entire glass surface.
4. Select the area from the preview scan that you want to capture as a high-quality image. Usually, this involves drawing a rectangle around your image or photograph using the mouse. You can then instruct the scanning program to perform a high-quality scan of just your selection.
5. The scanning program can then either save the scanned image directly as a file, or load it into one of the image manipulation programs like Adobe Photoshop.
6. You can now do whatever you like with the scanned image: resize it, crop the edges, save it as a different image format, or whatever strikes your fancy.
7. Lastly, don't forget to remove your original from the scanner!

One thing you'll notice is that the scanned image seems to be much larger on the computer screen than the original. This is because the scanner's resolution is usually much higher than the computer's. One of the steps you'll likely take in a tool like Photoshop is to shrink the image (I find 40% of the scanned size is often about right).

You may also find that the default raster format for the scanned image is TIFF (file extension .TIF). TIFF is a popular raster format used by scanners and digital cameras. You can use Photoshop to convert it to one of the more common formats such as JPEG or PNG.

Manipulating Images in Adobe Photoshop

Adobe Photoshop[35] is *the* name in image-manipulation software tools. It is incredibly powerful, but also discombobulatingly complex. In this course, we'll only touch on a couple of its most simple features. Feel free to play with it though--it's installed on all of the lab computers.

In order to	Do this
View the Actual Size of an Image	Choose **Actual Pixels** from the **View** menu.
Resize an Image	Choose **Image Size** from the **Image** menu. You can then enter the new size as either a percentage of the current size, or as a specific number of pixels.
Crop an Image	Choose the selection tool (looks like a dashed rectangle) and draw the border of the portion of the image you want to keep. Then choose **Crop** from the **Image** menu.
Save an Image in a Web-Friendly Format (e.g. GIF, JPEG, PNG)	Choose **Save for Web** from the **File** menu. If you then choose the tab labelled **2-Up** you can see a side-by-side comparison of the original image and how it will appear in the selected web format. The controls on the right side of the screen let you pick the format and adjust the quality. Once you're satisfied with the results, you can click the **Save** button.

Inserting Images into HTML Documents

To insert an image into an HTML document, you just reference it using the `<img>` tag. The referenced image can be located in the same folder as the HTML file, or across the world at an Internet address.

To reference an image located somewhere on the Internet, simply type the image's full location in the `src` attribute of the `<img>` tag. (You can find the location of an image by right-clicking it in Internet Explorer and then choosing Properties from the pop-up menu. The Address field shows the full location of the image.) For example, if you want to display the Yukon Government's logo on your page:

```
<img src="http://www.gov.yk.ca/images/ytgcolor.gif">
```

Which, when rendered by a browser, looks like this:

If the image you want to display is located in the same folder as your HTML file, then all you have to put in the `src` attribute is the name of the file (without any of that "http://" gobbledygook):

```
<img src="my_image.jpg">
```

Consult the online HTML references, or the books on loan in the library, for the other attributes of the `<img>` tag: borders, alignment, sizing, alternate wording for people with text-only browsers, and so forth.

Inserting Images into Microsoft Word Documents

To insert an image into a Word document, first choose **Picture** from the **Insert** menu, and then choose the **From File** from the submenu. Locate the image you want to insert, and then click the **Insert** button.

You can also simply paste an image from the clipboard directly into a Word document.

---

[35] www.adobe.com/products/photoshop/main.html

The image will now be visible in the document and will move around in the text as you type. If you prefer to have the text flow around the image, right-click the image and choose **Format Picture** from the pop-up menu. Choose the **Layout** tab and then one of the "Wrapping Style" options, such as Square or Tight.

Click the **Advanced** button on that tab to view even more possibilities, and to precisely control the way the text will wrap around your image.

Microsoft Word will now display the page in "Print Layout" view mode, so that you can see how the text wraps around the image. You can also left-click and drag the image about the page until it's positioned just as you like. If you later return to the "Normal" view mode, the image will just be displayed alone on its own line, but returning to the "Print Layout" view mode will again show the proper positioning of the image and its surrounding text.

HTML Images and Colour

Images

Web images must use the GIF, JPEG, or PNG bitmap format. Once you have an image in one of those formats, you simply reference its file using the standalone `<img>` tag. The referenced image may be located in the same folder as the HTML file, or across the world at an Internet address.

To reference an image located somewhere on the Internet, simply type the image's full location in the `src` attribute of the `<img>` tag. (You can find the location of an image by right-clicking it in Internet Explorer and then choosing Properties from the pop-up menu. The Address field shows the full location of the image.) For example, if you want to display Mayo's logo:

```
<img src="http://www.yukonweb.com/community/mayo/logo.gif">
```

Which, when rendered by a browser, looks like this:

If the image you want to display is located in the same folder as your HTML file, then all you have to put in the `src` attribute is the name of the file (without any of that "http://" gobbledygook):

```
<img src="my_image_file.jpg">
```

You can also resize images in the browser by using either the `width` or `height` attributes of the `<img>`, or by using both together. *It's important to remember that changing an image's size using these attributes doesn't physically change the file size of the original image file. To put it another way, a very large 1 MegaByte-sized image file will always take the same long time to download over a dialup internet connection, even if shrunk down to a width of 1 pixel.*

For example, to shrink Mayo's logo to just 80 pixels high, you'd use this code:


```
<img src="http://www.yukonweb.com/community/mayo/logo.gif"
height="80">
```

If you use both the `width` and `height` attributes together, you can stretch an image out of shape.


```
<img src="http://www.yukonweb.com/community/mayo/logo.gif" width="40"
height="80">
```

You can add a border to your image with the `border` attribute. The value the attribute takes is the width of the border.


```
<img src="http://www.yukonweb.com/community/mayo/logo.gif"
height="80" border="4">
```

And you can turn your image into a hyperlink simply by "wrapping" it in an `<a>` tag. If you wish to remove the border, set the `border` value to 0.


```
<a href="http://www.yukonweb.com/community/mayo">
 <img src="http://www.yukonweb.com/community/mayo/logo.gif"
height="80" border="0">
</a>
```

Note that by hovering your mouse over the image, the cursor turns into a hand, indicating that it is a link.

You can also use a tiled image for the background of a web page. To do this, you specify the filename of the image in the `background` attribute of the document's `<body>` tag. Visit grsites.com/textures to find a tremendous assortment of images suitable for web page backgrounds.

Colour

Many of HTML's tags accept either a `color` or `bgcolor` attribute, meaning the foreground and background colour, respectively. Note that HTML uses the American spelling: "color".

Colours in HTML can be specified by name or by hexadecimal (base 16) number. You can find a complete list of HTML colour names and numbers in the online notes for this week. When using a hexadecimal number to specify a colour, don't forget to prefix it with "#".

For example, to set the background colour of an entire web page to light blue, simply set the `bgcolor` attribute of the `<body>` tag as follows:

```
<body bgcolor="lightblue">
```

And to insert a reddish-purple horizontal rule on the page:

```
<hr color="#FF0099">
```

If you want to change the **colour** of text, use the `color` attribute of the `<font>` tag:

```
<font color="indigo">This is <b>Indigo</b>.</font>
```

This is **Indigo**.

There are many courses that will teach you the appropriate uses of colour. This isn't one of them, so feel free to go bananas.

Introduction to Microsoft Word

Microsoft Word is *the* dominant word processor--both on the Windows and Macintosh platforms--owning somewhere north of 90% of the market. In the lab, we'll use Word 2003, but if you use an older (e.g., 6.0, 95, 97, XP) version elsewhere (or the newer Word 2007 or 2008 for the Mac), the principles and techniques covered here will be similar, if not identical.

So far in class we've just used text editors (e.g. Notepad) rather than true word processors. The dividing line between the two is hard to nail down, but in general, the word processors of today make full use of the What You See Is What You Get (WYSIWYG, pronounced "Whizeewig") metaphor. In other words, word processors combine the content of the document (the words, sentences, and paragraphs) together with the formatting of the document (fonts, layout, headers & footers, etc.). Text editors just handle the content part (languages such as HTML can be used in a text document to spruce up its format, but there's no semblance of WYSIWYG in Notepad).

Document Views

When you first open an existing or new document in Word, it is displayed in one of five views: Normal, Web Layout, Print Layout, Reading Layout, or Outline. You may change the view by selecting one of the five options from the **View** menu, or by clicking one of the five very small buttons on the bottom-left of the screen. If you can't tell which of the small buttons corresponds to which view by the button icon alone, just "hover" the mouse pointer over the buttons for a moment and a little message (called a **tooltip**) describing the button will appear.

The five views do not change the contents of the document, but are instead used to preview either its appearance or its organization:

- The Normal View is the best for editing documents (although any of the views can be used for editing) in that it hides some of the more distracting WYSIWYG elements such as: multiple columns, some embedded images, complicated wrapping of text around tables or images, headers & footers, etc. All of the hidden formatting is still in place--it's just not displayed.
- The Web Layout View shows how the document would look if it was converted to the HTML format.
- The Print Layout View is the true WYSIWYG display of the document, exactly how it will appear when printed. In fact, whenever you add a special layout feature that can't be displayed in the Normal View, Word will automatically switch to Print Layout View so that you can see the precise effect of your layout change.
- The Reading Layout View maximizes the onscreen legibility of a document and also adds a couple of toolbars to the screen to make it easier to move around in the document.
- The Outline View is different from the others in that it displays the overall organization of your document, broken down by the various headings you have used (see the Numbering section below for more about headings). You may expand or collapse these heading sections to display sub-headings or the actual content, as well as reorganize your document by dragging and dropping whole sections from one place to another.

Although unrelated to the document views, a handy feature of Word's will display special symbols for the otherwise hidden paragraph marks, spaces, tabs, and so forth. Displaying these symbols does not change any of the content in your document, and they can be displayed in any of the five document views. To display these hidden characters, click the button on the Standard toolbar that looks like a backwards "P" (Â¶, technically called a **pilcrow** and used by publishers to indicate the end of a paragraph). Clicking that button again will hide the special symbols. If the button isn't visible, you can also show the symbols by choosing **Options** from the **Tools** menu, choose the **View** tab, and then select one or more of the checkboxes in the **Formatting Marks** section.

Word Document Structure

At its simplest, a Word document is composed of sections which are in turn composed of paragraphs. Note that Word documents are not organized at all by page.

By default, a new Word document contains only a single section. Whenever you need to change the overall layout of a portion of your document, you will need to create a new section to hold that new layout. The types of layout adjustments that require their own section are as follows: margins, paper size or orientation, paper source for a printer, page borders, vertical alignment, headers and footers, columns, page numbering, line numbering, and footnotes and endnotes. Sections are delimited by one of three types of Section Breaks: Continuous, Next Page, or Odd/Even Page. A continuous section break allows two sections to occupy the same page. The other two breaks force the section following the break to begin on either the next page, or on the next odd or even page (e.g. each chapter in a document has its own section, and the first page of a chapter must always begin on an odd-numbered page). In some cases, Word will automatically insert the section breaks on your behalf. The section breaks are only visible in the Normal View. To insert a section break, choose **Break** from the **Insert** menu and then pick the appropriate type of break.

The paragraph is Word's fundamental unit of text content. Whenever you press the **Enter** or **Return** key on the keyboard, you are signaling the end of one paragraph and the beginning of another (Â¶). Therefore, anything that is followed by a paragraph mark is considered a paragraph: titles, headings, single lines, or traditional paragraphs (a series of sentences expressing a single idea). As we'll see, most formatting when using styles is done at the paragraph level.

Formatting Styles

Word makes it easy to format you document piece by piece: select a section of text, change the font, make it bold, increase the vertical spacing between it and the next line, and so forth. But instead of making these formatting adjustments one at a time, we can take advantage of Word's Style feature. A style is just a bundle of formatting characteristics that can be applied all at once to a section of text (sometimes just a few characters, but usually to whole paragraphs). Best of all, once the sections of our document have been assigned styles (by default, the whole document gets the Normal style), we can change a style's formatting characteristics and every single section of the document that uses that style will instantly change to match.

When beginning a new document, Word will also allow you to choose an overall "theme" for your document. These themes are created from something called a Style Template: essentially a list of pre-fabricated styles that together give any document that uses them a certain look. You can also create your own style templates so that all of your business letters, for example, based on that template share the same appearance.

To define new styles, or change existing ones, choose **Styles and Formatting** from the **Format** menu.

Paragraph vs. Character Styles

There are two main types of styles: paragraph and character. (We won't cover the new table and list styles in this course.) A paragraph style applies to an entire paragraph (i.e. everything up to and including the Â¶ mark). A character style applies to any text selection (e.g. a letter, a word, many words).

To apply a paragraph style, simply place the caret (the flashing vertical cursor-thing) anywhere in the paragraph (you do not have to select the whole paragraph) and then choose the desired style from the dropdown list on the left-hand side of the Formatting toolbar, or from the Styles and Formatting sidebar, if visible. To view the Styles and Formatting sidebar, choose **Styles and Formatting** from the **Format** menu. In either case, paragraph styles are identified by Â¶ and character styles by **a**.

To apply a character style, the procedure is much the same, but you must first select the portion of text to receive the new style.

Format Options

A style's formatting options are divided into several sections: Font, Paragraph, Tabs, Border, Language, Frame, and Numbering. Of these, we'll cover Font, Paragraph, Tabs, and Numbering. Note that, of these four, only the Font section applies to character styles. The settings for each section apply to the *entire* style. In other words, a style can't have one half appear in boldface with extra vertical spacing and the other half in inch-high purple italics--two separate styles are required to pull off that feat.

Font

As expected, the font section allows you to choose the font for your style: typeface (e.g. Arial, Times New Roman, Courier New), version (e.g. normal, italic, bold), point size, colour, underline, super/subscript, small caps, and so forth. You may also choose the horizontal spacing between characters (called "kerning"), and any number of annoying special effects.

Paragraph

The paragraph section describes both the paragraph's overall shape, and its relation to adjoining paragraphs. The shape of the paragraph is defined by its horizontal alignment (e.g. left, centered, right, full justified), its line spacing (e.g. single, double), and whether the paragraph should be indented (either the entire paragraph or just the first line). The paragraph's vertical distance from both the previous and following paragraphs can be set, as can its association with the following paragraphs (e.g. titles should "stick" to the following paragraph and not be stranded at the bottom of a page). You may also specify that paragraphs of this style must appear at the top of the page, or that they are not allowed to straddle pages (if the paragraph would partially spill over onto the next page, the whole paragraph is instead moved to the top of the following page).

Tabs

Just as on a typewriter (!), you may set specific tab stops for a paragraph style. By default, tabs are spaced every half-inch. You may also set left-, center-, or right-aligned tab stops, with or without leaders (e.g. .....1).

To view the current tab stops for any selected paragraph, choose **Ruler** from the **View** menu.

Numbering

Numbering is likely the most perplexing of all of the style sections. In essence, it controls the format of lists: bulleted lists, numbered lists, and numbered headings. The first two are fairly self explanatory, but the latter ain't.

Word pre-defines a number of styles. Of these, nine are used for headings: **Heading 1** through **Heading 9**. When you begin a new document, usually only the first three of these are listed in the style dropdown box on the Formatting toolbar or in the Styles and Formatting sidebar. Should you choose to organize your document in some hierarchical fashion (e.g. parts, chapters, sections, subsections), these heading styles will come in handy: use **Heading 1** for the top of the hierarchy (e.g. parts), **Heading 2** for the next level down, and so on.

Once your headings are in place, you may want to change the numbering characteristics of these so that parts are "numbered" starting with "A", chapters starting at "1", sections starting at "a", and subsections starting at "i" (roman numerals). Or, you may want to use a legal style: parts start at "1", chapters at "1.1", sections at "1.1.1", and subsections at "1.1.1.1". The magic happens when you decide to reorganize your document (through the Outline View or otherwise): all of the headings will automatically renumber themselves. And you are also free to change the numbering style at any time, or to dispense with it altogether, and the headings in your document will immediately reflect your change of mind.

A Field Guide to Word's Style Screens

Setting up and changing styles in Microsoft Word involves navigating through a hodgepodge of window screens (a.k.a. "forms" or "dialog boxes"). It'll take some practice before you're familiar with all of these, but this guide should at least help explain the purpose of many of these screens.

Styles and Formatting Sidebar

To display the sidebar, choose **Styles and Formatting** from the **Format** menu. The contents of the sidebar will vary depending upon the selection shown in the **Show** listbox at the the bottom of the sidebar:

- Available formatting: all styles and manual formatting changes that *can be used* in the document. For a new document, the default list of styles is Normal, Heading 1, Heading 2, and Heading 3.
- Formatting in use: all styles and manual formatting changes that *have been used* in the document.
- Available styles: all styles--but not manual formatting changes--that can be used in the document.
- All styles: all styles that come built in to Word.
- Custom: you can customize what appears in the sidebar by choosing Custom, but we'll leave that one alone.

To apply a style to the current paragraph or selection, click the style name.

To create a new style, click the **New Style** button at the top-right of the sidebar. To modify an existing style, click the downward-pointing arrow that appears to the right of the style name when you hover the mouse cursor over it, and choose **Modify** from the pop-up menu. The pop-up menu will also let you select every paragraph of that style, or delete the style entirely.

New Style and Modify Style Screens

The New Style and Modify Style screens are essentially the same. The New Style screen allows you to create a new style based on an existing style (usually Normal), and the Modify Style screen allows you to change an existing style. Click the **Format** button at the bottom-left of the screen to change a style's characteristics.

The **Based on** list shows the style on which the current style is based; the current style will "inherit" all of the properties of the other style, selectively overriding some of them. The **Style for following paragraph** list is used to choose the style that will automatically be applied to the subsequent paragraph after you type Enter (i.e. after typing in a heading and pressing Enter, the style changes to Normal for the next paragraph).

You may also select a shortcut key for the current style by choosing **Shortcut key** from the menu that appears when you click the **Format** button. This is very handy if you need to apply a certain style to many paragraphs spread throughout the document. A common convention is to use Shift+Ctrl+*digit* to apply the Heading *digit* style to a paragraph (i.e. pressing Shift+Ctrl+2 together applies the Heading 2 style).

Font

You reach the Font screen by clicking the **Format** button on the New/Modify Style screen and then choosing **Font** from the pop-up menu. The Font screen allows you to select a font family (e.g. Arial, Courier, Times New Roman), font styles (e.g. regular, italic, bold), sizes, colours, special effects (e.g. underline, superscript, shadow, small caps), horizontal and vertical spacing, and a variety of annoying animated effects.

Paragraph

You reach the Paragraph screen by clicking the **Format** button on the New/Modify Style screen and then choosing **Paragraph** from the pop-up menu. The Paragraph screen's **Indents and Spacing** tab permits you to control the "shape" of the style's paragraphs: horizontal alignment, left and right indentation, special indentations (e.g. first line only, or hanging indents), the vertical spacing before and after each paragraph, and the overall line spacing (e.g. single, double).

The **Line and Page Breaks** tab allows you to control the "behaviour" of the paragraphs. **Widow/Orphan control** prevents single lines from being abandoned at either the top (orphans) or the bottom (widows) of the page. **Keep lines together** prevents a paragraph from being split across a page break (if possible). **Keep with next** ensures that paragraphs of this style are always "glued" to the paragraph that follows (good for headings). **Page break before** inserts a manual page break before each and every paragraph of this style (useful for top-level headings).

Tabs

You reach the Tabs screen by clicking the **Format** button on the New/Modify Style screen and then choosing **Tabs** from the pop-up menu. The Tabs screen allows you to set the horizontal tab stops for this style. Tabs can be aligned in a variety of ways at that position: left (default), right, center, decimal (on the decimal point of a number), or bar (inserts a vertical line at the tab stop--who knows why). You can also set leaders (repeated characters that fill up the horizontal space of the tab character) for fancy table of contents-style formatting.

To set a tab, type the distance of the tab stop from the left margin (usually expressed in inches), select the alignment and leader, and press the **Set** button. Select an existing tab stop and click the **Clear** button to remove that tab stop, or click **Clear All** to remove all tabs. If tab stops are not defined, Word simply uses left-aligned leaderless tab stops spaced every half-inch.

Borders and Shading

You reach the Borders and Shading screen by clicking the **Format** button on the New/Modify Style screen and then choosing **Border** from the pop-up menu. This screen allows you to specify a border for this style's paragraph (on the **Borders** tab) or a background shade or pattern (**Shading** tab). A widely-held design belief is that these sorts of special effects should be kept to an absolute minimum.

Bullets and Numbering

You reach the Bullets and Numbering screen by clicking the **Format** button on the New/Modify Style screen and then choosing **Numbering** from the pop-up menu. The **Bulleted** tab lets you select a type of bullet shape for an unordered list. If you first select a bullet type and then click the **Customize** button, you can choose any character from any font to be the bullet shape (try the Wingdings font).

The **Numbered** tab lets you select a type of numbering for an ordered list (e.g. step-by-step instructions). Click the **Customize** button to change the digits used for the list or to change the starting point (the first number to use).

The **Outline Numbered** tab lets you choose a numbering scheme for Word's built-in Heading *X* styles. Choosing one of these options for a heading style often mangles any previous formatting selections, but these can be restored (therefore, it's a good idea to set up outline numbering early on in the process). By clicking the **Customize** button, you can tailor the numbering scheme to fit your requirements, but be warned that doing so can be quite a mind-boggling operation--better to use the default options if you can.

Word Objects

Aside from formatted text, word processor applications like Microsoft Word can contain many other types of content: images, text boxes, charts & graphs, tables, and even audio files and videos. Managing these different types of contents, or "objects" as Word calls them, is very similar--mostly you're just concerned about their placement on the page and how they interact with the surrounding text.

In-Line versus Floating Objects

When you first insert an object into Word (either through the clipboard or via a menu selection), it is usually inserted "in-line" (wherever the caret, or flashing I-bar is located). This means that it behaves very much like any other character you type at the keyboard. It can even be formatted using Word styles or manually (pretty much everything but the Font and Language formatting options apply).

However, if you want the object to behave independently of the surrounding text (perhaps you want the text to flow around the object), then you'll need a "floating" object. To make an object float, right-click it and then choose the menu item that allows you to change its properties (the menu item's name will vary depending upon the type of object: "Format Picture", "Format Text Box", "Format Object", or "Table Properties"). Then choose the **Layout** tab and select any wrapping style other than **In line with text**.

Once an object has a floating layout, it will not be visible in Word's Normal View. To see the object, make sure you choose **Print Layout** from the **View** menu.

Object Specifics

Images

To insert an image, choose **Picture** from the **Insert** menu and then choose the appropriate item from the submenu. If your image file is on disk, choose **From File** from the submenu. You can also paste an image from the clipboard directly into Word. Word understands most bitmap image formats.

By default, a new image will be inserted "in-line" with the surrounding text (wherever your caret was). To make the image float, right-click the image and choose **Format Picture** from the pop-up menu. Then choose the **Layout** tab and select any **Wrapping Style** other than **In line with text**.

Image objects are stored as part of the Word document, so you do not need to keep the original file together with Word's document file (.doc).

Text Boxes

To insert a text box (a piece of text that is separate from the main document), choose **Text Box** from the **Insert** menu. *If a large box labelled "Create your drawing here" appears, ignore it and draw your text box normally where you want it to go. It will not appear again.* Your cursor will now change to a cross-hair icon. Click and hold the left mouse button and drag out the area for your text box.

Text boxes are always inserted as floating objects (an in-line text box would simply be normal paragraph text).

Whenever manipulating text boxes, make sure your cursor is over the shaded border of the text box before you begin an action (like right-clicking to choose the **Format Text Box** menu item). The cursor will change from an I-beam to a combination of a small pointer and little compass arrows when the mouse is over the border.

Tables

To insert a new table, you have two options: draw it by hand, or define its parameters through a screen.

If you have a very irregular table in mind, then draw it by hand by choosing **Draw Table** from the **Table** menu. Your cursor will change to a small pencil icon which you can use to drag out the outline of the table and then draw in the lines to divide rows or columns.

To insert a more regular table, choose **Insert** from the **Table** menu, and then choose **Table**. A screen will appear that lets you select the number of columns and rows for the new table, together with a number of other formatting options. Once you click **OK**, the new table will be inserted.

By default, a new table will be inserted in-line with the surrounding text (wherever the caret was). To make the table float, click anywhere in the table and choose **Select** from the **Table** menu, and then choose **Table**. You can also click the small square (with compass-style arrows inside) that appears at the top left of the table if your mouse cursor is somewhere on top of the table. Then right-click the selected table and choose **Table Properties** from the pop-up menu. From the **Table** tab, you can now select a column alignment for your table, as well as a **Text wrapping** style.

Objects from other Applications

The most common objects that you're likely to insert from other applications are Microsoft Excel spreadsheet tables and charts, but pretty much anything that can be put in the clipboard can be inserted into Word. To insert such an object, copy or cut the object from the originating application so that it is in the clipboard, and then paste it into your Word document.

By default, a new object will be inserted "in-line" with the surrounding text (wherever your caret was). To make the object float, right-click the image and choose **Format Object** from the pop-up menu. Then choose the **Layout** tab and select any **Wrapping Style** other than **In line with text**.

Objects from other applications are stored as part of the Word document, so you do not need to keep the original file together with Word's document file (.doc). It is possible to insert objects so that they will automatically update if the original application changes the underlying data. To see how this is done, consult the Word Help article called "About linked and embedded objects".

Heading and Page Numbering in Word

Microsoft Word offers a few special features that will save you a great deal of time and effort if you are working with documents with numbered chapters, sections, subsections, and so forth.

Heading Style Numbering

As we've seen, Word has built-in heading styles called Heading 1 through Heading 9 (usually only levels 1 through 3 or 4 are needed, but you have 9 if you need them). These styles are composed of more than just big and bold fonts; they can also be used to organize your document using some system of numbering.

For instance, legal documents often begin with section 1, under which is section 1.1, 1.2, and then under that is 1.2.1, 1.2.2, and so on deeper and deeper until you see things like section 3.5.7.14.2.9. Textbooks often use a mixed Roman-letter-Arabic system with chapters numbered by capital Roman numerals (e.g. I, II, III, IV), subchapters by capital letter (e.g. A, B, C), sub-subchapters by Arabic numerals (e.g. 1, 2, 3), and so on until you see things like section VI.F.13.ix.q.

Word lets you number your document automatically by manipulating the Heading *X* styles. To apply numbering, choose **Numbering** from the **Format** pop-up menu on the **Modify Style** page. On the Bullets and Numbering screen, choose the Outline Numbered tab and then select from any of the options that name the heading styles in the preview window (these will automatically apply the appropriate numbering to all of the Heading *X* styles at once). You can control the specific numbering system by clicking the **Customize** button.

Once these styles are applied, Word will automatically keep the numbering in the correct order as you change the content of your document or shuffle the order of the existing sections. Remember that the Outline view can make reorganizing your document a snap.

If you're going to use this technique, it's better to do it earlier than later because applying outline styles changes all of the Heading *X* styles at once. You can always reformat individual heading styles afterwards.

Page Numbering

Word calculates page breaks on its own (keeping in mind any style settings that force page breaks, or manually-inserted page breaks). Therefore Word also keeps track of page numbers on its own. Because of this behaviour, it's never a good idea to refer to an absolute page number within your document--it may be incorrect the next time the document is printed (look up Word's cross-reference feature if you must refer to page numbers).

To show the page numbers on a page, choose **Page Numbers** from the **Insert** menu. A screen will appear that allows you to select the place on the page to display the number, and you can also click the **Format** button to make changes to the manner in which pages are numbered.

A document section must have a single page numbering style, so if you wish to use two different numbering styles (e.g. "i, ii, iii" for the introductory section, and "1, 2, 3" for the main body of the document), you must divide your document into at two sections by inserting a section break (choose **Break** from the **Insert** menu).

To change the page numbering format for a document or section, pretend there is no page numbering already and choose **Page Numbers** from the **Insert** menu again. Instead of inserting new page numbers, Word will change the existing ones to match your preferences. The change will only apply to the section where your caret (insertion point) is currently positioned.

Inserting a Table of Contents

A Table of Contents (TOC) is where heading style numbers and page numbers collide. Keeping a manually-created TOC synchronized with your document is a horribly tedious and error-prone task. Fortunately, Word will do all of the work for you, provided you've used the Heading *X* styles to structure your document.

To insert a Table of Contents, place the caret where you wish the table to be inserted and choose **Reference** from the **Insert** menu, and then choose **Index and Tables** from the sub-menu. On the Index and Tables screen, choose the **Table of Contents** tab. The **Show levels** box allows you to choose how many levels (starting at Heading 1) you wish to display in the TOC.

You can choose from a variety of formats for your TOC from the **Formats** list. If you choose the "From template" option, you may edit the individual styles controlling the TOC by clicking the **Modify** button. The TOC *X* styles can be modified just as any other style, although they never show up in the Styles and Formatting sidebar. (Earlier versions of Word don't make this distinction.)

Once the TOC is inserted in your document, it will not update automatically until you either print your document or save it and then open it again. To manually update the TOC, select the entire document (Ctrl+A) and press the refresh key (F9). You will be asked if you wish to update just the page numbers or the entire table. Selecting the former is acceptable (and quicker) if you didn't make changes to any of the document's headings, but the latter is the safer choice, particularly if you can't remember whether any of the headings have been changed.

To make changes to the TOC, click somewhere in it and choose **Reference** from the **Insert** menu, and then choose **Index and Tables**. After you make your changes, you will be asked whether you wish to replace the current TOC.

Spreadsheets

Spreadsheet History

In the days of yore, ere the arrival of yon computerized spreadsheet, there were General Ledgers. Ledgers were just big pieces of paper with a grid of lines that could be used to record accounting figures. Large companies (e.g. banks, insurance companies, IBM) had mainframe programs that could do these sorts of calculations, but these behemoths only allowed certain kinds of input, only displayed textual reports on reams of tractor-fed paper, and required work by a programmer if any of the business rules changed or a new report was required. Smaller businesses kept Ledger books.

In 1978, VisiCalc for the Apple II personal computer was released. It had pretty much all the elements of today's spreadsheet applications except for charting capabilities. Business people realized the utility of the spreadsheet immediately, and so it became the first "killer app" (an application program so desirable that it alone is the reason for the purchase of a computer system--email could be considered the second true killer app).

The feature that set VisiCalc apart from a simple computerized Ledger was its ability to automatically recalculate totals if any figure on the page changed. This enabled true "What If?" analysis for the first time: "if we sell 5,000 more widgets this quarter, how will our profit margin change?", and so forth.

Lotus 1-2-3 arrived in 1983 and, together with the new IBM PC, took over the market. (Lotus eventually bought out VisiCalc and shut it down.) Lotus 1-2-3 added database functionality (2) and simple charting (3) to the spreadsheet (1), but the basic spreadsheet functionality was no different than VisiCalc.

In 1985, Microsoft released Excel for the Macintosh, but didn't release a DOS version until 1987. By the mid-1990s, Excel combined with Windows had pretty much taken over all of the market and now Lotus is a tiny subsidiary of IBM.

Terminology

Cell: a single value in the spreadsheet at the intersection of a row and a column. Cells are named using both row and column (e.g. A1, G456). A cell may contain either a label (text, up to 32KB), a number (literal number or date), or a formula.

Row: a horizontal sequence of cells. Rows are numbered starting with 1 through 65536.

Column: a vertical sequence of cells. Columns are lettered starting with A through IV.

Range: a rectangular group of cells. Ranges are delimited by the top-left and bottom-right cells, and are specified by naming the top-left cell, a colon, and then the bottom-right cell (e.g. B4:F16).

Worksheet: the 2-dimensional grid of cells from A1 to IV65536.

Workbook: a set of worksheets (by default, 3) that are saved together as a single .XLS file.

Name Box: located at the top-left of the screen, the Name Box displays the name (column/row) of the selected cell. You may also name a cell or range by first making a selection, and then typing in the name inside the Name Box.

Formula Bar: the textbox at the top of the screen, to the right of the Name Box. Cell values may be changed in this space, or directly in the cell itself. Selecting a cell with a formula will display the formula in the Formula Bar, while the numeric result of the formula is displayed in the cell.

Navigation

Arrow keys, PageUp and PageDown work as expected.

Enter: saves any changes to the current cell and moves down one row.

Home: moves to the first column of a row.

Ctrl+Home: moves to cell A1.

Ctrl+End: moves to the bottom-right cell of the current work area.

Ctrl+PageUp, Ctrl+PageDown: moves to the previous, or next, worksheet in the workbook, respectively.

End, arrow keys: (press and release the End key and then press one of the arrow keys) moves as far as possible in the selected direction until either an empty cell is reached, or the first non-empty cell is reached. Holding the Ctrl key and then pressing one of the arrow keys does the same thing.

Ctrl+G (or choose Go To from the Edit menu): moves to the specified cell, or to a named cell or range chosen from a list. You may also move to named cells or ranges by selecting the name from the Name Box dropdown list.

Selection

Cell: click on the cell, or use one of the navigation techniques to move to the cell. A thick black border highlights the currently-selected cell.

Range: click and drag to select a rectangular range of cells.

Row: click the row number (far left of screen) to select an entire row or rows.

Column: click the column letter (top of worksheet) to select an entire column or columns.

Worksheet: click the box at the intersection of the column and row headings to select the entire worksheet.

Contiguous selection: hold down the Shift key while navigating to select a contiguous range of cells.

Non-contiguous selection: hold down the Ctrl key and select multiple non-contiguous cells or ranges with the mouse.

Clipboard: individual cells or ranges can be cut, copied, and pasted just as usual. Excel will highlight a clipboard selection with a flashing marquee. Press the Esc[ape] key to hide the marquee. Note that when cutting a selection to clipboard, Excel does not actually cut the selection until it is pasted.

AutoFill: The selection border highlight has a small square at the bottom-right. Click and drag this square to copy formulas or labels, or automatically populate a range with a sequence of numbers or dates.

AutoSum: Select a range of cells with numbers and Excel will display their sum at the bottom-right of the window.

Editing: to edit the contents of the cell: double-click the cell and you can edit it directly, select the sell and then click in the Formula Bar to edit it there, or select the cell and press F2 to edit it directly.

To insert a new column (or row), select the column to the right (or below, for rows) of the new column you wish to insert. Right-click and choose Insert from the pop-up menu.

To delete a column (or row), select the column, right-click and choose Delete from the pop-up menu.

Formatting

You may format individual cells, ranges, rows, columns, or the entire worksheet.

To resize column widths (or row heights) click and drag the line separating column letter headings (or row numbers). Double-clicking this line will automatically resize the column (or

row) to fit the largest cell in that column (row). You can also choose Column (or Row) from the Format menu, and then choose Width (Height) to enter a numeric width (or height), or choose Autofit to resize the column (or row) to fit the largest cell in that column (row).

To format the appearance of any selection, first make the selection and then right-click and choose Format Cells from the pop-up menu (or choose Cells from the Format menu). You may then make changes to the: number or date style, cell alignment, font, border, background pattern, and so forth.

If you have a contiguous "table" of cells, you may select any cell in that table and then choose AutoFormat from the Format menu to apply a pre-defined style to the entire table.

Merge a range of cells into a single cell by selecting the range and clicking the Merge and Center button on the toolbar (a little "a" with arrows pointing left and right). To un-merge a cell, select it and right-click. Choose Format Cells from the pop-up menu and then the Alignment tab. Uncheck the Merge Cells checkbox.

Naming

You may use the Name Box to give your own meaningful name to a cell or range. This name may then be used in formulas which makes them much easier to read. Spaces are not allowed in cell or range names, so use the underscore (_) character to delimit words in your name.

You may name a cell or range simply by selecting it and then entering the new name in the Name Box. Make sure to press the Enter key when you're done or the name won't be stored.

To delete an existing name, choose Name from the Insert menu, and then choose Define. Select the name to delete from the list of defined names and click the Delete button.

Formulas (Formulæ?)

All formulas start with the equals sign (=).

You may use simple algebraic operators such as: + (plus), - (minus), / (divide), * (multiply). Normal rules of operator precedence apply (PEMDAS: Parentheses, Exponents, Multiplication & Division, Addition & Subtraction).

To reference a cell, use either its column/row designation, or a pre-defined name. When entering a formula, you may also use the mouse or the keyboard navigation techniques to select the cell or range you wish to use in your formula and the name will be entered when you press the Enter key, or type the next operator symbol.

To display all formulas in a worksheet, press Ctrl+` (the back-quote at the top-left of the keyboard). Press the same key combination to hide the formulas. Selecting a cell will also display its formula in the Formula Bar.

Handout: Common Excel Functions

When performing calculations in a spreadsheet, you must use a formula (begins with "="). Within that formula, you may use a combinations of cell references (e.g. A5, B7, or named cells), mathematical operators (+, -, *, /), as well as functions.

A function is a kind of pre-defined formula that calculates a value based on one or more inputs (called "arguments"). The type and number of arguments that a function will accept vary from function to function, and some functions don't require any arguments. The most common type of argument for Excel functions is a cell or range reference, but literal number values, formulas, or even other functions may also be used. Most functions also accept a list of these arguments (shown as "*list/range*") in the examples that follow.

To write a function, first write the function name (you don't have to use uppercase letters), followed by an opening parenthesis, the arguments to the function (separated by commas if there are more than one), and a closing parenthesis. For example, to use the `SUM()` function to calculate the total of the numbers in the cells C5 and D7, together with the numbers in the range G3:J8, together with value 47.19, you would write (note the equals sign that signifies that the cell is a formula):

```
=SUM(C5,D7,G3:J8,47.19)
```

In most cases the order of the arguments does not matter (the `IF()` function is one notable exception). And because functions themselves can be arguments to other functions, you can write some pretty complicated expressions. For instance, the following--unnecessarily complicated--formula is equivalent to the above example:

```
=47.19+SUM(G3:J8,SUM(C5, D7))
```

When in doubt, use the Paste Function feature of Excel (choose **Function** from the **Insert** menu). This feature will walk you through the steps required to use Excel's built-in functions (of which there are hundreds).

Function Name	Description
`SUM(`*list/range*`)`	Returns the added total (i.e. sum) of all of the arguments (skipping references to empty cells or cells containing labels). This is such a common formula that Excel includes a toolbar button called AutoSum (looks like a capital greek Sigma--kinda like a big zigzag "E") that, when you select a range of cells, will insert a SUM() function into the cell below that range, complete with arguments to add up all of the numbers in that range.
`AVERAGE(`*list/range*`)`	Returns the statistical mean (the sum of the values divided by the number of values) of all of the arguments (skipping references to empty cells or cells containing labels).
`COUNT(`*list/range*`)`	Returns the count of all of the cells referenced in the arguments (skipping references to empty cells or cells containing labels).
`COUNTA(`*list/range*`)`	Returns the count of all of the cells or values referenced in the arguments, including any non-empty cells (even those that contain labels).
`MIN(`*list/range*`)`	Returns the minimum value from the cells referenced in the arguments (skipping references to empty cells or cells containing labels).
`MAX(`*list/range*`)`	Returns the maximum value from the cells referenced in the arguments (skipping references to empty cells or cells containing labels).

IF(*logical_test*, *value_if_true*, *value_if_false*)	Evaluates the *logical_test* expression, and if the result is true, returns the *value_if_true* value, otherwise the *value_if_false* value. The *logical_test* expression usually contains one of Excel's comparison operators: = (equals), < (less than), > (greater than), <= (less than or equal to), >= (greater than or equal to), or <> (not equal to). Labels can be used in the *value_if_true* or *value_if_false* arguments if surrounded by double-quotes ("). For example, the following evaluates the contents of cell A5, presumably a student's mark, and interprets the results accordingly: =IF(A5>=50,"Passing Grade","Try Again")

Spreadsheet Tips

Conditional Formatting

Excel's Conditional Formatting feature allows you to apply certain formatting characteristics (e.g. fonts, colours, borders, shading) based on the current value of a particular cell. If that cell's value changes, then so might the format. This technique is useful for highlighting cells that meet a specified criteria (e.g. use a red font for all negative numbers, display all altitudes greater than 1,000m in italics and a blue background but use a green background and bold text for altitudes below sea-level).

To apply a conditional format to a cell or range of cells, follow these steps:

1. Select the cell or range of cells that you wish to format.
2. Choose **Conditional Formatting** from the **Format** menu.
3. Specify the condition to be met before the special format is applied. Use the "Cell Value Is" condition specifier. In the third (and sometimes fourth) box, type a number or date to use for the comparison (e.g. "Cell Value Is less than 0" or "Cell Value Is between 1000 and 1500").
4. Choose a format to apply if the condition is met (click the **Format** button). The screen will display a preview of your selected format.
5. You may add up to 2 more conditions (if more than one condition is met simultaneously, Excel chooses the first when applying formats) by clicking the **AddÂ»** button.
6. Click the **OK** button to apply the conditional formatting to the selected cell or range of cells.

The appropriate format will now be displayed depending upon the value of the cell. If the value changes, then so might the format.

Text File Import

It's often necessary to create spreadsheets from data stored in a text file. There are three common formats for such text files: comma-separated values, tab-delimited, and fixed-width. The data must be formatted in some way, so that the different fields (cells) in a row can be distinguished. For example, in English we use spaces to format, or "delimit", the words in a sentence. Ifwedidn'titwouldbeawfullyhardtoreadanything.

Format	Delimiter	File Extension	Import Difficulty	Description
Comma-Separated Values (CSV)	Comma	.CSV	Automatic	Most spreadsheets can read a CSV file directly without any human intervention, but they are difficult to make sense of when opened in a text editor (Notepad). Fields are also sometimes surrounded by double-quotes ("). The double-quotes are necessary in case the field value happens to contain a comma.
Tab-Delimited	Tab	.TAB, .TXT	Easy	Occasionally, human intervention is required to import a tab-delimited text file, but not often. Double-quotes aren't typically required in this format because it's rare to have a field that contains a tab character.
Fixed-Width, or Space-Delimited	One or more Spaces	.TXT, .PRN	Sometimes Tricky	This format vertically "lines up" the columns of data in an easy-to-read structure. Unfortunately, while easy for people to read, it's much harder for a computer spreadsheet to digest. Therefore, human intervention is usually required to specify the start or stop point for each column.

Excel's "Text Import Wizard" (which is automatically launched whenever you try to open a file that is not in Excel's native workbook format) is very good at recognizing these formats and guiding you through the conversion process.

Dates in Excel

To Excel, dates and times are merely special formats applied to numbers. For instance, the number 1 represents January 1, 1900 (so zero represents December 31, 1899). Adding 1 to any date effectively adds a single day, so the number 2 represents January 2, 1900. Hours, minutes, and seconds are represented by fractions: 32.5 is noon on February 1, 1900.

Most of the time, Excel automatically understands that a cell should be formatted as a date if you enter the value in a common date format (e.g. `2006-10-31`, `31-Oct-06`, `October 31, 2006`).

If you wish to display the current date, use the `TODAY()` function. If you want both the current date and time, use the `NOW()` function. Neither of these two functions accept any arguments between their parentheses, but you must still type the parentheses.

Handout: Charts

Charts, sometimes called graphs, are Excel's way of displaying numerical spreadsheet data in a graphical format. Excel has a bewildering variety of chart types, but the most common ones are probably familiar to you: pie charts, bar or column charts, line charts, scatter plots, and so forth.

Due to the flexibility of Excel's charting features, explaining all of the possibilities in one course, let alone a couple of lab sessions, is impossible. Instead, I encourage you to play with charts in the lab and on your own. You can create some pretty funky results if you put your mind to it.

Terminology

A few terms will help use interpret Excel's charting features.

Term	Definition
Axis	Except for pie charts, each chart type has two axes: an X or horizontal axis (Excel calls this the "Category Axis"), and a Y or vertical axis (Excel calls it the "Value Axis"). Often, the category axis represents the "things" that are measured, while the value axis represents the value of those measurements. Excel allows you to name the axes, change the measurement high and low points (scale), add a second vertical axis, and apply all sorts of formatting options.
Data Series	A data series is a list of values that are usually plotted against the value axis of a chart. A pie chart may only include a single data series, but other chart types may use two or more simultaneously.
Legend	A legend, usually drawn within a box to the right of the chart, identifies multiple data series in a chart so that the viewer can tell them apart (e.g. one series is coloured red, the next blue with white stripes, and so on).
Embedded Charts vs. Chart Sheets	When you create a new chart, you have the option of inserting (or "embedding") it into the same worksheet that holds the numerical data (an "embedded chart"), or inserting it as a new and completely separate worksheet, called a "chart sheet". The choice is entirely up to you, and at any point in the future you can convert an embedded chart into a chart sheet, or vice versa.

The Excel Chart Wizard

Excel will walk you through the procedure of creating a chart using a tool called the "Chart Wizard". In general, the procedure is as follows:

1. Select the data from your worksheet that you wish to chart. Don't include any title cells in your selection (titles are more easily added to the chart in a later step).
2. Click the Chart Wizard button on the toolbar (looks like a tiny bar chart) or choose **Chart** from the **Insert** menu. This will open the Chart Wizard screen.
3. Choose the chart type from the list, and then optionally the chart sub-type (remember that pie charts cannot display more than one data series). The **Custom Types** tab lists even more options. Click (and hold) the **Press and Hold to View Sample** button to get a quick preview of your chart. Click the **Next >** button to proceed to the next step.
4. If you selected your data in step 1, all of the Data Range and Series fields should be populated. If not, it's probably easier to click the **Cancel** button and start again with your selected data; this step in the process is very tricky to deal with manually. Click the **Next >** button to proceed to the next step.
5. You may now set a variety of chart options: Titles, Axes, Gridlines, Legend, Data Labels, and a Data Table. The screen shows a preview of your chart, so feel free to experiment with the settings. Click the **Next >** button to proceed to the next step.
6. Select whether you wish to insert the new chart as a new chart sheet, or embedded into an existing worksheet (labelled **As object in** on the screen). Click the **Finish** button to insert the chart into your workbook, either as an embedded chart or as a separate worksheet.

Data Series in Charts

Deciding which cells to select (the data series) before starting Excel's Chart Wizard can be a bit confusing at times. Whenever you're unsure, just select the range of cells that have value you want to plot against the Value axis (Y or vertical). If you're creating a pie chart, then these values will represent the proportional size of each pie slice. It may help to remember that the data series always consists of just numbers (never text labels, and only rarely dates).

Once you reach the second step if the Chart Wizard, you will need to select the range of cells that represent the Category axis (X or horizontal). Choose the **Series** tab, and click in the box labelled **Category (X) axis labels**. You can then select the range of cells in the spreadsheet that represent the labels that should appear across the bottom of the chart (or as the titles of each slice of a pie chart). The Category axis cells are often dates or non-numerical data (but not always). There should always be just as many cells in the Category axis range as there are values in the Value axis range (so that each data value has a category or label, or each Y has an X).

Chart Trend Lines

When displaying a data series (in a non-pie chart type) that has an obvious trend (i.e. over a period of time), you can add a "trend line" to the chart that more clearly demonstrates the progression.

To do this, right-click on one of the data series in the chart and choose **Add Trendline** from the pop-up menu. You will then be given a choice of different trend types. Choosing the most appropriate type is a bit of an art, but experiment with the options until you achieve the desired results. Once you click the **OK** button, Excel will display the new line on the chart (right-click the line to format or remove it).

The Can't-Fail Guide to Creating Excel Charts

Excel's Chart Wizard usually does the right thing. But sometimes you just need total control over how your chart--particularly one with multiple data series--is created. The following instructions should always work, for either simple or complex charts, at the expense of being a tad complicated.

1. Don't bother selecting any data. Click on an empty cell if you like.
2. Click the Chart Wizard button on the toolbar or choose **Chart** from the **Insert** menu.
3. In step 1 of the Wizard, choose the type of chart you like. Click **Next**.

In step 2 of the Wizard:

a) Completely delete the contents of the **Data Range** field visible on the **Data Range** tab.
b) Choose the **Series** tab.
c) Click the **Add** button to add a new data series to your chart.
d) Fill in the Name field with the name for your data series. You may also select the name from a cell on the worksheet.
e) In the **Values** field, delete the "`={1}`" that sometimes appears, and select the range containing your data series numbers. If the Wizard screen is obscuring the numbers, click the "Collapse Dialog"  button to the right of the field to shrink the Wizard screen to a single line. Once you've made your selection, click the "Restore Dialog"  button to the right of the field, or just press the Enter key.
f) In the **Category (X) axis labels** field, select the range containing your category labels, which must have as many elements as your data series range.
g) If you want to add another data series, just click the **Add** button again. You will need to fill in the **Name** and **Values** fields for each new data series, but the **Category (X) axis labels** field will remain the same for all data series. Each data series must have the same number of elements.

Once you've added all of your data series, click **Next**.

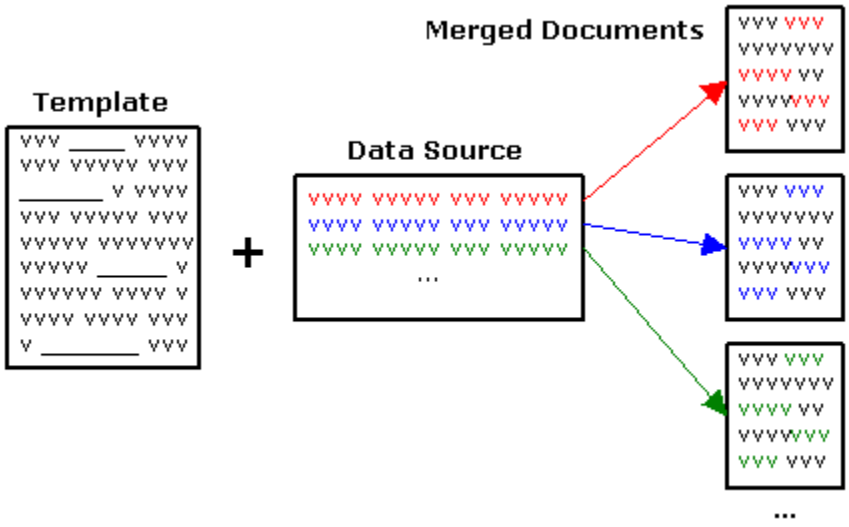
4. Step 3 of the Wizard is the Wild West. Fill in whatever options you think are appropriate and click **Next**.
5. In step 4, we usually want to insert the chart as a "new sheet". You can name the chart's tab by filling in the the **As new sheet** field. Click **Finish**.

Later on, if you should decide to add more data series to your chart, right-click on any of the data series in the chart and choose **Source Data** from the pop-up menu. Choose the **Series** tab to take you right back to complicated step 2 of the Chart Wizard where you can add or remove data series to your heart's content.

Mail Merge

Microsoft Word's Mail Merge feature is useful whenever you want to generate multiple copies of the same document with a few changes made to each copy. The most common use of this is to generate form letters or address labels, but Mail Merge is also often used to generate reports with complex formatting.

The overall Mail Merge process is shown in the following diagram. A template document, containing **merge fields** (placeholders for fields from the data source) is combined with a **data source** (e.g. database, spreadsheet, delimited text file, contact list) to produce multiple merged documents, one for each record in the data source.



Most of the template document's contents will be copied to each of the merged documents intact; only the merge fields will be replaced with the corresponding field from the data source records. Therefore, all of the customizations must appear in the data source (e.g. name, address, employer, personalized greeting).

With just a single template file, Mail Merge will produce as many merged documents as their are records in the data source: 10, 500, 20,000, whatever.

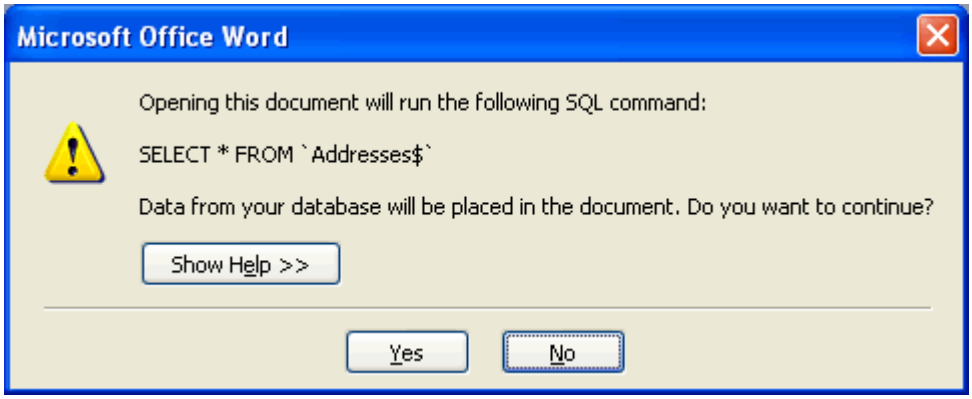
To begin, choose **Letters and Mailings** from the **Tools** menu, and then choose **Mail Merge** from the submenu. We'll assume that the template document is already written and is open in Word, and that a data source (perhaps an Excel spreadsheet) is standing by. The Mail Merge wizard (a sidebar) will now assist you through the 6-step process in Word 2003 (earlier versions used a more complicated 3-step process):

1. Select the **Letters** option from the **Select document type** list. Click **Next** to continue.
2. Select the **Use the current document** option from the **Select starting document** list. Click **Next**.
3. Select the **Use an existing list** option from the **Select recipients** list. Click **Browse**, and in the **Select Data Source** screen, find the data source file and click the **Open** button. Select the table from the data source you wish to use (the default usually is fine for a spreadsheet) and click the **OK** button. To merge using every record in the data source, click **OK** on the **Mail Merge Recipients** screen. You may also exclude certain recipients at this stage. Click **Next**.
4. Insert the appropriate merge fields into the template document by clicking the **More items** link in the sidebar's **Write your letter** section. You may also wish to have Word insert entire **Address block**s or **Greeting line**s. The merge fields will appear in the template document surrounded by Â«chevronÂ» symbols. At any time, you can preview

the merged documents by clicking the **Next** link (click **Previous** to return to step 4). Click **Next** once all of the fields have been inserted.

5. Preview the merged documents to make sure everything's okay. You may cycle through the merged documents using the arrow buttons in the **Preview your letters** section. If there are problems, return to step 4 by clicking the **Previous** link. Otherwise, click **Next**.
6. You may optionally click the **Edit individual letters** link to view the merged documents in a new Word file and make further changes to individual merged documents. Or, you may print the merged documents directly by clicking the **Print** link. In either case you will first be asked if you want to merge all of the records in a pop-up screen. Click **OK**.

At any point in the future, you can re-open the template document to make changes or run the Mail Merge process again. When you open the document, Word will display a message screen similar to the following:



This means that Word wants to re-attach the template document to the same data source that was used previously. Click **Yes** unless you want to use a different data source. If the data source file can't be found, Word will ask you to either locate it again, or to not bother with a data source for now.

You may then start the Mail Merge process over again--usually from step 3--by choosing **Letters and Mailings** from the **Tools** menu, and then choosing **Mail Merge** from the submenu.

Exercise: Working with Files and Folders

In-class exercises give you a chance to practise new techniques. They are not marked. I'll go over some of the answers in the next class, but feel free to ask if you're on the right track at any time.

1. Take a look at all of the logical drives to which you have access. Which ones are physically located on the lab computer, and which are somewhere on the network?
2. How much space remaining do you have for your documents, e-mail messages, and personalized settings? *Hint: all of this is stored on the `P:` drive.*
3. Create a new folder under My Documents called `cpsc100`. Note that this will show up in both your My Documents and `P:\My Documents` folders at the same time.
4. Create a new folder within the `cpsc100` folder named "`storage`".
5. Copy the contents of the `S:\cpsc100\storage` folder to your new `storage` folder.
6. What happens if you try to create a new file from scratch in the `S:\cpsc100` folder? What system is protecting that folder?
7. Try to create another folder in your `cpsc100` folder that is also named "`storage`". Why can't you do that?
8. Copy the complete `S:\cpsc100\editing` folder to your `cpsc100` folder.
9. Rename the file named "`1984.txt`" in your `editing` folder to "`1984.html`". What happened to the icon? What happens when you double-click the `1984.html` file? Try renaming it to "`1984.exe`". What happens when you double-click it this time?
10. Delete the `1984.exe` file to the Recycle Bin.
11. Recover the `1984.exe` file from the Recycle Bin and change its name back to "`1984.txt`".
12. Create a shortcut to the `S:\cpsc100` folder on your desktop. Rename the shortcut, "CPSC100 Course Files".
13. Delete the complete contents of your `cpsc100` folder (but not the folder itself) without using the Recycle Bin. Verify that the deleted folders are not in the Recycle Bin.
14. Open the new `CPSC100 Course Files` shortcut and copy everything there to the `cpsc100` folder under My Documents.
15. Compare the contents of the `P:\Desktop` folder with the items that appear on your Windows desktop. What is the Windows desktop, really?

By the end of all of this, you'll be left with copies of the storage and editing files, as well as a shortcut to the `S:\cpsc100` folder that you can use for the rest of the term to copy other files we'll use in class. From now on, try to keep the contents of your `cpsc100` folder organized by topic.

Exercise: HTML File Practice

For this exercise, you'll take an existing web page, save the HTML source code to your computer, modify the HTML a little and add an image or two. The exercise will require techniques that we have covered since the beginning of the term. If you get stuck, just ask.

1. Using Internet Explorer, visit `http://www.driftinn.ca/home.html` and use the browser's View Source feature to look at the HTML code. The HTML will be displayed in a Notepad window.
2. Save the contents of the Notepad window to a file named `driftinn.html` on your desktop.
3. Close Internet Explorer and Notepad.
4. Create a new folder named `html-practice` somewhere under your My Documents folder. Where exactly? That's up to you.
5. Move the `driftinn.html` file from step #2 into the new `html-practice` folder.
6. Open `driftinn.html` in a web browser. Don't worry if none of the images appear; you'll fix that soon.
7. Also open `driftinn.html` in Notepad. The HTML code is all clumped together, so enable Word Wrap if it's not already: choose Word Wrap from the Format menu.
8. The <img> tags in the HTML all refer to image files on the driftinn.ca web server under the `Quickstart/ImageLib/` folder, but since your computer doesn't have this folder, you'll need to change the <img> tag URLs to refer to the driftinn.ca server. To do this, use the Find & Replace feature in Notepad to replace every instance of `src="` with `src="http://www.driftinn.ca` (pay very careful attention to the double-quotes). Save your changes, refresh the page in your browser, and you should see all of the images again.
9. The page needs a heading, so just after the opening <body> tag, add an <h1> container tag with the text "Drift Inn at Telegraph Cove". Feel free to give youself some extra space in the HTML by adding blank lines.
10. The web page at `http://www.telegraphcove.ca` has a nice sunset photograph of the cove along the top of the page. Insert an <img> tag below the new heading that references that photograph's URL (*hint: right-click on the image and choose Properties from the pop-up menu to see the image's URL*).
11. You can go whale-watching at Telegraph Cove, so go to Wikipedia (`http://en.wikipedia.org`) and search for "Killer Whale". There's a nice photo of two of them jumping on the top-right of the page. Click on the image to see the full-size version. Download the full-size image to your desktop (*hint: right-click on the image and choose Save Picture As from the pop-up menu*).
12. Move the downloaded image file to the `html-practice` folder you created in step #4.
13. Rename the downloaded image file to `orcas.jpg`.
14. Insert an <img> tag below the Telegraph Cove banner that references the `orcas.jpg` file (*hint: because the image file is in the same folder as `driftinn.html`, the `src` attribute only has to contain the name of the image file*). Make sure the whales show up in the browser.

If you can get through all that, even if it takes a while and you have to refer to your notes, you're in good shape. If you got stuck more than a couple of times with no idea how to proceed, then let me know and we'll schedule some extra practice sessions.

Exercise: Formatting with Styles

In this exercise we'll breathe life into a very dull document by re-formatting it in an eye-pleasing fashion using Word's document styles.

General Instructions

1. To begin, copy the file `Black Bear.doc` in the `S:\cpsc100\word-styles` folder to a working folder of your choice.
2. Open your copy of the document in Microsoft Word. Note that every paragraph is using the Normal style. We'll soon change that.
3. Reformat the document using just Word's paragraph styles following the suggestions below.
4. Insert the picture of a black bear found at the YTG Department of Environment site environmentyukon.gov.yk.ca/wildlifebiodiversity/mammals/blackbear.php by copying it to the clipboard and then pasting it into your document. If you then right-click on the image and choose **Format Picture** from the pop-up menu, you can make the text flow around the image by playing with the selections under the **Layout** tab.

Step #3 Paragraph Styles

All of the following styles are "built-in" to Microsoft Word, but you can also create new ones if you wish. If you can't find the built-in style on the Style screen, select "All Styles" from the list near the bottom left of the screen.

Normal

Change the paragraph formatting of the built-in Normal style to left-indent each paragraph by one-quarter of an inch. Change the Normal style's font to something other than "Times New Roman" (use a "serif" font: one that has tiny lines at the edges of the letter, rather than a block style font).

Change the paragraph formatting so that 12 points (pt) of space follow each paragraph. Turn on the display of hidden characters (the ¶ button) and delete any lines that only consist of paragraph marks (¶). Doing this spaces out the document nicely.

Heading 1

Set the style of each paragraph that begins with all-capital letters (e.g. "THE SPECIES...") to the built-in Heading 1 style (there should be 9 such paragraphs). Change the paragraph formatting of this style so that it has a zero left-indent. Choose a sans-serif (block-style) font for Heading 1 other than Arial.

Heading 2

Set the style of each sub-title paragraph ("Winter", "Spring", ..., "Claws", "Face", ...) to the built-in Heading 2 style. Change the Heading 2 style so that it is "based on" Heading 1 (note that all Heading 2 paragraphs now take on the left-indent and font of Heading 1). Change the font of Heading 2 to the same one as you used for Heading 1 and the font style to bold (no italics).

List Bullet

Set the style of the paragraphs that follow the "DID YOU KNOW?" title to the built-in List Bullet style. Also apply this style to the paragraphs that follow each of the "Claws", "Face", etc. subsections. Change the left-indent of this style to one-quarter of an inch. Check the box labelled "Don't add space between paragraphs of the same style" on the Paragraph screen.

Title

Set the style of the very first paragraph ("Black Bear") to the built-in Title style. Change the font of this style to match the one you used for Heading 1 and 2 and give it a larger point size. Remove any left-indent from the paragraph formatting.

Exercise: Combining Text, Images, Tables, and Charts in Word

In this exercise we'll once again breathe life into a very dull document by first formatting using styles and then inserting graphical images, a table of data, and an Excel chart into the mix.

The sample document for this exercise is taken from an Environment Canada report found at www.ecoinfo.ec.gc.ca/env_ind/region/caribou/caribou_e.cfm. Feel free to consult this page for hints on formatting the exercise document (although the document you'll produce is a little different).

Instructions

1. Copy the four files `porcupine-caribou-herd-report.txt`, `porcupine-caribou-herd-population-and-harvest.xls`, `porcupine-caribou-herd-calving-ground-1985.gif`, and `porcupine-caribou-herd-calving-ground-1987.gif` from the `S:\cpsc100\word-objects\` folder to a working folder of your choice.
2. Start Microsoft Word (from the Windows' **Start** menu), and from within Word, open the text file `porcupine-caribou-herd-report.txt`. Save the file as a Word Document (Select "Word Document (*.doc)" from the Save As screen's **Save as type** list) named `porcupine-caribou-herd-report.doc`.
3. Select all of the text in the document (Ctrl+A) and apply the Normal style. Change the paragraph spacing of the Normal style so that every paragraph is followed by 12 points of space. Feel free to make any other formatting change to the Normal style that strike your fancy.
4. Apply the built-in style, Title, to the first paragraph of the document ("Porcupine Caribou") and apply the built-in style, Subtitle, to the second paragraph of the document ("An indicator..."). Again, spice up these styles if the urge strikes you.
5. Apply the built-in style, Heading 1, to the four headings in the document (each one is phrased as a short question). Change the paragraph behaviour so that Word automatically inserts a page break before each occurrence of the Heading 1 style.
6. Unfortunately, if the first heading ("What is Happening") is on a new page, then only the title and subtitle of the document will appear on the first page of the document. To correct this, so that the first heading appears on first page of the document, manually change the paragraph formatting of the heading so that there is no page break (place the cursor on the heading, choose **Paragraph** from the **Format** menu and uncheck the box labelled **Page break before**).
7. Create a new style called "Image Caption", based on the Normal style, that uses a 10-point italicized font with center-aligned paragraphs. Apply the new style to the two paragraphs in the document begin with the word "Source" and that are surrounded by square brackets ([]). Remove the square brackets from these two paragraphs.
8. Create a new style called "Caption Title", based on the existing Image Caption style, that uses a 12-point bold font but without italics. Also change the left and right indent of paragraphs of this style to one inch. Apply this new style to the two paragraphs still surrounded by square brackets ("Estimated Total Population..." and "Porcupine Caribou Herd...") and remove the square brackets.
9. Open the Microsoft Excel file named `porcupine-caribou-herd-population-and-harvest.xls` and choose the tab along the bottom of the screen labelled "Chart". Move the mouse so that when you pause the cursor, a "tooltip" labelled "Chart Area" appears (near any border of the chart). Click the left mouse button once to select the entire chart and copy it to the clipboard (Ctrl+C). Return to Microsoft Word, and insert a new line below the caption title labelled "Estimated Total Population...". In that new paragraph, paste the chart copied from Excel (Ctrl+V).

10. Select the chart in Word, and apply the Image Caption style (because we haven't changed the layout of it, the copied chart is still treated just like text, and the Image Caption style will center it horizontally on the page). Right-click the copied chart, choose **Format Object** from the pop-up menu, and then choose the **Size** tab from the screen that appears. Change the width of the chart to exactly 6 inches (the height will automatically be adjusted).
11. Insert a new paragraph after the second caption title, "Porcupine Caribou Herd...", and in that space insert the diagrams from the two files `porcupine-caribou-herd-calving-ground-1985.gif` and `porcupine-caribou-herd-calving-ground-1987.gif` (choose **Picture** from the **Insert** menu and then choose the **From File** item from the submenu). Insert a couple of spaces between the two images so that there is some separation between the two (you can simply place the caret between the two images and hit the space bar a few times).
12. Insert a table at the end of the "Why is it Significant?" portion of the document. The table should contain five rows and two columns with the cell contents as shown below:

Year	Population
1992	160,000
1994	152,000
1998	129,000
2001	123,000

13. Add the title "Recent Estimated Herd Population" before the new table, using the Caption Title style.
14. Create a new style called "Table Data" based on the Normal style but *without* the 12 points of space following each paragraph. Apply this new style to the entire table. Manually change the font of the heading row to boldface.
15. Select the table and choose **Table Properties** from the **Table** menu. Change the horizontal alignment of the table to "Center".
16. Select the table and right-click. Choose **Autofit > Autofit to Contents** from the pop-up menu.

Exercise: Outline and Page Numbering

In this exercise we'll use Word's heading and page numbering features to organize a long and complex document.

Instructions

1. To begin, copy the file `xhtml-1.0-specification.doc` in the `S:\cpsc100\word-numbering\` folder to a working folder of your choice and open your copy of the document in Microsoft Word.
2. Take a look through the document. There are a number of styles in use, including Heading 1 through Heading 3. Note that the first use of Heading 1 doesn't occur until the paragraph "What is XHTML?" Feel free to change the definition of any of the existing styles to meet your tastes.
3. Change the Heading 1 style so that it uses one of the four Outline Numbering options that appear along the bottom row of the **Outline Numbered** tab on the **Bullets and Numbering** screen. By selecting one of the options that mention heading styles in the preview window, Word will automatically apply the appropriate numbering style to each level of heading.
4. In applying an outline numbering style, Word will also change the formatting of the heading styles (particularly tabs and indents), so go back and change any of the heading styles to suit your tastes. Notice that the titles at the beginning of the document were not numbered. This is because they are not formatted with the built-in Heading *X* styles, but rather using the custom styles Cover Title and Cover Subtitle (this is so that they won't show up in the Table of Contents).
5. Erase the text near the beginning of the document labelled "<Insert Table of Contents Here>" and with the caret in the same place, choose **Reference** from the **Insert** menu and then **Index and Tables**. Choose the **Table of Contents** tab and select a format for the Table of Contents from the **Formats** list. Make sure that the **Show levels** box has the number 3 (so that all headings from 1 through 3 will be displayed in the table) and click **OK**.
6. When first inserted, the Table of Contents calculates the page number at that moment, but does not automatically update the page numbers until the document is printed, saved, or the numbers are manually updated. To manually update the page numbers, select the entire document (Ctrl+A) and then press Word's update button: **F9**. A screen may appear asking whether you wish to update just the page numbers or the entire Table of Contents; select **Update entire table** to be on the safe side and press **OK**.
7. Manually format the single paragraph titled "Table of Contents" so that a page break is inserted before that paragraph.
8. Change the TOC 1, TOC 2, and TOC 3, styles so that there are zero points of space before and after each paragraph. To change these styles, click somewhere in the TOC and choose **Reference** from the **Insert** menu and then **Index and Tables**. Change the **Formats** selection to "From template" (if it wasn't already). You can now click the **Modify** button and modify each of the first three TOC styles.
9. Immediately after the Table of Contents, and before the "What is XHTML?" heading, insert a section break: choose **Break** from the **Insert** menu and then select the **Next Page** section break type. Reminder: a section break divides the document into two sections that will have very different layout or numbering formats.
10. Click the mouse somewhere in the document above the section break. Choose **Page Numbers** from the **Insert** menu and then click the **Format** button. Select the **Number format** called "i, ii, iii" and press the **OK** button on each screen. You can now see the page numbers at the bottom of each page of the first section if you switch to the Print Layout view (which should happen automatically).

11. Click the mouse somewhere in the document below the section break. Choose **Page Numbers** from the **Insert** menu and then click the **Format** button. Select the **Number format** called "1, 2, 3", make sure that the **Start At** option is selected and the number shown is 1 (so that page numbering starts over at page 1) and press the **OK** button on each screen. You can now see the page numbers at the bottom of each page of the second section if you switch to the Print Layout view. (This entire step sometimes happens automagically in Word, but this is how you apply different page numbering schemes to a document if you have to.)
12. Update the Table of Contents to reflect the new numbering (Ctrl+A, F9).
13. Switch to the Outline View and select "Show Level 3" from the list on the toolbar that appears (this will show all headings from 1 to 3). Click on the plus sign to the left of any of the headings and drag it to another part of the document. Notice how the section numbers automatically change to match. You can also click on any heading and then click either the left (Promote) or right (Demote) arrow icons on the toolbar to decrease or increase, respectively, the heading level of the selected heading. Note that the section numbers automatically change to match this time too.
14. Return to the Normal or Print Layout view and update the Table of Contents to match your new document organization (Ctrl+A, F9).

Exercise: Spreadsheet Formulas

Instructions

You don't necessarily have to follow these instructions in the same order as they are listed below. Generally speaking, it's better to leave the formatting until last.

1. Copy the `S:\cpsc100\excel-formulas\alaska-weather-statistics.xls` workbook file to a working directory of your choice and then open it in Microsoft Excel.
2. Resize the columns so that you can view the full contents of every cell. If you wish, you may also abbreviate the column titles so that they don't take up quite so much horizontal space.
3. Add a new column to the right of the worksheet titled "Rainfall (mm)". For each weather station, use a formula (starting with "=") to calculate the rainfall in millimetres from the given figure in inches (conversion: 1 inch = 25.4 millimetres).
4. Add another new column titled "Snowfall (mm)" and calculate the snowfall for each station in millimetres.
5. Add a column titled "Precipitation (mm)" and calculate the total precipitation (rainfall and snowfall) for each station in millimetres.
6. Add a column titled "Temperature (deg C)" and calculate the temperature for each station in degrees Centigrade. To convert from Fahrenheit (**`F`**) to Celsius (**`C`**), the formula is:

   ```
   C = 5 / 9 * (F - 32)
   ```

   Leave spaces out of your formulas when you enter them into Excel.
7. Add a column titled "Windspeed (m/s)" and calculate the windspeed for each station in metres per second. To help convert the units, remember that there are 1.61 kilometres in a mile.
8. Add a column titled "Windchill (deg C)" and calculate the approximate windchill using temperature in degrees Celsius (**`T`**) and windspeed in metres per second (**`W`**) in the following (slightly out of date) formula:

   ```
   Windchill = 0.045 * (5.49 * SQRT(W) + 5.81 - 0.56 * W) * (1.8 * T
   - 59.4) + 33
   ```

   ("SQRT()" means take the square root of the given value. Use Excel's `SQRT()` function to do just that.)
9. Format all of the numbers in the worksheet so that only one decimal place is displayed.
10. Dress-up the table a little using fonts, borders, shading, and alignment, as desired. You may also want to experiment with Excel's AutoFormat feature (choose **AutoFormat** from the **Format** menu).

Exercise: Spreadsheet Functions

In this exercise, we'll add formulas and functions to an existing spreadsheet to calculate college marks and, based upon the final average, which universities will offer placement. We'll also calculate potential scholarship awards to judge the total cost of tuition for each university.

Instructions

You don't necessarily have to follow these instructions in the same order as they are listed below. Generally speaking, it's better to leave the formatting until last.

1. Copy the `S:\cpsc100\excel-functions\university-planning-exercise.xls` workbook file to a working directory of your choice and open the file in Microsoft Excel.
2. Add a formula for each course in the Weighted Mark column that is the product of the course Weight and the Final Mark. Remember that to enter a formula into a cell, you must begin with an equals sign (=). Copy this formula to the other rows by using either the clipboard or the AutoFill feature (drag the little square at the bottom-right of the selection border).
3. Add formulas in the row below the last course that calculate: the total weight of all of the courses taken (use the `SUM()` function), the unweighted average of the final marks (use the `AVERAGE()` function), and the final weighted average mark (think about how to calculate this one).
4. Give the cell that contains the final weighted average mark the name "`finalmark`". You'll use this in subsequent calculations. To name a cell, type the name (no spaces allowed) into the name box at the top left of the worksheet screen and then press Enter.
5. Add formulas to calculate the Total Tuition for each university by multiplying Yearly Tuition by Program Years. But we only want this value calculated if the `finalmark` is greater than the university's Entrance Mark. Use the `IF()` function to compare the Entrance Mark with the `finalmark`: if `finalmark` is greater or equal, then calculate the Total Tuition, otherwise return 0.
6. Add formulas to calculate the Scholarship Reimbursement values (Scholarship Yearly Amount multiplied by Scholarship Years) but again, only if `finalmark` is greater than the Scholarship Mark. Otherwise the value should be 0.
7. Add formulas to calculate the Total Cost by subtracting the Scholarship Reimbursement from the Total Tuition amount.
8. Add conditional formatting to the Total Cost column that changes the background colour to red if the value is $0, green if it is between $1 and $6000, and blue if over $6000.
9. Format the number styles of the rest of the values using currency and percentage formats as appropriate.
10. Play with the numbers in the Final Mark column to change the weighted average and observe the effect on the Total Cost (and possibly the background colour).

Exercise: Date Calculations

In this exercise, we'll calculate the date for the Yukon river breakup each year in Dawson City going all the way back to 1900.

Instructions

You don't necessarily have to follow these instructions in the same order as they are listed below. Generally speaking, it's better to leave the formatting until last.

1. Copy the `S:\cpsc100\excel-dates\dawson-city-breakup-history.txt` text file to a working directory of your choice. From within Microsoft Excel, open the fixed-width text file through the Text Import Wizard. Save the file as an Excel workbook.
2. Insert column titles for the two columns: "Year" and "Julian Date", respectively. (The values in the Julian Date column represent the number of days until the breakup in that year, counting from January 1st.)
3. Add a new column, "Calculated Date", that displays the date of each year's river breakup formatted like "March 14, 1998". To perform the (approximate) calculation, remember that Excel starts numbering dates on January 1, 1900 (equivalent to one), and that there are approximately 365.25 days in a year. To check your answer, the breakup was on May 2nd in 1979, and on May 9th in 1987. You may find that you need to adjust your formula a tad to produce these dates.
4. To perform a more accurate calculation of the calendar date, add a new column called "Actual Date" that uses the **`DATE()`** function. The **`DATE()`** function takes three arguments: year, month, and day of the month. Use the value in the Year column for the year, "1" for the month (January), and the value in the Julian Date column for the day of the month. (This seems to use imply very strange dates like the 129th of January, 1900, but Excel is smart enough to figure out what we mean.) Notice that every fourth year is different between the Calculated Date and Actual Date columns. Why do you suppose that is the case?
5. Apply conditional formatting to the "Julian Date" column. If the breakup date is on the 135th day or later, colour the cell background blue (it was a cold year). If the breakup date is before the 125th day, colour the cell background orange (a warm year). Otherwise use a green background colour.
6. Note the number of warm years that have occurred recently. Check out the chart at `www.taiga.net/coop/indics/dawbkup.html` to see the overall trend.

Exercise: Charting

In this exercise, we'll display a history of fur pelt prices for various species in a chart. The data provided is from YTG Renewable Resources with updates from Yukon College's own Scott Gilbert.

Instructions

You don't necessarily have to follow these instructions in the same order as they are listed below. Generally speaking, it's better to leave the formatting until last.

1. Copy the `S:\cpsc100\excel-charts\fur-prices-at-auction.xls` workbook to a working directory of your choice. Open the workbook in Microsoft Excel.
2. Select the range of cells starting with "Bear, Black" down to the 1987 price for the Wolverine (most likely A5:B18). Choose **Chart** from the **Insert** menu and proceed through the Chart Wizard steps, using a pie chart to display the relative prices of the various pelts in 1987. Insert the chart as a new chart sheet.
3. Right-click on various portions of the chart to see what sort of options you can change. Add a title to the chart.
4. Return to the worksheet with the price data and select the range of cells starting with "Bear, Black" through to the price for a black bear pelt in 1990 (A5:E5). Start the Chart Wizard and select a line chart type. In step 2 of the wizard, select the **Series** tab and click inside the box labelled **Category (X) axis labels**. Return to the worksheet and select the range from B4 to E4 (1987 through 1990). Finish the wizard steps and insert the chart as a new sheet.
5. Right-click on the data series line and choose **Format Data Series** from the pop-up menu. Check the box labelled **Smoothed line** on the **Patterns** tab and click the **OK** button. Change the scale of the value axis (vertical) to only show prices between $50 and $90.
6. Return to the worksheet with the price data and select the range of cells starting with "Bear, Black" through to the price a coloured fox pelt in 2001 (A5:P9). Start the Chart Wizard and select a column chart type. In step 2 of the wizard, select the **Series** tab and click inside the box labelled **Category (X) axis labels**. Return to the worksheet and select the range from B4 to P4 (1987 through 2001). Finish the wizard steps and insert the chart as a new sheet.
7. Visit `www.taiga.net/coop/indics/fur.html` to view this same fur price data charted in different ways.

Exercise: Mail Merge

In this exercise we'll generate a series of job application cover letters from a Word document that serves as the Mail Merge template, and an Excel spreadsheet that will serve as the data source.

Instructions

1. Copy the files `resume-cover-letter.doc` and `job-list.xls` from the `S:\cpsc100\mail-merge\` folder to a working folder of your choice. Take a peek at the Excel file to see the data that will be used to populate the cover letters.
2. Open the Word document. Any text that is printed in ALL CAPITALS will be replaced during the exercise by either Mail Merge fields, or other Word fields.
3. The DATE field is not part of the data from the Excel file (and if you think about it, the date of the cover letter has nothing to do with job listings) so you will need to replace it with a Word field that displays today's date. To do this, select the text that you want to replace ("DATE") and choose **Field** from the **Insert** menu. From the **Categories** list, select "Date and Time" and from the **Field Names** list select "Date". The default format of the date is kind of dull, so choose one of the long date formats. You should now see today's date displayed (possibly with a grey background to indicate that this is a Word field).
4. You might as well replace "NAME" with your name while you're at it. If you were running Word from home, you could also enter this as a Word field (**Category** = "User Information", **Field Name** = "UserName") but here in the lab doing so would produce something like "Lab Account" so just type in your name directly instead.
5. We're now ready to begin the Mail Merge process, so choose **Letters and Mailings** from the **Tools** menu, and then choose **Mail Merge** from the submenu.
6. In step 1 of the process, select the **Letters** option from the **Select document type** list. Click Next to continue.
7. In step 2, select the **Use the current document** option from the **Select starting document** list. Click Next.
8. In step 3, select the **Use an existing list** option from the **Select recipients** list. Click **Browse**, and in the **Select Data Source** screen, find the `job-list.xls` file that you copied earlier and click the **Open** button. Click **OK** on the **Select Table** screen and **OK** again on the **Mail Merge Recipients** screen. Click Next.
9. In step 4, select the various uppercase placeholders in the document and replace them with merge fields by clicking the **More items** link in the sidebar. At any time, you can preview the merged documents by clicking the Next link (click Previous to return to step 4). Click Next once all of the fields have been replaced.
10. In step 5, preview the eight letters to make sure everything's okay. Click Next.
11. In step 6, you may optionally click the **Edit individual letters** link to view the eight letters in a new document. Don't bother printing the merged documents though.
12. You may also notice that the merged fields don't always fit quite right as far as the grammar is concerned. For instance, you may see something like "...your advertisement for a Accounts Manager..." which calls out for "an" rather than "a". Correct this problem and any other misfits that you spot (hint: change both the template and the data source). Once the corrections are made, you can re-run the mail merge to test your results.