Complete and comprehensive notes according to syllabus of Pakistan Pharmacy Council

Adnan's Computer for Pharmacist

Adnan S. Chaudhary & Saad M. Rustam

An easy approach to computer for health professionals and researchers.

bi-smi llāhi r-raḥmāni r-raḥīm

بِسْمِ اللّٰهِ الرَّحْمٰنِ الرَّحِيْمِ

In the name of God, the Most Gracious, the Most Merciful.

DEDICATIONS

"To Our Parents, Brothers, Teachers, Friends, and Class Fellows"

TABLE OF CONTENTS

HISTORY OF COMPUTERS: A BRIEF TIMELINE

The computer was born not for entertainment or email but out of a need to solve a serious number-crunching crisis. By 1880, the U.S. population had grown so large that it took more than seven years to tabulate the U.S. Census results. The government sought a faster way to get the job done, giving rise to punch-card based computers that took up entire rooms.

Today, we carry more computing power on our smartphones than was available in these early models. The following brief history of computing is a timeline of how computers evolved from their humble beginnings to the machines of today that surf the Internet, play games and stream multimedia in addition to crunching numbers.

1801: In France, Joseph Marie Jacquard invents a loom that uses punched wooden cards to automatically weave fabric designs. Early computers would use similar punch cards.

1822: English mathematician Charles Babbage conceives of a steam-driven calculating machine that would be able to compute tables of numbers. The project, funded by the English government, is a failure. More than a century later, however, the world's first computer was actually built.

1890: Herman Hollerith designs a punch card system to calculate the 1880 census, accomplishing the task in just three years and saving the government $5 million. He establishes a company that would ultimately become IBM.

1936: Alan Turing presents the notion of a universal machine, later called the Turing machine, capable of computing anything that is computable. The central concept of the modern computer was based on his ideas.

1937: J.V. Atanasoff, a professor of physics and mathematics at Iowa State University, attempts to build the first computer without gears, cams, belts or shafts.

1939: Hewlett-Packard is founded by David Packard and Bill Hewlett in a Palo Alto, California, garage, according to the Computer History Museum.

1941: Atanasoff and his graduate student, Clifford Berry, design a computer that can solve 29 equations simultaneously. This marks the first time a computer is able to store information on its main memory.

1943-1944: Two University of Pennsylvania professors, John Mauchly and J. Presper Eckert, build the Electronic Numerical Integrator and Calculator (ENIAC**)**. Considered the grandfather of digital computers, it fills a 20-foot by 40-foot room and has 18,000 vacuum tubes.

1946: Mauchly and Presper leave the University of Pennsylvania and receive funding from the Census Bureau to build the UNIVAC, the first commercial computer for business and government applications.

1947: William Shockley, John Bardeen and Walter Brattain of Bell Laboratories invent the transistor. They discovered how to make an electric switch with solid materials and no need for a vacuum.

1953: Grace Hopper develops the first computer language, which eventually becomes known as COBOL. Thomas Johnson Watson Jr., son of IBM CEO Thomas Johnson Watson Sr., conceives the IBM 701 EDPM to help the United Nations keep tabs on Korea during the war.

1954: The FORTRAN programming language, an acronym for FORmula TRANslation, is developed by a team of programmers at IBM led by John Backus, according to the University of Michigan.

1958: Jack Kilby and Robert Noyce unveil the integrated circuit, known as the computer chip. Kilby was awarded the Nobel Prize in Physics in 2000 for his work.

1964: Douglas Engelbart shows a prototype of the modern computer, with a mouse and a graphical user interface (GUI**)**. This marks the evolution of the computer from a specialized machine for scientists and mathematicians to technology that is more accessible to the general public.

1969: A group of developers at Bell Labs produce UNIX, an operating system that addressed compatibility issues. Written in the C programming language, UNIX was portable across multiple platforms and became the operating system of choice among mainframes at large companies and government entities. Due to the slow nature of the system, it never quite gained traction among home PC users.

1970: The newly formed Intel unveils the Intel 1103, the first Dynamic Access Memory (DRAM) chip.

1971: Alan Shugart leads a team of IBM engineers who invent the "floppy disk," allowing data to be shared among computers.

1973: Robert Metcalfe, a member of the research staff for Xerox, develops Ethernet for connecting multiple computers and other hardware.

1974-1977: A number of personal computers hit the market, including Scelbi & Mark-8 Altair, IBM 5100, Radio Shack's TRS-80 — affectionately known as the "Trash 80" — and the Commodore PET.

1975: The January issue of Popular Electronics magazine features the Altair 8080, described as the "world's first minicomputer kit to rival commercial models." Two "computer geeks," Paul Allen and Bill Gates, offer to write software for the Altair, using the new BASIC language. On April 4, after the success of this first endeavor, the two childhood friends form their own software company, Microsoft.

1976: Steve Jobs and Steve Wozniak start Apple Computers on April Fool's Day and roll out the Apple I, the first computer with a single-circuit board, according to Stanford University.

1977: Radio Shack's initial production run of the TRS-80 was just 3,000. It sold like crazy. For the first time, non-geeks could write programs and make a computer do what they wished.

1977: Jobs and Wozniak incorporate Apple and show the Apple II at the first West Coast Computer Faire. It offers color graphics and incorporates an audio cassette drive for storage.

1978: Accountants rejoice at the introduction of VisiCalc, the first computerized spreadsheet program.

1979: Word processing becomes a reality as MicroPro International releases WordStar. "The defining change was to add margins and word wrap," said creator Rob Barnaby in email to Mike Petrie in 2000. "Additional changes included getting rid of command mode and adding a print function. I was the technical brains — I figured out how to do it, and did it, and documented it. "

1981: The first IBM personal computer, code-named "Acorn," is introduced. It uses Microsoft's MS-DOS operating system. It has an Intel chip, two floppy disks and an optional color monitor. Sears & Roebuck and Computerland sell the machines, marking the first time a computer is available through outside distributors. It also popularizes the term PC.

1983: Apple's Lisa is the first personal computer with a GUI. It also features a drop-down menu and icons. It flops but eventually evolves into the Macintosh. The Gavilan SC is the first portable computer with the familiar flip form factor and the first to be marketed as a "laptop."

1985: Microsoft announces Windows, according to Encyclopedia Britannica. This was the company's response to Apple's GUI. Commodore unveils the Amiga 1000, which features advanced audio and video capabilities.

1985: The first dot-com domain name is registered on March 15, years before the World Wide Web would mark the formal beginning of Internet history. The Symbolics Computer Company, a small Massachusetts computer manufacturer, registers Symbolics.com. More than two years later, only 100 dot-coms had been registered.

1986: Compaq brings the Deskpro 386 to market. Its 32-bit architecture provides as speed comparable to mainframes.

1990: Tim Berners-Lee, a researcher at CERN, the high-energy physics laboratory in Geneva, develops HyperText Markup Language (HTML), giving rise to the World Wide Web.

1993: The Pentium microprocessor advances the use of graphics and music on PCs.

1994: PCs become gaming machines as "Command & Conquer," "Alone in the Dark 2," "Theme Park," "Magic Carpet," "Descent" and "Little Big Adventure" are among the games to hit the market.

1996: Sergey Brin and Larry Page develop the Google search engine at Stanford University.

1997: Microsoft invests $150 million in Apple, which was struggling at the time, ending Apple's court case against Microsoft in which it alleged that Microsoft copied the "look and feel" of its operating system.

1999: The term Wi-Fi becomes part of the computing language and users begin connecting to the Internet without wires.

2001: Apple unveils the Mac OS X operating system, which provides protected memory architecture and pre-emptive multi-tasking, among other benefits. Not to be outdone, Microsoft rolls out Windows XP, which has a significantly redesigned GUI.

2003: The first 64-bit processor, AMD's Athlon 64, becomes available to the consumer market.

2004: Mozilla's Firefox 1.0 challenges Microsoft's Internet Explorer, the dominant Web browser. Facebook, a social networking site, launches.

2005: YouTube, a video sharing service, is founded. Google acquires Android, a Linux-based mobile phone operating system.

2006: Apple introduces the MacBook Pro, its first Intel-based, dual-core mobile computer, as well as an Intel-based iMac. Nintendo's Wii game console hits the market.

2007: The iPhone brings many computer functions to the smartphone.

2009: Microsoft launches Windows 7, which offers the ability to pin applications to the taskbar and advances in touch and handwriting recognition, among other features.

2010: Apple unveils the iPad, changing the way consumers view media and jumpstarting the dormant tablet computer segment.

2011: Google releases the Chromebook, a laptop that runs the Google Chrome OS.

2012: Facebook gains 1 billion users on October 4.

2015: Apple releases the Apple Watch. Microsoft releases Windows 10.

2016: The first reprogrammable quantum computer was created. "Until now, there hasn't been any quantum-computing platform that had the capability to program new algorithms into their system. They're usually each tailored to attack a particular algorithm," said study lead author Shantanu Debnath, a quantum physicist and optical engineer at the University of Maryland, College Park.

2017: The Defense Advanced Research Projects Agency (DARPA) is developing a new "Molecular Informatics" program that uses molecules as computers. "Chemistry offers a rich set of properties that we may be able to harness for rapid, scalable information storage and processing," Anne Fischer, program manager in DARPA's Defense Sciences Office, said in a statement. "Millions of molecules exist, and each molecule has a unique three-dimensional atomic structure as well as variables such as shape, size, or even color. This richness provides a vast design space for exploring novel and multi-value ways to encode and process data beyond the 0s and 1s of current logic-based, digital architectures."

HISTORY OF DATA PROCESSING

Beginning in the summer of 1980, the data processing department used an IBM 34 computer as the college's administrative computing system. It had 10 times more storage, was twice as fast, and was cardless. The new equipment used diskettes, a much more efficient data storage. The machine was the latest in the area of computer science.

Oh, how times have changed! From the System 34, the college upgraded to the IBM System 36 in the mid-1980s. It used a cartridge with 10 eight-inch floppy disks, and it was upgraded to a whopping 96 megabytes of memory.

The first personal computer lab for student use was in what is now the Computer Center. In it were two Apple IIs and two Radio Shack TRS 80s. There also was a Radio Shack Model 1 for show-and-tell. A short while later, 18 computers were crammed into the north half of KTB Room 111, while the office education instructor had typewriters set up on the south half of the room. The 18 computers were Radio Shack Model IIIs. They had 32 kilobytes of memory and no hard drive. They operated on diskettes.

After a couple of years, the computers were upgraded to IBM PCs. Still, the machines did not have hard drives, and they had 4.77-megahertz processors inside. Eventually, the processors became faster on the machines and reached 16 MHz at one point. The first Model III that had an external hard drive was placed on an employee's desk in the mid-1980s. The hard drive itself cost $2,000 and had 5 megabytes of storage. Dot matrix printers were used at the time.

The first Introduction to Microcomputers class offered was in the fall semester of 1985. Fifty-nine students were enrolled. Not long after Dr. Pat McAtee arrived in July 1987 as the college's third president, a new administrative computing system was purchased. The VAX system was ordered in November 1990, and the college completed the installation of the system in 1993 as student records and administrative functions were put on the computer.

Today, the college has two VAX machines that handle e-mail, administrative functions, and student records. Through the years, IBM 386 and 486 machines have been upgraded to Pentium processors of 66 and 233 MHz, to 1.2 gigahertz. The college has approximately 800 computers for student and employee use and has a full-time computer services staff of 10. The latest computers ordered for student use are 2.4-GHz machines with 128 MB of random access memory (which will be boosted prior to installation), a compact disc burner and 20-gigabyte hard drives. To give you an idea of the speed of the processors, 2.4 GHz is equivalent to 2,400 MHz.

TYPES OF COMPUTERS

THE COMPUTER DEFINED

In basic terms, ***a computer is an electronic device that processes data, converting it into information that is useful to people***. Any computer—regardless of its type—is controlled by programmed instructions, which give the machine a purpose and tell it what to do.

The earliest computers were analog systems, and today's digital systems owe a great deal to their analog ancestors. Analog and digital computers differ in many respects, but the most important distinction is the way they represent data. Digital systems represent data as having one distinct value or another; with no other possibilities. Analog systems, however represent data as variable points along a continuous spectrum of values. This makes analog computers somewhat more flexible than digital ones, but not necessarily more precise or reliable.

TYPES OF COMPUTERS

Computers can be categorized in several ways. For example, some computers are designed for use by one person, some arc meant to be used by groups of people, and some are not used by people at all. They also can be categorized by their power, which means the speed at which they operate and the types of tasks they can handle. Within a single category, computers may be subcategorized by price, the types of hardware they contain, the kinds of software they can run, and so on.

COMPUTERS FOR INDIVIDUAL USERS

Most computers are meant to be used by only one person at a time. Such computers are often shared by several people (such as those in your school's computer lab), but only one user can work with the machine at any given moment

The six primary types of computers in this category arc

» Desktop computers

» Workstations

» Notebook computers

» Tablet computers

» Handheld computers

» Smart phones

These systems are all examples of personal computers (PCs) — ***a term that refers to any computer system that is designed for use by a single person***. Personal computers arc also called microcomputers, because they are among the smallest computers created for people to use. Note, however, that the term personal computer or PC is most often used to describe desktop computers.

Desktop Computers

The most common type of personal computer is the desktop computer—a PC that is designed to sit on (or under) a desk or table. These are the systems you sec all around you, in schools, homes, and offices, and they are the main focus of this book.

Today's desktop computers are far more powerful than those of just a few years ago, and are used for an amazing array of tasks. Not only do these machines enable people to do their jobs with greater ease and efficiency, but they can be used to communicate, produce music, edit photographs and videos, play sophisticated games, and much more. Used by everyone from preschoolers to nuclear physicists, desktop computers arc indispensable for learning, work, and play

As its name implies, a desktop computer is a full-size computer that is too big to be carried around. The main component of a desktop PC is the system unit, which is the case that houses the computer's critical parts, such as its processing and storage devices. There are two common designs for desktop computers. The more traditional desktop model features a horizontally oriented system unit, which usually lies flat on the top of the user's desk. Many users place their monitor on top of the system unit.

Vertically oriented tower models have become the more popular style of desktop system. This design allows the user to place the system unit next to or under the desk, if desired.

Workstations

A workstation is a specialized; single-user computer that typically has more power and features than a standard desktop PC. These machines are popular among scientists, engineers, and animators who need a system with greater-than-average speed and the power to perform sophisticated tasks. Workstations often have large, high-resolution monitors and accelerated graphics-handling capabilities, making them suitable for advanced architectural or engineering design, modeling, animation, and video editing.

Notebook Computers

Notebook computers, as their name implies, approximate the shape of an 8.5-by-ll-inch notebook and easily fit inside a briefcase. Because people frequently set these devices on their lap, they are also called laptop computers. Notebook computers can operate on alternating current or special batteries. These amazing devices generally weigh less than eight pounds, and some even weigh less than three pounds! During use, the computer's lid is raised to reveal a thin monitor and a

keyboard. When not in use, the device folds up for easy storage. Notebooks arc fully functional microcomputers; the people who use them need the power of a full-size desktop computer wherever they go. Because of their portability, notebook PCs fall into a category of devices called mobile computers—systems small enough to be carried by their user.

Some notebook systems are designed to be plugged into a docking station, which may include a large monitor, a full-size keyboard and mouse, or other devices. Docking stations also provide additional ports that enable the notebook computer to be connected to different devices or a network in the same manner as a desktop system.

Tablet PCs

The tablet PC is the newest development in portable, full-featured computers. Tablet PCs offer all the functionality of a notebook PC, but they are lighter and can accept input from a special pen—called a stylus or a digital pen—that is used to tap or write directly on the screen. Many tablet PCs also have a built-in microphone and special software that accepts input from the user's voice. A few models even have a fold-out keyboard, so they can be transformed into a standard notebook PC. Tablet PCs run specialized versions of standard programs and can be connected to a network. Some models also can be connected to a keyboard and a full-size monitor.

Handheld PCs

Handheld personal computers are computing devices small enough to fit in your hand. A popular type of handheld computer is the personal digital assistant (PDA). A PDA is no larger than a small appointment book and is normally used for special applications, such as taking notes, displaying telephone numbers and addresses, and keeping track of dates or agendas. Many PDAs can be connected to larger computers to exchange data. Most PDAs come with a pen that lets the user write on the screen. Some handheld computers feature tiny built-in keyboards or microphones that allow voice input.

Many PDAs let the user access the Internet through a wireless connection, and several models offer features such as cellular telephones, cameras, music players, and global positioning systems.

Smart Phones

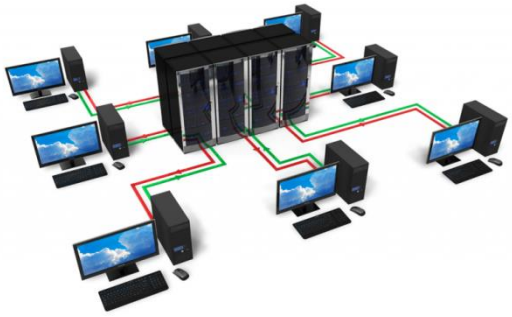
Some cellular phones double as miniature PCs. Because these phones offer advanced features not typically found in cellular phones, they are sometimes called smart phones. These features can include Web and e-mail access, special software such as personal organizers, or special hardware such as digital cameras or music players. Some models even break in half to reveal a miniature keyboard.

COMPUTERS FOR ORGANIZATIONS

Some computers handle the needs of many users at the same time. These powerful systems are most often used by organizations, such as businesses or schools, and are commonly found at the heart of the organization's network.

Generally, each user interacts with the computer through his or her own device, freeing people from having to wait their turn at a single keyboard and monitor. The largest organizational computers support thousands of individual users at the same time, from thousands of miles away. While some of these large scale systems are devoted to a special purpose, enabling users to perform only a few specific tasks, many organizational computers are general purpose systems that support a wide variety of tasks.

Network Servers



Today, most organizations' networks are based on personal computers. Individual users have their own desktop computers, which are connected to one or more centralized computers, called network servers. A network server is usually a powerful personal computer with special software and equipment that enable it to function as the primary computer in the network.

PC-based networks and servers offer companies a great deal of flexibility. For example, large organizations may have dozens or hundreds of individual servers working together at the heart of their network. When set up in such groups—sometimes called clusters or server farms—network servers may not even resemble standard PCs. For example, they may be mounted in large racks or reduced to small units called "blades," which can he slid in and out of a case. In these large networks, different groups of servers may have different purposes, such as supporting a certain set of users, handling printing tasks, enabling Internet communications, and so on.

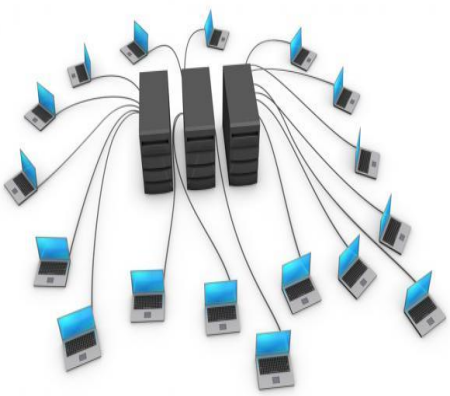
A PC-based server gives users flexibility to do different kinds of tasks. This is because PCs are general-purpose machines, designed to be used in many ways. For example, some users may rely on the server for e-mail access, some may use it to perform accounting tasks, and others may use it to

perform word-processing or database- management jobs. The server can support these processes, and many others, while storing information and programs for many people to use.

Depending on how the network is set up, users may be able to access the server in multiple ways. Of course, most users have a standard desktop PC on their desk that is permanently connected to the network. Mobile users, however, may be able to connect a notebook PC or a handheld device to the network by wireless means. When they are away from the office, users may be able to use the Internet as a means of connecting to the company's network servers.

Mainframe Computers

Mainframe computers are used in large organizations such as insurance companies and banks, where many people frequently need to use the same data. In a traditional mainframe environment, each user accesses the mainframe's resources through a device called a terminal. There are two kinds of terminals. A dumb terminal does not process or store data; it is simply an input/output (I/O) device that functions as a window into a computer located somewhere else. An intelligent terminal can perform some processing operations, but it usually does not have any storage. In some mainframe environments, however, workers can use a standard personal computer to access the mainframe.



Mainframes are large, powerful systems. The largest mainframes can handle the processing needs of thousands of users at any given moment. But what these systems offer in power, they lack in flexibility. Most mainframe systems are designed to handle only a specific set of tasks. In your state's Department of Motor Vehicles, for example, a mainframe system is probably devoted to storing information about drivers, vehicles, and driver's licenses, but little or nothing else. By limiting the number of tasks the system must per-form, administrators preserve as much power as possible for required operations.

You may have interacted with a mainframe system with-out even knowing it. For example, if you have visited an airline's Web site to reserve a seat on a flight; you probably conducted a transaction with a mainframe computer.

Minicomputers

First released in the 1960s, minicomputers got their name because of their small size compared to other computers of the day. The capabilities of a minicomputer are somewhere between those of mainframes and personal computers. For this reason, minicomputers are often called midrange computers.

Like mainframes, minicomputers can handle much more input and output than personal computers can. Although some ‘‘minis’’ are designed for a single user, the most powerful minicomputers can serve the input and output needs of hundreds of users at a time. Users can access a central minicomputer through a terminal or a standard PC.

Supercomputers

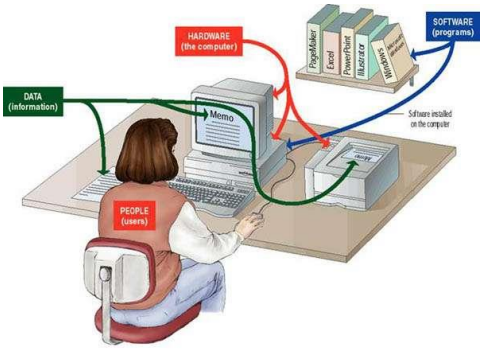
Supercomputers arc the most powerful computers made, and physically they are some of the largest. These systems can process huge amounts of data, and the fastest supercomputers can perform more than one trillion calculations per second. Some supercomputers can house thousands of processors. Supercomputers are ideal for handling large and highly complex problems that require extreme calculating power. For example, supercomputers have long been used in the mapping of the human genome, forecasting weather, and modeling complex processes like nuclear fission.

EMBEDDED COMPUTERS

Embedded computers are components of larger products that usually have a digital interface. These computers use a specially programmed microprocessor to perform a set of predefined tasks, and may require little or no input from the user. Microwave ovens, digital cameras, programmable thermostats, and airbags and antilock braking systems for cars are just a few examples of products that use embedded computers.

COMPONENTS OF A COMPUTER

Computers come in many varieties, from the tiny computers built into household appliances, to the astounding supercomputers that have helped scientists map the human genome. But no matter how big it is or how it is used, every computer is part of a system. A complete computer system consists of four parts.



» Hardware

» Software

» Data

» User

HARDWARE

The mechanical devices that make up the computer are called hardware. Hardware is any part of the computer you can touch. A computer's hardware consists of interconnected electronic devices that you can use to control the computer's operation, input, and output. (The generic term device refers to any piece of hardware.)

SOFTWARE

Software is a set of instructions that makes the computer perform tasks. In other words, software tells the computer what to do. (The term program refers to any piece of software.) Some programs exist primarily for the computer's use to help it perform tasks and manage its own resources. Other types of programs exist for the user, enabling him or her to perform tasks such as creating documents. Thousands of different software programs are available for use on personal computers

DATA

Data consist of individual facts or pieces of information that by themselves may not make much sense to a person. A computer's primary job is to process these tiny pieces of data in various ways, converting them into useful information. For example, if you saw the average highway mileages of six different cars, all the different pieces of data might not mean much to you. However; if someone created a chart from the data that visually compared and ranked the vehicles' mileages, you could probably make sense of it at a glance. This is one example of data being processed into useful information.

USER

People are the computer operators, also known as users. It can be argued that some computer systems are complete without a person's involvement; however, no computer is totally autonomous. Even if a computer can do its job without a person sitting in front of it, people still design, build, program, and repair computer systems. This lack of autonomy is especially true of personal computer systems, which arc the focus of this book and are designed specifically for use by people.

THE INFORMATION PROCESSING CYCLE

Using all its parts together, a computer converts data into information by per- forming various actions on the data. For example, a computer might perform a mathematical operation on two numbers, and then display the result. Or the computer might perform a logical operation such as comparing two numbers, then display that result. These operations are part of a process called the information processing cycle, which is a set of steps the computer follows to receive data, process the data according to instructions from a program, display the resulting information to the user, and store the results.

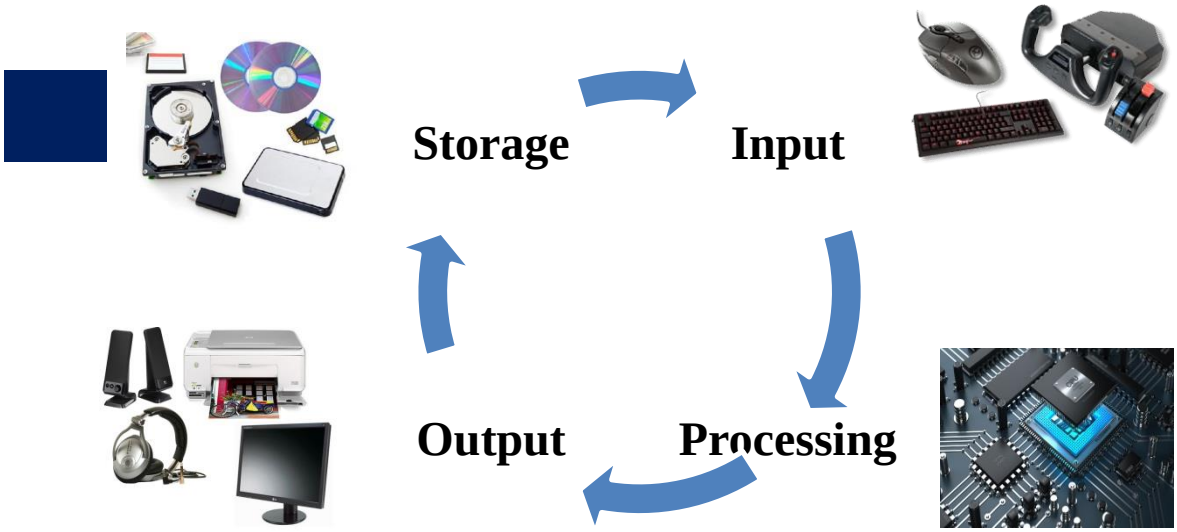
The information processing cycle has four parts, and each part involves one or more specific components of the computer:

» **Input** During this part of the cycle, the computer accepts data from some source, such as the user or a program, for processing.

» **Processing** During this part of the cycle, the computer's processing components perform actions on the data, based on instructions from the user or a program.

» **Output** Here, the computer may be required to display the results of its processing. For example, the results may appear as text, numbers, or a graphic on the computer's screen or as sounds from its speaker. The computer also can send output to a printer or transfer the output to another computer through a network or the Internet. Output is an optional step in the information processing cycle but may be ordered by the user or program.

» **Storage** In this step, the computer permanently stores the results of its processing on a disk, tape, or some other kind of storage medium. As with output, storage is optional and may not always be required by the user or program.



ESSENTIAL COMPUTER HARDWARE

Hardware is the computer and any equipment connected to it. Hardware devices are the physical components of the computer. Items such as the monitor, keyboard, mouse, and printer are also known as peripherals because they attach to the computer.

The computer itself is known as the system unit, and it contains many of the critical hardware and electrical components. The system unit is sometimes referred to as the tower, box, or console. When the system unit is combined with the appropriate peripheral devices, the system can perform the four basic computer functions: input, process, output, and storage. Peripheral devices are used to input and output data and information, and the system unit processes and stores the data.

A computer's hardware devices fall into one of four categories.

1. Processor / System Unit
2. Memory
3. Input and output devices
4. Storage

PROCESSOR / SYSTEM UNIT

The procedure that transforms raw data into useful information is called processing. To perform this transformation, the computer uses two components: the processor and memory.

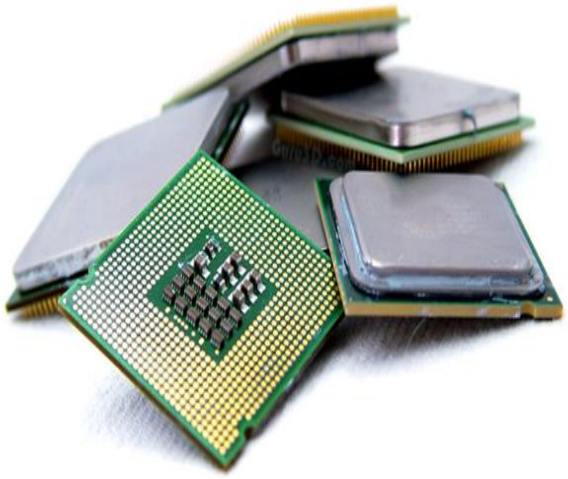
The **processor** is like the brain of the computer; it organizes and carries out instructions that come from either the user or the software. In a personal computer; the processor usually consists of one or more specialized chips, called microprocessors, which are slivers of or other material etched with many tiny electronic circuits. To process data or complete an instruction from a user or a program, the computer passes electricity through the circuits.

The microprocessor is plugged into the computer's motherboard. The **motherboard** is a rigid rectangular card

containing the circuitry that connects the processor to the other hardware. The motherboard is an example of a circuit board. In most personal computers, many internal devices such as video cards, sound cards, disk controllers, and other devices—are housed on their own smaller circuit boards, which attach to the motherboard. In many newer computers, these devices are built directly into the motherboard. Some newer microprocessors are large and complex enough to require their own dedicated circuit boards, which plug into a special slot in the mother-board. You can think of the motherboard as the master circuit board in a computer.

A personal computer's processor is usually a single chip or a set of chips contained on a circuit board. In some powerful computers, the processor consists of many chips and the circuit boards on which they are mounted. In either case, the term central processing unit (CPU) refers to a computer's processor. People often refer to computer systems by the type of CPU. They contain. A "Pentium 4" system, for example, uses a Pentium 4 microprocessor as its CPU. "Core i5" system contain Core i5 Processor as its CPU.

The CPU is the brain of the computer, and is responsible for controlling the commands and tasks that the computer performs. It has two main parts—the control unit and the arithmetic logic unit (ALU). The **control unit** is responsible for obtaining instructions from the computer's memory. It then interprets these instructions and executes them, thereby coordinating the activities of all the other computer components. **The arithmetic logic unit**, or ALU, performs all the arithmetic and logic functions for the computer. The ALU handles addition, subtraction, multiplication, and division, and also makes logical and comparison decisions. This enables the CPU to perform tasks such as sorting data alphabetically or numerically and filtering data to locate specific criteria. Third is the **cache**, which serves as high-speed memory where instructions can be copied to and retrieved.



Inside a processor, we can store zeros and ones using **transistors**. These are microscopic switches that control the flow of electricity depending on whether the switch is on or off. So the transistor contains binary information: a one if a current passes through and a zero if a current does not pass through.

Transistors are located on a very thin slice of silicon. A single silicon chip can contain thousands of transistors. A single CPU contains a large number of chips. Combined, these only cover about a square inch or so. In a modern CPU, however, that square inch can hold several hundred million transistors - the very latest high-end CPUs have over one billion.

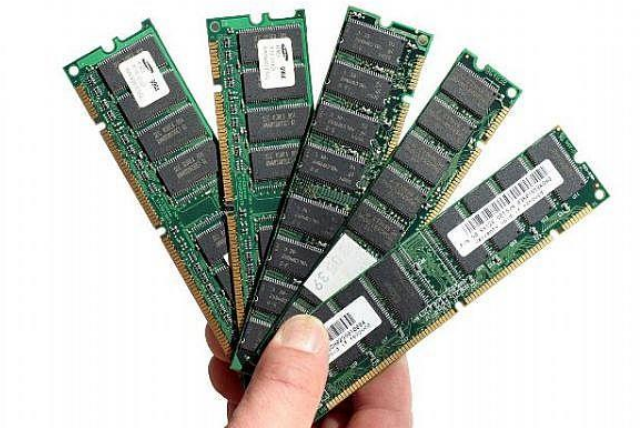
MEMORY DEVICES

In a computer, memory is one or more sets of chips that store data and/or program instructions, cither temporarily or permanently. Memory is a critical processing component in any computer. Personal computers use several different types of memory, but the two most important are called random

access memory (RAM) and read-only memory (ROM). These two types of memory work in very different ways and perform distinct functions.

Random Access Memory

The most common type of memory is called random access memory (RAM). As a result, the term memory is typically used to mean RAM. RAM is like an electronic scratch pad inside the computer. RAM holds data and program instructions while the CPU works with them. When a program is launched, it is loaded into and run from memory. As the program needs data, it is loaded into memory for fast access. As new data is entered into the computer, it is also stored in memory—but only temporarily. Data is both written to and read from this memory. (Because of this, RAM is also sometimes called read/write memory.)



Like many computer components, RAM is made up of a set of chips mounted on a small circuit board.

RAM is volatile, meaning that it loses its contents when the computer is shut off or if there is a power failure. Therefore, RAM needs a constant supply of power to hold its data. For this reason, you should save your data files to a storage device frequently, to avoid losing them in a power failure.

RAM has a tremendous impact on the speed and power of a computer. Generally, the more RAM a computer has, the mote it can do and the faster it can perform certain tasks. The most common measurement unit for describing a computer's memory is the byte—the amount of memory it takes to store a single character such as a letter of the alphabet or a numeral. When referring to a computer's memory, the numbers are often so large that it is helpful to use terms such as kilobyte (KB), megabyte (MB), gigabyte (GB), and terabyte (TB) to describe the values. For newer systems, a minimum of 512 MB to 1 GB is recommended.

Dynamic RAM (DRAM) and **static RAM (SRAM)** are the types of RAM. DRAM is the main memory in a computer and SRAM is used for high-speed caches and buffers. Both types are "byte addressable," which means that data can be read and written one byte at a time.

Read-Only Memory

Unlike RAM, Read-only memory (ROM) permanently stores its data, even when the computer is shut off. ROM is called nonvolatile memory because it never loses its contents. ROM holds instructions that the computer needs to operate. Whenever the computer's power is turned on, it checks ROM for directions that help it start up, and for information about its hardware devices. The information on a ROM chip can't be changed, removed, or rewritten and is generally inaccessible to the computer user.

Permanent (ROM and PROM): Data and instructions in ROM and PROM chips can never be changed. ROMs are manufactured, while PROMs are programmed in an external device like EPROMs.

EPROM (Erasable Programmable ROM): A rewritable memory chip that holds its content without power. EPROM chips are written on an external programming device before being placed on the circuit board. The chip requires an expensive ceramic chip package with a small quartz window that is covered with opaque, sticky tape. For reprogramming, the chip is extracted from the circuit board, the tape is removed, and it is placed under an intense ultraviolet (UV) light for approximately 20 minutes.

EEPROM (Electrically Erasable Programmable ROM): A rewritable memory chip that holds its content without power. EEPROMs are bit or byte addressable at the write level, which means either the bit or byte must be erased before it can be re-written.

INPUT AND OUTPUT DEVICES

A personal computer would be useless if you could not interact with it because the machine could not receive instructions or deliver the results of its work. **Input devices** accept data and instructions from the user or from another computer system (such as a computer on the Internet). **Output devices** return processed data to the user or to another computer system.

The most common input device is the **keyboard**, which accepts letters, numbers, and commands from the user. There are actually several different kinds of keyboards. The QWERTY keyboard is the one most people are familiar with. It is based on the original typewriter keyboard and is named for the arrangement of the letters on the top-left alphabetic row of keys. Another style is the Dvorak keyboard, which arranges the letters and numbers in a different pattern for increased typing speed. Some ergonomic keyboards use a split keyboard arrangement, offsetting each half at an angle to reduce the incidence of repetitive stress injuries such as carpal tunnel syndrome.

Keyboard size and layout on notebook and tablet computers can differ slightly from a standard desktop keyboard. Keyboards usually send information to the computer through a cable connected to a USB port; however, wireless or remote keyboards are gaining in popularity. In addition to the standard alphanumeric keys originally found on typewriters, computer keyboards have a variety of keys that provide additional functionality.

- Control keys, such as the Ctrl, Alt, and Windows keys, often provide shortcuts or increased functionality to the keyboard when used in combination with another key. For example, pressing Ctrl and the letter P opens the Print dialog box.
- The numeric keypad, located at the right of the keyboard, provides an alternate method of quickly entering numbers.
- Function keys are located above the standard row of number keys. Numbered F1 through F12, these keys are generally associated with certain software-specific commands. Pressing the F1 key will usually open the Help menu for a program.
- Arrow keys are the keys located at the bottom of the keyboard between the standard keys and the numeric keypad. These keys enable the user to move the insertion point around the window one space at a time.
- Toggle and other keys, which are located just above the arrow keys, are used for various purposes, including navigation and editing. The Insert, Num Lock, and Caps Lock keys are all examples of toggle keys.
- Multimedia and Internet control keys are typically found at the top edge of the keyboard. The precise placement and function of these keys usually depends on the keyboard manufacturer.

Another important type of input device is the **mouse**, which lets you select options from on-screen menus. You use a mouse by moving it across a flat surface and pressing its buttons. Optical mice use a laser beam, instead of a rollerball, to control the pointer movement. Notebook and tablet computers can use mice, or they may use a built-in touchpad, trackball, or track point to move the insertion point. The traditional mouse has two buttons and a scroll wheel. The palm of your hand should rest comfortably over the mouse. The following provides a brief description of

some of the ways the mouse can be used:

- Click—By default, the left mouse button is considered the primary button. When instructed to click the mouse, it is understood this means that the left mouse button should be pressed one time.
- Double-click—Double-clicking is performed by pressing the left mouse button two times in rapid succession. It is important that the mouse does not move while double-clicking or the command will not produce the expected results. Example, you double-click to open a file or start a program.
- Drag—To carry out this action, press the left mouse button and continue to hold it while dragging, or moving, the mouse. This action can be used to select large blocks of text.
- Right-click—Pressing the right mouse button one time will open a shortcut menu. Shortcut menus are usually context-sensitive, which means they will vary depending on what you've clicked and what program you are using.
- Scroll wheel—If your mouse is equipped with a scroll wheel, it can be used to quickly move a page up or down in a window.

A variety of other input devices work with personal computers.

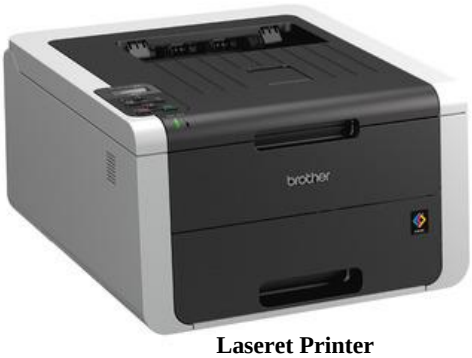
The trackball and touchpad are variations of the mouse and enable you to draw or point on the screen. The joystick is a swiveling lever mounted on a stationary base that is well suited for playing video games. A scanner can copy a printed page of text or a graphic into the computer's memory, freeing you from creating the data from scratch. A digital camera can record still images, which you can view and edit on the computer .A microphone enables you to input your voice or music as data.

The function of an output device is to present processed data to the use. The most common output devices are the **monitor** and the printer. The computer sends output to the monitor (the display screen) when the user needs only to see the output. Text, video, and graphics are displayed on a monitor. When a monitor outputs data or information, it is called soft copy—you can view it, but you can't touch it. It sends output to the printer when the user requests a paper copy—also called a hard copy—of a document. Monitors come in a variety of sizes and styles, but, there are just two main categories: **cathode-ray tube (CRT) and liquid crystal display (LCD)**. A CRT monitor resembles a traditional television set and uses a cathode-ray tube to produce the picture on the screen. The glass screen of a CRT monitor can be curved or flat (sometimes called a flat screen monitor). Flat

panel or LCD monitors use a liquid crystal display and are much thinner and lighter than CRT monitors. They are also more expensive than CRTs, although they have become more affordable in recent years. A monitor's display is made up of millions of tiny dots, known as pixels. The number of pixels on the screen determines a monitor's sharpness and clarity, also known as its resolution. A standard screen resolution might be expressed as 1024 x 768, which means there are 1,024 columns, each containing 768 pixels. Dot pitch is another display characteristic and refers to the diagonal distance between two pixels of the same color. The speed at which the pixels are reilluminated is called the refresh rate, which is measured in cycles per second, expressed as hertz (Hz).

Using a monitor is a good way to view the information on your computer, but sometimes a soft copy isn't sufficient for your needs. The ability to generate a hard copy—a permanent record of your work—is the primary benefit of a **printer.**

There are two categories of printers: impact and nonimpact. **Impact printers** have small keys, similar to a typewriter's, that strike an ink ribbon against paper, leaving behind an image of the character on the key. The dot matrix printer is an impact printer**. Nonimpact printers** do not actually touch the paper when printing. There are a variety of nonimpact printers, but the two most commonly used with home computers are the ink-jet printer and the laser printer. The ink-jet printer uses a special nozzle and ink cartridges to spray ink in small droplets onto the surface of the paper. Ink-jet printers are able to easily print in color and in black and white, produce good quality copy, and are relatively inexpensive to buy. Laser printers use the same process as photocopiers to produce their output. They use a special cylinder known as a drum, dry ink or toner, and a laser.



Laseret Printer

Just as computers can accept sound as input, (hey can use stereo speakers or headphones as output devices to produce sound. Multimedia projectors are used to conduct presentations and training sessions. Imagine how difficult it would be to have a room full of students or conference attendee's crowd around a single monitor to view a presentation. A multimedia projector allows information to

be projected onto a larger screen so it can easily be viewed by a group.

Some types of hardware can act as both input and output devices. A touch screen, for example, is a type of monitor that displays text or icons you can touch. When you touch the screen, special sensors detect the touch and the computer calculates the point on the screen where you placed your finger. Depending on the location of the touch, the computer determines what information to display or what action to take next. A good example of such a device is the telephone, which allows you to both speak (output) and listen (input) to another person. Other examples include the touchscreen monitor at a convenience store or ATM or the "all-in-one" printer, which combines a printer with a scanner, copier, and fax machine.

Communications devices are the most common types of devices that can perform both input and output. These devices connect one computer 10 another—a process known as networking. The most common kinds of communications devices are modems, which enable computers to communicate through telephone lines or cable television systems, and network interface cards (NICs), which let users connect a group of computers to share data and devices.

STORAGE DEVICES

Storage devices are used to store the data and information used by or created with the computer. Such storage is often referred to as permanent memory because, unlike data that is in RAM, data saved to a storage device remains there until the user deletes or overwrites it. A computer can function with only processing, memory, input, and output devices. To be really useful, however; a computer also needs a place to keep program files and related data when they are not in use. The purpose of storage is to hold data permanently, even when the computer is turned off. Data can be stored within internal hardware devices located within the system unit or in removable external units.

Additionally, storage can be fixed or portable, depending on whether the data saved remains within the system unit or is saved on removable units and accessed elsewhere.

You may think of storage as an electronic file cabinet and RAM as an electronic worktable. When you need to work with a program or a set of data, the computer locates it in the file cabinet and puts a copy on the table. After you have finished working with the program or data, you put it back into the file cabinet. The changes you make to data while working on it replace the original data in the file cabinet (unless you store it in a different place).

Novice computer users often confuse storage with memory. Although the functions of storage and memory arc similar; they work in different ways. There are three major distinctions between storage and memory:

- There is more room in storage than in memory, just as there is more room in a file cabinet than there is on a tabletop.
- Contents are retained in storage when the computer is turned off, whereas programs or the data in memory disappear when you shut down the computer.
- Storage devices operate much slower than memory chips, bur storage is much cheaper than memory.

Magnetic Storage

Magnetic storage uses tape or film covered in a thin, magnetic coating that enables data to be saved as magnetic particles. There are many types of computer storage, but the most common is the magnetic disk. A disk is a round, flat object that spins around its center. (Magnetic disks are almost always housed inside a case of some kind, so you can't see the disk itself unless you open the case.) Read/write heads, which work in much the same way as the heads of a tape recorder or VCR, are used to read data from the disk or write data onto the disk.

The device that holds a disk is called a disk drive. Some disks are built into die drive and are not meant to be removed; other kinds of drives enable you to remove and replace disks. In addition, there is also a diskette drive, which allows you to use removable diskettes (or floppy disks). Diskettes are used to load data onto the hard disk, to trade data with other users, and to make backup copies of the data on the hard disk.

Magnetic disks are divided into tracks and sectors. Just like an old vinyl record, tracks form rings around the circumference of the media. Sectors divide the tracks into pie-shaped wedges extending from the center to the outer edge of the disk. Data is stored magnetically within the sectors.

Optical Storage

Storage uses flat plastic discs coated in a special reflective material. Data is saved by using a laser beam to burn tiny pits into the storage medium. Storage uses flat plastic discs coated in a special reflective material. Data is saved by using a laser beam to burn tiny pits into the storage medium.

The CD-ROM drive is the most common type of optical storage device. Compact discs (CDs) arc a type of optical storage, identical to audio CDs. Until recently, a standard CD could store about 74 minutes of audio or 650 M B of data. A newer breed of CDs can hold 80 minutes of audio or 700 M B of data. The type used in computers is called Compact Disc Read-Only Memory (CD-ROM). As the name implies, you cannot change the information on the disc, just as you cannot record over an audio CD.

If you purchase a CD-Recordable (CD-R) drive, you have the option of creating your own CDs. A CD-R drive can write data to and read data from a compact disc. To record data with a CD-R

drive, you must use a special CD-R disc, which can be written on only once, or a CD-Rewritable (CD-RW) disc, which can be written to multiple times, like a floppy disk.

An increasingly popular data storage technology is the Digital Video Disc (DVD), which is revolutionizing home entertainment. Using sophisticated compression technologies, a single DVD (which is the same size as a standard compact disc) can store an entire full-length movie. DVDs can hold a minimum of 4.7 GB of data and as much as 17 GB. Future DVD technologies promise much higher storage capacities on a single disc. DVD drives also can locate data on the disc much faster than standard CD-ROM drives. There are currently two competing formats—DVD-R/RW, known as "DVD dash," and DVD+R/RW, known as "DVD plus." The R/RW suffix indicates the DVD can be used to record and can also be rewritten.

DVDs require a special player. Many DVD players, however, can play audio, data, and DVD discs, freeing the user from purchasing different players for each type of disc. DVD drives are now standard equipment on many new personal computers. Users not only can install programs and data from their standard CDs, but they also can watch movies on their personal computers by using a DVD.

Flash memory

Flash memory uses solid-state technology. It is completely electronic and has no moving mechanical parts. Flash memory is a quick and easy form of rewritable storage, capable of exceeding the storage capacity of magnetic or optical media. Flash memory cards are often used in mobile devices such as PDAs, digital cameras, and MP3 players. Depending on the manufacturer, flash memory cards may be called Memory Stick, CompactFlash, Secure Digital, or MultiMediaCard. Typically, a device can use only one style of memory card; however, a computer equipped with the appropriate card reader can read any of them. Small, removable storage devices known as flash drives also use flash technology and have become increasingly popular.

SOFTWARE

The ingredient that enables a computer to perform a specific task is software, which consists of instructions. ***A set of instructions that drive a computer to per- form specific tasks is called a program.*** These instructions tell the machine's physical components what to do; without the instructions, a computer could not do anything at all. When a computer uses a particular program, it is said to be running or executing that program.

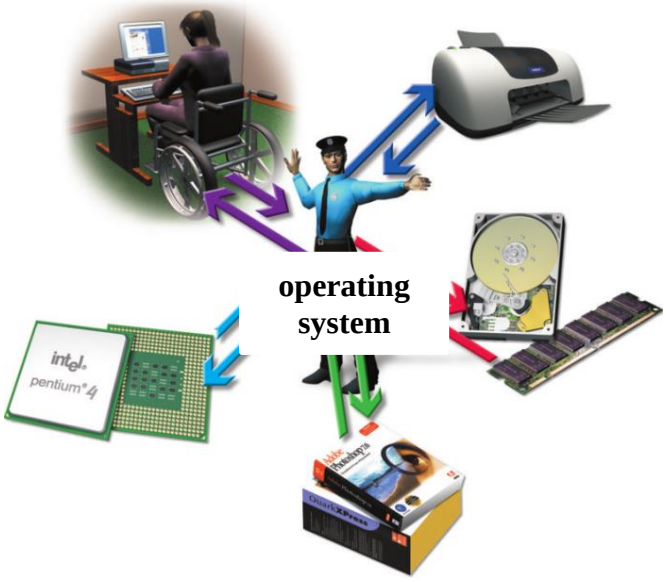
Although the array of available programs is vast and varied, most software falls into two major categories: system software and application software.

SYSTEM SOFTWARE

System software is any program (hat controls the computer's hardware or that can be used to maintain the computer in some way so that it runs more efficiently. There are three basic types of system software:

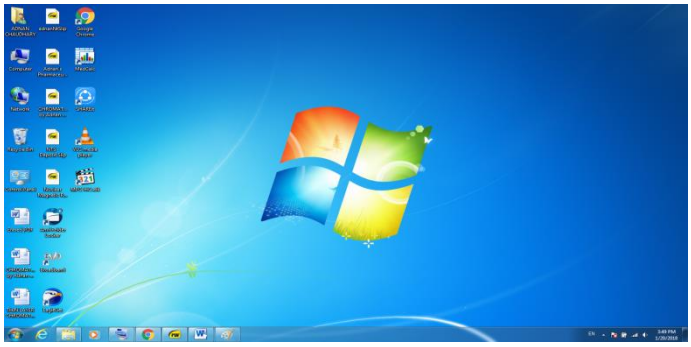
Operating Systems

An operating system tells the computer how to use its own components. Examples of operating systems include Windows, the Macintosh Operating System, and Linux. An operating system is essential for any computer; because it acts as an interpreter between the hardware, application programs, and the user. When a program wants the hardware to do something, it communicates through the operating system. Similarly, when you want the hardware to do something (such as copying or printing a file), your request is handled by the operating system.



Although the operating system communicates with the computer and its peripherals, it also includes a user interface that you can use to interact with the computer. Early operating systems used a DOS-based interface, which required knowledge of special commands that had to be typed accurately to achieve the desired results. As you can imagine, this type of system was not very user-friendly. Most current operating systems use a point-and-click format known as a graphical user interface (GUI). GUIs are more user-friendly and intuitive than DOS systems. Rather than typing specific commands, you can use a mouse to point to and click on an icon (a graphical depiction of an object such as a file or program) or a menu (a list of commands) to perform a task. GUI operating systems display information on the monitor in the form of rectangular boxes called screens or windows.

Microsoft Windows has the largest market share of the three main operating systems and is found on most of today's desktop and notebook computers. There have been many versions of Microsoft Windows, including Windows 3.0, Windows 95, Windows 98, Windows ME, and



Windows Vista. Although a previous version of Windows might be found on an older computer, Windows 7, Window 8 and Window 10 is the current version installed on most computers.

Mac OS is an operating system designed specifically for Apple's Macintosh computers. The current version is Mac OS 4.8 Mac OS X 10.6 Snow Leopard, & 4.9 Mac OS X 10.7 Lion. As you can the Mac OS appears similar to Windows, because it also uses a GUI. In fact, Apple was the first company to introduce a GUI operating system for commercial sale. But, because of the overwhelming popularity of the Windows-based PC, Mac OS has a much smaller market share. There are also significant differences in the way the Mac OS performs. Mac users tend to be very loyal and believe their system is far superior to the Windows system, although there are many Windows users who disagree.

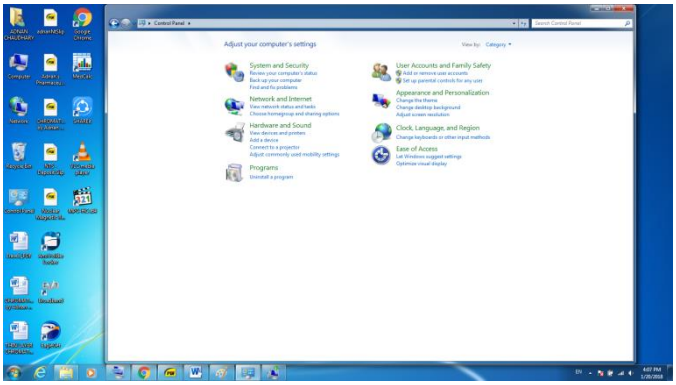
Linux is an alternative operating system. Based on the UNIX operating system developed for mainframe computers, it also has a dedicated group of users. Linux is an open-source operating system, which means it is not owned by a single company and some versions are available at no cost.

Network operating system

A network operating system allows computers to communicate and share data across a network while controlling network operations and overseeing the network's security.

Utility Programs

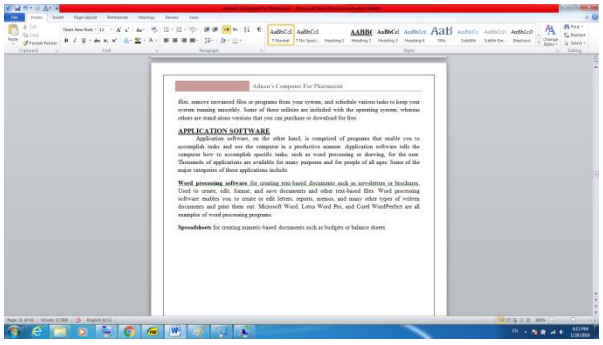
A utility is a program that makes the computer system easier to use or performs highly specialized functions. Utilities arc used to manage disks, troubleshoot hard ware problems, and perform other tasks that the operating system itself may not be able to do. Utility programs can be used to help back up important



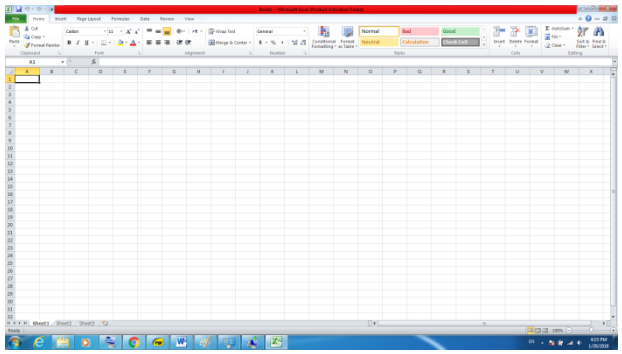
files, remove unwanted files or programs from your system, and schedule various tasks to keep your system running smoothly. Some of these utilities are included with the operating system, whereas others are stand-alone versions that you can purchase or download for free.

APPLICATION SOFTWARE

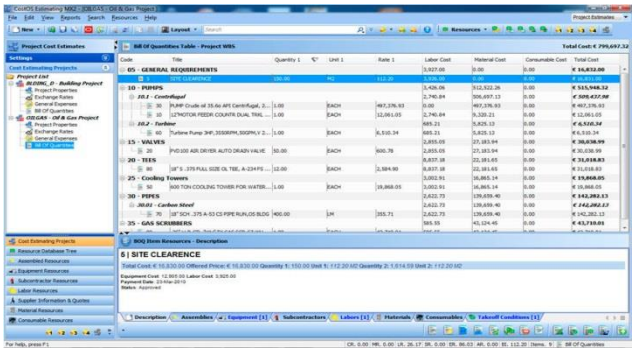
Application software, on the other hand, is comprised of programs that enable you to accomplish tasks and use the computer in a productive manner. Application software tells the computer how to accomplish specific tasks, such as word processing or drawing, for the user. Thousands of applications are available for many purposes and for people of all ages. Some of the major categories of these applications include.



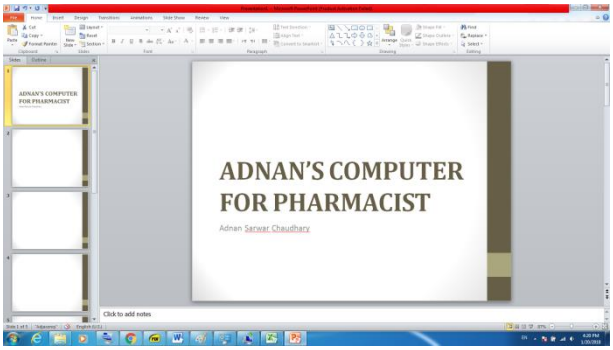
Word processing software for creating text-based documents such as newsletters or brochures. Used to create, edit, format, and save documents and other text-based files. Word processing software enables you to create or edit letters, reports, memos, and many other types of written documents and print them out. Microsoft Word, Lotus Word Pro, and Corel WordPerfect are all examples of word processing programs.



Spreadsheets for creating numeric-based documents such as budgets or balance sheets. Spreadsheet software enables you to perform calculations and other mathematical tasks. Similar to the documents used by accountants, spreadsheets contain data entered in columns and rows and enable you to perform calculations, create scenarios, per- form "what-if" analyses, chart and graph data, and format the worksheet layout.



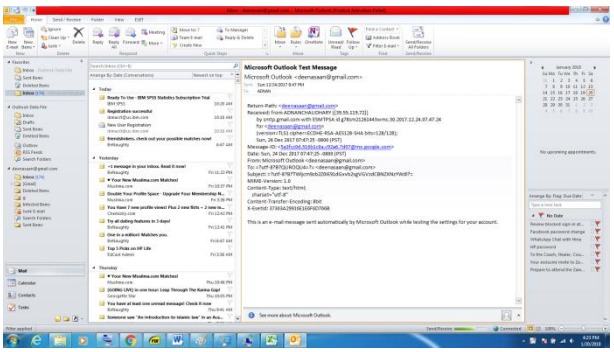
Database software—Databases are used to store and organize large amounts of data. Typically, database software can be used to manage various types of information, such as that found in large mailing lists, inventories, order histories, and invoicing. Databases help you to enter, store, sort, filter, retrieve, and summarize the information they contain and then generate meaningful reports.



Presentation software—Because of presentation software, lecturers no longer need to rely on flip charts, slide projectors, or overhead transparencies for their presentations. This software is used to create graphic presentations, known as slide shows, that can

be shown to large groups by means of an overhead projector or displayed on the Web. Microsoft PowerPoint, Lotus Freelance Graphics, and Corel Presentations are examples of presentation software programs.

Communication and organizational software—Communication software can cover a broad range of tasks including videoconferencing and telephony. However, applications within the productivity category are most often used to send and receive e-mail. These applications typically include an address book, a calendar, and task functions, which help users organize their personal and professional responsibilities. Microsoft Outlook, Lotus Notes, and Corel WordPerfect Mail are examples of communication and organizational software.



COMPUTER DATA

Data is any piece of information or fact that, taken by itself, may not make sense to a person. For example, you might think of the letters of the alphabet as data. Taken individually, they do not mean a lot. But when grouped into words and sentences, they make sense; that is, they become information. Similarly, basic geometric shapes may not have much meaning by themselves, but when they are grouped into a blueprint or a chart, they become useful information.

Within the computer, data is organized into files. A file is simply a set of data that has been given a name. A file that the user can open and use is often called a document. Although many people think of documents simply as text, a computer document can include many kinds of data. For example, a computer document can be a text file (such as a letter), a group of numbers (such as a budget), a video clip (which includes images and sounds), or any combination of these items.

COMPUTER USERS

Personal computers, which are the focus of this book, arc designed to work with a human user. In fact, the user is a critical part of a complete computer system, especially when a personal computer is involved. This may seem surprising, since we tend to think of computers as intelligent devices, capable of performing amazing tasks. People also sometimes believe that computers can think and make decisions, just like humans do. But this is not the case. Even the most powerful supercomputers require human interaction—if for no other reason than to get them started and tell them which problems to solve.

THE USER'S ROLE

When working with a personal computer; the user can take on several roles, de-pending on what he or she wants to accomplish:

Setting up the System.

If you want to change something about the system (a process called configuration), you will likely do it yourself, whether you want to add a new hardware device, change the way programs look on your screen, or customize the way a program functions.

Installing Software.

Although your new computer probably came with an operating system and some applications installed, you need to install any other programs you want to use. This may involve loading software from a disk or downloading it from a Web site. Either way, it is usually the user's responsibility to install programs, unless the computer is used at a school or business. In that case, a system administrator or technician may be available to do the job.

Running Programs

Whenever your computer is on, there are several program running in the background, such as the software that runs your mouse and printer. Such programs do not need any user input; in fact, you may not even be aware of them. But for the most part, if you want to use your computer to per- form a task, you need to launch and run the software that is designed for the task. This means installing the program, learning its tools, and working with it to make sure it gives you the results you want.

Managing Files

As you have already learned, a computer saves data in files. If you write a letter to a friend, you can save it as a file, making it available to open and use again later. Pictures, songs, and other kinds of data are stored as files. But it is the user's job to manage these files, and this means setting up a logical system for storing them on the computer It also means knowing when to delete or move files, or copy them to a disk for safekeeping.

Maintaining the System

System maintenance does not necessarily mean opening the PC and fixing broken parts, as you would repair a car's engine. But it could! In that case, you might call a qualified technician to do the job, or roll up your sleeves and tackle it yourself. PC maintenance, however, generally means

running utilities that keep the disks free of clutter and ensure that the computer is making the best use of its resources.

UNIT OF MEMORY

Memory unit is the amount of data that can be stored in the storage unit. This storage capacity is expressed in terms of Bytes.

The following table explains the main memory storage units.

1	**Bit(Binary Digit)**	A binary digit is logical 0 and 1 representing a passive or an active state of a component in an electric circuit. The smallest unit of memory is the "bit". A bit can be in one of two states— on vs. off, or alternately, 1 vs. 0. Technically any object that can have two distinct states can remember one bit of information. The ASCII code defines 128 characters and a mapping of those characters onto the numbers 0...127. For example, the letter 'A' is assigned 65 in the ASCII table. Expressed in binary, that's 2 6 + 2 0 (64 + 1), and so the byte that represents 'A' is: 0 1 0 0 0 0 0 1
2	**Nibble**	A group of 4 bits is called nibble.
3	**Byte**	A group of 8 bits is called byte. A byte is the smallest unit, which can represent a data item or a character.
4	**Word**	A computer word, like a byte, is a group of fixed number of bits processed as a unit, which varies from computer to computer but is fixed for each computer. The length of a computer word is called word-size or word length. It may be as small as 8 bits or may be as long as 96 bits. A computer stores the information in the form of computer words.
5	**Kilobyte (KB)**	1 KB = 1024 Bytes
6	**Megabyte (MB)**	1 MB = 1024 KB
7	**Gigabyte (GB)**	1 GB = 1024 MB
8	**Terabyte (TB)**	1 TB = 1024 GB
9	**PetaByte (PB)**	1 PB = 1024 TB

VIRUSES

VIRUS

A virus is a Program that can infect other program by modifying them to include possible evolved version of itself.

A computer virus is a special case of malicious logic (programs that act in violation of the security policy).

Computer viruses are malicious codes or programs that are usually installed on your computer without your knowledge and against your wishes.

The Trojan horse is the most general form of malicious logic. Viruses require a host, and their goal is to infect other files so that the virus can "live" longer. Some viruses perform destructive actions although this is not necessarily the case. Many viruses attempt to hide from being discovered. All viruses make copies of themselves, infecting boot sectors, programs, or "data files" as the opportunity arises.

Viruses can be distributed in several ways. In the early days of computers, viruses were spread by sharing infected floppy disks. Now, due to the ease in which files can be shared over the Internet, viruses are able to spread much more quickly. One of the most common ways to send a virus is through e-mail attachments. Security experts recommend that you never open an e-mail attachment unless you have first scanned it with antivirus software to determine that it is virus-free. Experts also recommend that unless you know the sender and have been expecting the e-mail attachment, it is best to delete the attachment without ever opening it. File-sharing services are another source for these types of problems.

Worms are similar to viruses because they are also malicious programs that spread from computer to computer; however, unlike viruses, worms are able to do this without any human interaction and are able to replicate themselves so numerous copies can be sent. Worms can burrow into your e-mail address book, or locate e-mail addresses on files saved on your hard drive, then send themselves out without any help from you.

Viruses can be programmed to do many kinds of harm, including the following:

» Copy themselves to other programs or areas of a disk.

» Replicate as rapidly and frequently as possible, filling up the infected system's disks and memory, rendering the system useless.

» Display information on the screen.

» Modify, corrupt or destroy selected files.

» Erase the contents of entire disks.

TYPES OF COMPUTER VIRUSES

Trojans

Trojans are programs that perform some unwanted action while pretending to be useful. Most Trojans activate when they are run and sometimes destroy the structure of the current drive (FATs, directories, etc.).

backdoor trojan

A special type is the backdoor Trojan, which often does not do anything overtly destructive, but sets your computer open for remote control and unauthorized access.

Spyware

Spyware is software designed to capture personal and confidential information that resides on your system and send it elsewhere. It has quickly become as large a problem as viruses. Spyware's primary threat is to your privacy and confidentiality. Although spyware is not usually intended to harm your system, it can sometimes have that effect on it.

Adware

Adware is spyware that tracks your Internet browsing and can install malicious cookies on your computer.

Bimodal, Bipartite, or Multipartite Viruses

This type of virus can infect both files and the boot sector of a disk.

Bombs

The two most prevalent types of bombs are time bombs and logic bombs. A time bomb hides on the victim's disk and waits until a specific date (or date and time) before running. A logic bomb may be activated by a date, a change to a file, or a particular action taken by a user or a program. Many experts do not classify bombs as viruses, but some do. Regardless, bombs are treated as viruses because they can cause damage or disruption to a system.

Boot Sector Viruses

Regarded as one of the most hostile types of virus, a boot sector virus infects the boot sector of a hard or floppy disk. This area of the disk stores essential files the computer accesses during startup. The virus moves the boot sector's data to a different part of the disk. When the computer is started, the virus copies itself into memory where it can hide and infect other disks. The virus allows the actual boot sector data to be read as though a normal start-up were occurring.

Cluster Viruses

This type of virus makes changes to a disk's file system. If any program is run from the infected disk, the program causes the virus to run as well. This technique creates the illusion that the virus has infected every program on the disk.

Key loggers are another type of spyware. In this case, a software program records every keystroke made on the computer. Key loggers can capture all sorts of confidential information this way—

passwords, credit card numbers, bank account numbers, and so on—and then relay this information elsewhere. Entire e-mail messagesand instant messaging conversations can be recorded this way too.

Denial-of-service (DOS)

Denial-of-service (DOS) tools are used to bombard other computers with connection attempts to such a degree that the computer that is under attack cannot handle the traffic load, and legitimate requests are neglected. A special case of denial-of service is the so-called "Distributed Denial-of-Service", or DDOS. DDOS occurs when several machines start a coordinated attack against the same target.

Nukers

Nukers send malformed network requests to try to confuse the attacked machine and cause a crash.

Mail bombers

Mail bombers are pretty self-explanatory – they are used to annoy people by filling up their mailbox. E-mail viruses can be transmitted via email messages sent across private networks or the Internet. Some e-mail viruses are transmitted as an infected attachment—a document file or program that is attached to the message. This type of virus is run when the victim opens the file that is attached to the message. Other types of e-mail viruses reside within the body of the message itself. To store a virus, the message must be encoded in HTML format. Once launched, many e-mail viruses attempt to spread by sending messages to everyone in the victim's address book; each of those messages contains a copy of the virus.

File-Infecting Viruses.

This type of virus infects program files on a disk (such as .exe or .com files). When an infected program is launched, the virus's code is also executed.

Joke Programs.

Joke programs are not viruses and do not inflict any damage. Their purpose is to frighten their victims into thinking that a virus has infected and damaged their system. For example, a joke program may display a message warning the user not to touch any keys or the computer's hard disk will be formatted.

Macro Viruses

A macro virus is designed to infect a specific type of document file, such as Microsoft Word or Excel files. These documents can include macros, which are small programs that execute commands. (Macros are typically used to issue program-specific commands, but they also can issue certain operating-system commands.) A macro virus, disguised as a macro, is embedded in a document file and can do various levels of damage to data, from corrupting documents to deleting data.

Polymorphic, Self-Garbling, Self-Encrypting, or Self-Changing Viruses.

This type of virus can change itself each time it is copied, making it difficult to isolate.

Stealth Viruses

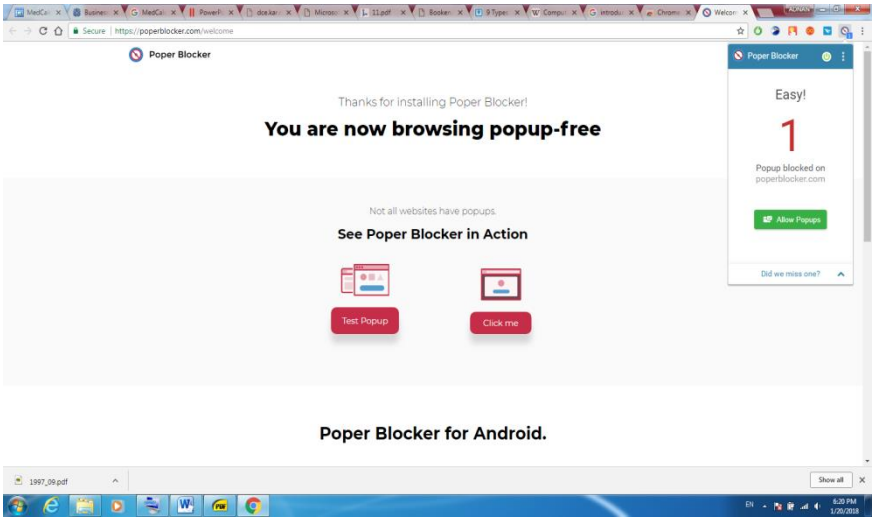
These viruses take up residence in the computer's memory, making them hard to detect. They also can conceal changes they make to other files, hiding the damage from the user and the operating system.

HOW CAN YOU AVOID BEING A VICTIM?

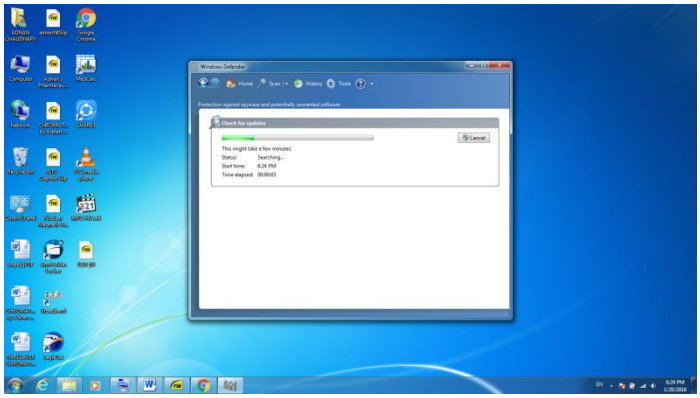
To minimize the risk of having spyware installed on your computer, there are some practical precautions you can take. One of the most prevalent methods of spreading spyware is through file-sharing services, such as Morpheus or Kazaa. Not only can the file-sharing software include spyware, but often the files you think you are downloading for free are infected too. Although it's tempting to get the newest song or video for free from such a site, don't risk it!

This problem can be avoided if you use one of the legitimate, pay-as-you- go file-sharing services such as iTunes or the reincarnated Napster. Additionally, be cautious when you download and install freeware or shareware software. Make sure you deal with a reputable software publisher, scan the downloaded software for viruses and spyware, and read the licensing agreement. Some licensing agreements actually include information about additional software that will be automatically installed if you accept it.

Another way to prevent spyware is to avoid pop-up and banner ads whenever possible. You should never click on them. Often the "No Thanks" button is just a ruse to get you to click it and enable the spyware installation. Close pop-up ads by clicking the Close button in the top right corner. Even better, installing pop-up blocking software can help to eliminate this risk almost entirely.



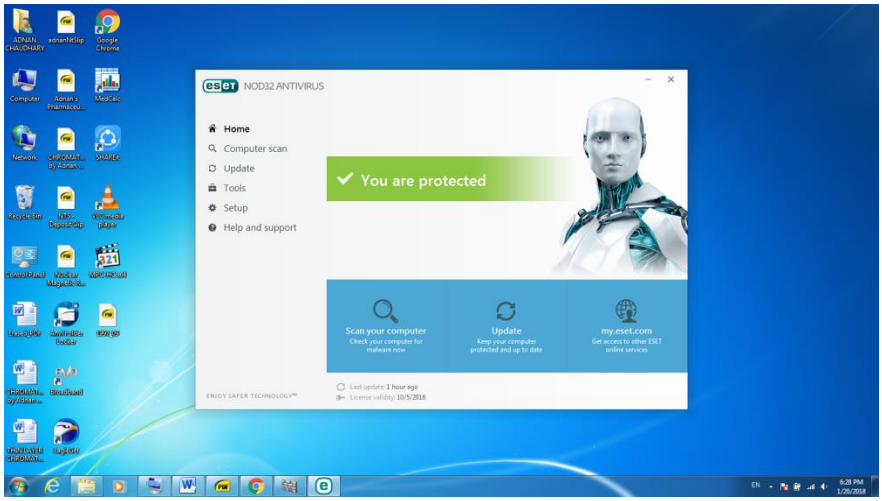
If you are running the most recent version of Windows you already have a pop-up blocker available to you. You can view the pop-up blocker settings for Windows XP in Figure and access this dialog box through Internet Explorer's Tools menu. Many popular search engines, such as Google and Yahoo!, also include pop-up blocking features in their tool- bars, which you can download at no charge. It is also wise to avoid questionable Web sites, because some of them can install spyware on your system just by visiting the



site.

Software updates and patches—Keeping your operating system and software up-to-date is critical. Software manufacturers are constantly on the lookout for security threats, and they issue updates and patches to help protect your system. Check for these and install them regularly. Software manufacturers have begun to implement auto-mated procedures to check and install such updates. If your computer has this capability, it's a good idea to use this feature.

Antivirus and antispyware software—Antivirus software is a utility program used to search your hard drive and files for viruses, and remove those that are found. Antispyware software works in a similar fashion, but searches for spyware rather than viruses. No computer should be without this protection. Many users erroneously think that because they aren't regularly online or use only a slow dial-up connection, they aren't a target. Nothing could be further from the truth! Recent studies show more than two-thirds of all computer users have some form of virus or spyware on their system.



There are a variety of antivirus and antispyware products available. ome well-known antivirus products include Norton AntiVirus (www.symantec.com), McAfee VirusScan (www.mcafee.com), and AVG Anti-Virus (www.grisoft.com). Antispyware products include eTrust PestPatrol (www.pestpatrol.com), Ad-Aware (www.lavasoft.com), and Spybot Search & Destroy (www.safer-networking.org). You can search for other products at popular download sites such as Download.com (www.download.com) or Tucows (www.tucows.com) but you should be sure to read the software reviews and evaluate their usefulness before downloading or installing them.

Personal firewalls—Firewalls may be software programs or hardware devices, although their purpose is the same—to prevent unauthorized access to your computer. When a firewall is installed properly, it can make your computer invisible to hackers and other invaders. Not only can a good firewall help prevent infections and identity theft; it can also prevent hackers from accessing your computer and turning it into a zombie. A zombie computer is one that can be controlled remotely and can be used to help spread viruses, spyware, or junk e-mail known as spam.

COMPUTING SCIENCE RESEARCH METHODOLOGIES

Computing Science researchers use several methodologies to tackle questions within the discipline. This discussion starts by listing several of these methodologies. The idea is not to classify researchers or projects in each of these methodologies or to be exhaustive. Tasks performed by a single researcher fall within different methodologies. Even the activities required to tackle a single research question may include several of these methodologies.

METHODOLOGIES

The following list of methodologies is intended to organize the discussion of the approach required by each of them.

Formal

In Computing Science, formal methodologies are mostly used to prove facts about algorithms and system. Researchers may be interested on the formal specification of a software component in order to allow the automatic verification of an implementation of that component. Alternatively, researchers may be interested on the time or space complexity of an algorithm, or on the correctness or the quality of the solutions generated by the algorithm.

Experimental

Experimental methodologies are broadly used in CS to evaluate new solutions for problems. Experimental evaluation is often divided into two phases. In an exploratory phase the researcher is taking measurements that will help identify what are the questions that should be asked about the system under evaluation. Then an evaluation phase will attempt to answer these questions. A well-designed experiment will start with a list of the questions that the experiment is expected to answer.

Build

A "build" research methodology consists of building an artifact — either a physical artifact or a software system — to demonstrate that it is possible. To be considered research, the construction of the artifact must be new or it must include new features that have not been demonstrated before in other artifacts.

Process

A process methodology is used to understand the processes used to accomplish tasks in Computing Science. This methodology is mostly used in the areas of Software Engineering and Man-Machine Interface which deal with the way humans build and use computer systems. The study of processes may also be used to understand cognition in the field of Artificial Intelligence.

Model

The model methodology is centered on defining an abstract model for a real system. This model will be much less complex than the system that it models, and therefore will allow the researcher to better understand the system and to use the model to perform experiments that could not be performed in the system itself because of cost or accessibility. The model methodology is

often used in combination with the other four methodologies. Experiments based on a model are called simulations. When a formal description of the model is created to verify the functionality or correctness of a system, the task is called model checking.

In the rest of this document we will attempt to provide advice about each of these methodologies. But first, lets consider some general advice for research. New and insightful ideas do not just happen on an empty and idle mind. Insightful research stems from our interaction with other people that have similar interests. Thus it is essential for all researchers to be very involved in their communities, to attend seminars, participate in discussions, and, most importantly, to read widely in the discipline.

It is also important to keep a record of our progress and to make notes of the ideas that we have along the way. The brain has an uncanny nag to come back to ideas that we have considered in the past — and often we do not remember what were the issues that led us not to pursue that idea at that time. Thus a good record keeping system — a personal blog, a notebook, a file in your home directory — is a very important research tool. In this log we should make notes of the papers that we read, the discussions that we had, the ideas that we had, and the approaches that we tried.

Once a student starts having regular meetings with a thesis supervisor, it is a good idea to write a summary of each of these meetings. The student may send the summaries to the supervisor or may keep it to herself. After several months, revisiting these meeting logs will help review the progress and reassess if she is on track with her plans towards graduation.

In research, as in other areas of human activities, an strategy to get things done is as important as great visions and insightful ideas. Whenever working with others, it is important to define intermediate milestones, to establish clear ways to measure progress at such milestones and to have clear deadlines for each of them. Collaborators will be multitasking and will dedicate time to the tasks that have a clear deadline.

1.1 FORMAL METHODOLOGY

A formal methodology is most frequently used in theoretical Computing Science. Johnson states that Theoretical Computer Science (TCS) is the science that supports the field of computing. TCS is formal and mathematical and it is mostly concerned with modeling and abstraction. The idea is to abstract away less important details and obtain a model that captures the essence of the problem under study. This approach allows for general results that are adaptable as underlying technologies and application changes, and that also provides unification and linkage between seemingly disparate areas and disciplines. TCS concerns itself with possibilities and fundamental limitations. Researchers in TCS develop mathematical techniques to address questions such as the following. Given a problem, how hard is it to solve? Given a computational model, what are its limitations? Given a formalism, what can it express?

TCS is not only concerned with what is doable today but also with what will be possible in the future with new architectures, faster machines, and future problems. For instance Church and Turing gave formalisms for computation before general-purpose computers were built.

TCS researchers work on the discovery of more efficient algorithms in many areas including combinatorial problems, computational geometry, cryptography, parallel and distributed computing. They also answer fundamental questions about computability and complexity. They have developed a comprehensive theoretical frame to organize problems into complexity classes, to establish lower bounds for time and space complexity for algorithms, and to investigate the limits of computation.

The best practical advice for new researchers in the area of formal research methods is to practice solving problems and to pay attention to detail. The general advice for researchers in computing science, know the literature, communicate with colleagues in the area, ask questions, think, applies to formal method research as well. Problem solving can be risky but also very rewarding. Even if you don't solve your original problem, partial results can lead to new and interesting directions.

The skills and the background knowledge that formal method researchers find useful include: problem-solving, mathematical proof techniques, algorithm design and analysis, complexity theory, and computer programming.

1.2 EXPERIMENTAL METHODOLOGY

The Computing Science literature is littered with experimental papers that are irrelevant even before they are published because of the careless fashion in which the experiments were conducted and reported. Often the authors themselves could not reproduce the experiments only a few weeks after they rushed to obtain the experimental results for a mid-night conference deadline. Here is some general advise to help preventing you from producing worthless experimental papers.

1.2.1 Record Keeping

Good record keeping is very important in experimental work. Computing-Science researcher's record keeping tend to be surprisingly lax. Because experiments are run on computers, inexperienced researchers have a tendency to think that they can rerun the experiments later if they need to. Thus they tend to not be as careful as they should be about labeling and filing the results in ways that will make it possible to retrieve and check them later. Sometimes it is even difficult for a researcher to reproduce his own experiments because he does not remember in which machine it was run, or which compiler was used, or which flags were on, etc.

The reality is that computers are very transient objects. The computer that is in the lab today is likely to not be available in this configuration in just a couple of months. Thus experimental computing science would greatly benefit if each experimental computing scientist would treat her experiment with the same care that a biologist treats a slow-growing colony of bacteria. Annotating, filing, and documenting are essential for the future relevance of an experimental scientist's work.

1.2.2 Experimental Setup Design

Speed is an important factor during the exploratory phase of an experimental work. Thus this phase usually proceeds with less care than it should. Once this exploratory phase is over, a researcher should stop and document the findings and carefully describe the experimental setup, as well as the characteristics of the hardware and software that will be used for the evaluation phase. What are the questions that the experimental work is expected to answer?

What are the variables that will be controlled? What variables may affect the results of the experiment but are not under the control of the researcher? What measures will be taken to account for the variance due to these variables — will the results be statistically significant? Is the experiment documented in a fashion that would allow other researchers to reproduce it?

1.2.3 Reporting Experimental Results

When reporting the results of an experimental evaluation, it is important to state clearly and succinctly, in plain English, what was learned from the experiments. Numbers included in a paper or written into a report should be there to provide an answer or make a point.

Graphical or table representations should be carefully selected to underscore the points that the researcher is making. They should not mislead or distort the data. Analyzing data based only on aggregation is very dangerous because averages can be very misleading. Thus, even if the raw data is not presented in a paper, the author should examine the raw data carefully to gain further insight into the results of the experiments. The numerical results presented in tables and graphs should be accompanied with a carefully written analytical discussion of the results. This discussion should not simply repeat in words the results that are already shown in the tables and graphs. This discussion should provide insight on those results, it should add knowledge that the researcher gained and that is not in those numbers. Alternatively this discussion should attempt to explain the results presented.

1.3 BUILD METHODOLOGY

Whenever a research question leads to the building of a software system, the researchers involved should consider the following set of good practices: Design the software system . No matter how simple the system is, do not allow it to evolve from small pieces without a plan. Think before you build. Most importantly, consider a modular approach - it simplifies testing. Testing is also simplified by choosing text-based data and communication formats. Defining small interfaces increases flexibility and reuse potential.

Reuse components. Are some needed software components already (freely) available? If yes, using such components can save time. When deciding which components to reuse consider the terms of use attached with them. The components that you reuse in the system can have implications on the software license under which the new system can be distributed. For instance, if a component distributed under the GNU Public License

(GPL) is used in a software system; the entire system will have to be distributed under GPL. Choose an adequate programming language. Often researchers want to use a programming language that they already know to minimize the time invested on learning a new language. However it may pay off to learn new languages that are more adequate for the building of an specific system. Important factors to consider when selecting a programming language include: required run-time speed (compiled vs. interpreted languages), expressiveness (imperative vs. functional vs. declarative languages), reliability (e.g. run-time checks, garbage collection), and available libraries.

Consider testing all the time. Don't wait to test the entire system after it is built. Test modules first. Keep a set of input/output pairs around for testing. This way future changes can be tested when they are introduced. Consider building an automated testing infrastructure that compares the program's output on a set of input data with correct outputs and also measures run time. Set this

automated testing infrastructure to run automatically daily/weekly to notify the builders about troublesome changes immediately.

Documentation is crucial in any software system. Good computer programs must be well documented. Supervisors, outside users, and fellow students who may extend the system in the future need to be able to understand the code without much trouble. Even when there is a single developer there are advantages to use of a version control system, such as Concurrent Versions System (CVS). CVS gives the developer, and anyone that needs casual access to the code and documentation, easy access to the set of current files from multiple locations. Moreover, CVS allows access to previous versions in case of changes that introduce bugs.

Once the software system is functional, researchers should compare its functionality and/or performance with that of existing systems to verify that the claim(s) that they want to make about the system still hold. Often, the runtime/space requirements dependent on input size are reported on commonly used test sets. Architecture-independent measures, such as the number of nodes visited on a graph per unit of time, should be reported based on wall-clock time or actual memory consumption to simplify comparison with other systems. Results should be reported using statistics - such as percentiles (e.g. quartiles min 25% median 75% max) - that don't depend on unjustified distribution assumptions.

1.4 PROCESS METHODOLOGY

Process methodologies are most useful in the study of activities that involve humans. Examples of such activities in Computing Science include the design and construction of software systems — large or small, the design and evaluation of human-computer interactions, and the understanding of cognitive processes. More recently the creation of interactive games has been studied extensively. These activities often involve studies with human subjects.

1.4.1 Software process

The construction and evolution of a large-scale software system is challenging for many reasons, including complexity, large teams, long lifetimes, changing needs, quality trade-offs, and emerging technology. One aim of software engineering research is to devise methods or processes to put more discipline into how such systems are developed and ultimately raise their quality. We can study how systems are currently put together or managed, and discover proven designs, repeated patterns, or recurring strategies. We might discover strong trends that could be useful predictive indicators about some system property, like defect density.

Alternatively, we can step back and abstract the system to better reason about some aspect, like resource usage, and devise a prescriptive software process to prevent certain classes of defects from occurring in the first place.

Such generalizations can be codified in various procedural ways — such as best-practice guides, pattern languages, application frameworks, process models, and development tools — to be applied by the software engineering practitioner. Indeed, many scientific techniques when applied to software can provide useful fodder for a process or tool to be applied in reality. Examples of techniques include the development of models of software structure and behavior; experiments of a system with users and developers; and the development of a proof-of-concept or useful component.

Experience with these techniques can lead to processes for model-driven development, methods to enable improved collaboration, and tool-based assistance for both. Processes and tools themselves may evolve together as we better understand the context where they are situated.

There are still difficult social, cognitive, and technical barriers to the adoption of a new process or tool in an existing setting. A tool that either is conceptually complex, forces a heavy-handed approach, requires significant training, or ignores legacy issues could be doomed to failure. In-the-field case studies may be needed to convincingly demonstrate real impact. In scoping out a research problem in software engineering, one needs to be aware of issues beyond the technical ones, without being too bogged down in non-essentials.

Important is having clear assumptions, a well-stated, verifiable research question, and a suitably designed evaluation. The definition of a research problems resembles the definition of a software module: it should have a clear interface, a well-defined function, and meaningful tests.

1.4.2 Methodological Issues

In designing a study to evaluate Human-Computer Interfaces (HCI) or to assess how a software engineering tool or methodology, it is important to appreciate that there is a large body of literature on effective ways to design surveys and to conduct structured interviews. Such literature should be consulted even if the survey is exploratory in nature and not aimed at hypothesis testing. For instance, psychologists describe "anchor effects" that may impact how a person chooses a rating on a rating scale to make a preference judgment. The psychometric literature should be explored before designing a study that employs surveys or interviews. Natural sources of information range from the web to meeting with someone from the Psychology Department for advice, especially if the data produced by the survey is a crucial part of the research.

The University of Alberta has a Population Research Lab that offers consultation on the design of reliable surveys. It may be useful to contact someone in that centre (there may be a fee, so discuss this with your supervisor). Once results are collected from a properly design survey or study, a researcher should ensure that appropriate statistical techniques are used for the analysis of these results.

1.4.3 Cognitive Modeling

Another area of Computing Science that centers around processes and human subjects is the study of cognition. Cognitive Scientists develop computer models of hypothesized cognitive processes, just as physicists develop computer models of hypothesized physics processes, or meteorological scientists develop computer models of hypothesized atmospheric processes. In all cases, there are observed empirical data, and the game is to develop a process model that accounts for the observed data, i.e., a set of proposed processes that operate in particular ways to produce the data of interest. Interesting and important models address a general issue that may lead to the understanding of underlying mechanisms that produce the existing data.

For instance, consider a data set containing some observed memory results such as accuracy and reaction time. It is possible to write a computer program that produces the same results given some particular input, and declare it to be a model. However, is it an interesting model? Does this model inform the community in some more general way? Are the observations obtained with this

model extendable to other results (in principle), that embodies some general assumptions that themselves are testable. Otherwise, it is a one-off computer program whose existence does not really contribute to any larger story. This is true whether one models using a neural network paradigm or a higher level architecture with various assumptions about how control, memory, and learning processes interact.

The principle that there must be a bigger story behind an implementation applies broadly to Computing Science enterprises. A researcher that decides to create a program that does X should be all along concerned with the research questions that this program will answer or will raise.

Researchers develop computer models to make explicit all the assumptions that constitute a given explanation. To say that information is retrieved "on the basis of similarity" requires one to offer some computational account of "similarity". A researcher must describe how it is computed, and what other factors might affect that computation. To say that certain kinds of information are "preferred" over other sorts of information may be appropriate for a descriptive account of observed results. However, a computational model of the processes by which preference is computed requires making explicit the factors, which are often context- dependent, that determine preference. The translation of assumptions into a particular computational process typically leads to predictions that themselves are verifiable in other experiments. In fact, a computational model that does not — as a byproduct — leads to new testable hypotheses is not very interesting or useful to the broader enterprise of understanding a complex system.

So how does one get around writing a one-off computer program to model some particular set of results? First, decide what kind of computational paradigm to operate in. This is a personal choice. You can operate with a neural net paradigm; you can operate within a higher-level architecture paradigm. Adopt a paradigm and situate your work within the community that uses that paradigm to increase the likelihood that the new research results produced will actually contribute to some larger understanding about an issue.

1.5 MODEL METHODOLOGY

Modeling is the purposeful abstraction of a real or a planned system with the objective of reducing it to a limited, but representative, set of components and interactions that allow the qualitative and quantitative description of its properties. Strictly speaking, modeling is a methodological aspect of science. Modeling is not the object of the research, it is part of an arsenal of instruments used by researchers to study and understand the research's object.

The action of building a model, called modeling is driven by the study that will use the model. Hence, there is no single modeling approach applicable to all systems. Rather, scientists build models that capture important aspects of a system and gloss over — either completely ignore or just approximate — the aspects that have lesser (or no) impact to their intended study. The decision of which aspects are important and which ones have lesser impact is itself part of the modeling strategy. Misleading outcomes are produced by models that eliminate what is important or that over-emphasize what is of lesser impact. Extensive arguments in the communities that depend on modeling center on decisions about the actual model used. Modeling should be seen as an evolving process that is coupled to a particular sub-discipline.

For instance, the workload — e.g. the form and pattern of requests — to a web server is usually modeled using particular distributions. These distributions are used to express time between request arrivals, the sizes of documents retrieved, the probabilistic relationship between successive item requests, and other such features. As the technologies around the World-Wide Web evolve, so does the corresponding traffic. Thus the issue what is a good model to express the traffic to web servers is by definition a moving target. The corresponding community of researchers needs to be in continuous dialog about the models used and needs to update them as the studied phenomena evolve. In some disciplines a good model is one that anticipate changes by incorporating degrees of freedom that allow it to be applied even when some aspects of the problem change over time. Models can be expressed in diverse ways. A model can be described by text and block diagrams to specify how a system is constructed from a collection of interacting components.

In Computing Science, a model is often described by a formal language designed specifically to describe particular kinds of models, or it is embedded in a computer program that simulates the behavior of a system. Formal models are usually geared toward specific qualitative and quantitative aspects of a system, while simulation is open ended. Simulation often lacks the power to make definite statements about properties of the system. For instance, the results of simulations may not be used to prove that a deadlock never develops in a concurrent system.

The best approach for new researchers in areas where models are required is to study publications that disclose the full details of the model used, or that describe models for similar systems. Unfortunately, often the presentation of models is abbreviated in papers due to page restrictions. Authors often assume that the reader is aware of usual modeling assumptions in a particular domain. A careful researcher that sees no details about the model used in a simulation based study will keep in mind that certain properties (qualitative or quantitative) of the study might be directly influenced by sloppy modeling. Such researcher will have a keen eye (developed through experience) to spot discrepancies that go hand-in- hand with sloppy modeling. It is also a good idea to reflect on, spot shortcomings, and to report them in the literature when models are found inadequate or incorrectly assume to "do their job" when they don't. Proposing a new (better) model is also a very important contribution to the community.

SYSTEMS ANALYSIS AND DESIGN

Systems development is systematic process which includes phases such as planning, analysis, design, deployment, and maintenance.

- Systems analysis
- Systems design

SYSTEMS ANALYSIS

It is a process of collecting and interpreting facts, identifying the problems, and decomposition of a system into its components. System analysis is conducted for the purpose of studying a system or its parts in order to identify its objectives. It is a problem solving technique that improves the system and ensures that all the components of the system work efficiently to accomplish their purpose. Analysis specifies what the system should do.

SYSTEMS DESIGN

It is a process of planning a new business system or replacing an existing system by defining its components or modules to satisfy the specific requirements. Before planning, you need to understand the old system thoroughly and determine how computers can best be used in order to operate efficiently. System Design focuses on how to accomplish the objective of the system.

System Analysis and Design (SAD) mainly focuses on:

- Systems
- Processes
- Technology

WHAT IS A SYSTEM?

The word System is derived from Greek word Systema , which means an organized relationship between any set of components to achieve some common cause or objective. A system is "an orderly grouping of interdependent components linked together according to a plan to achieve a specific goal."

Constraints of a System

A system must have three basic constraints:

1. A system must have some structure and behavior which is designed to achieve a predefined objective.

2. Interconnectivity and interdependence must exist among the system components.

3. The objectives of the organization have a higher priority than the objectives of its subsystems.

For example, traffic management system, payroll system, automatic library system, human resources information system.

PROPERTIES OF A SYSTEM

A system has the following properties:

Organization

Organization implies structure and order. It is the arrangement of components that helps to achieve predetermined objectives.

Interaction

It is defined by the manner in which the components operate with each other. For example, in an organization, purchasing department must interact with production department and payroll with personnel department.

Interdependence

Interdependence means how the components of a system depend on one another. For proper functioning, the components are coordinated and linked together according to a specified plan. The output of one subsystem is the required by other subsystem as input.

Integration

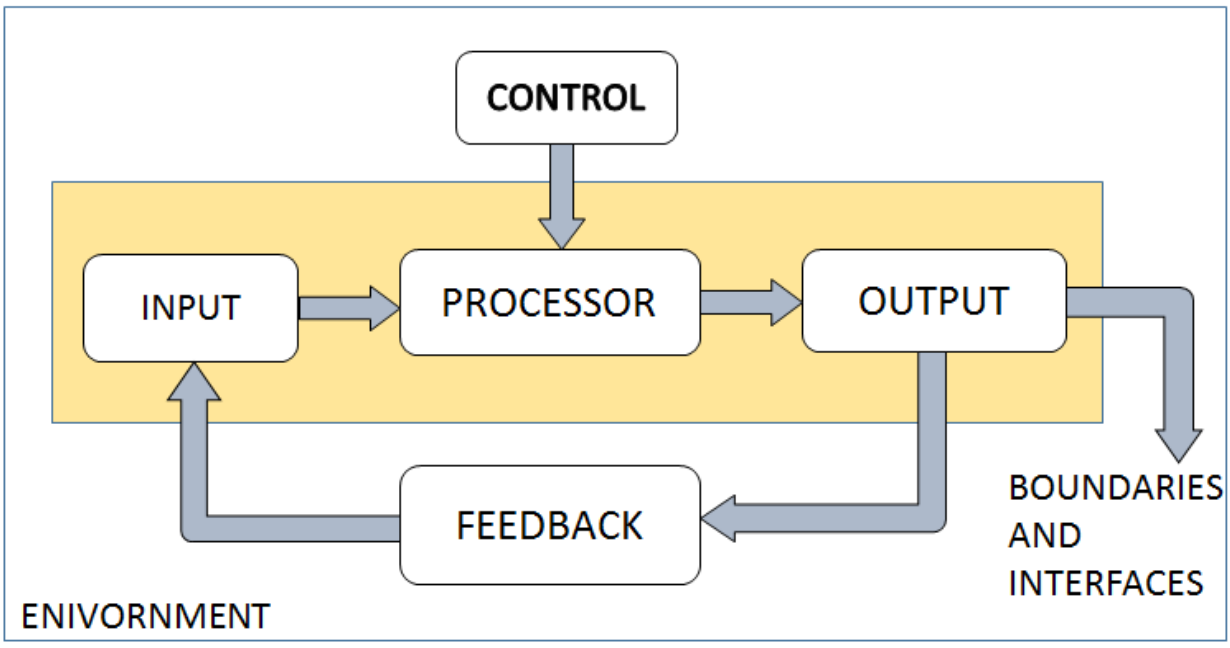
Integration is concerned with how system components are connected together. It means that the parts of the system work together within the system even if each part performs a unique function.

Central Objective

The objective of system must be central. It may be real or stated. It is not uncommon for an organization to state an objective and operate to achieve another. The users must know the main objective of a computer application early in the analysis for a successful design and conversion.

ELEMENTS OF A SYSTEM

The following diagram shows the elements of a system:



Outputs and Inputs

- The main aim of a system is to produce an output which is useful for its user.
- Inputs are the information that enters into the system for processing.
- Output is the outcome of processing.

Processor(s)

- The processor is the element of a system that involves the actual transformation of input into output.
- It is the operational component of a system. Processors may modify the input either totally or partially, depending on the output specification.
- As the output specifications change, so does the processing. In some cases, input is also modified to enable the processor for handling the transformation.

Control

- The control element guides the system.
- It is the decision–making subsystem that controls the pattern of activities governing input, processing, and output.
- The behavior of a computer System is controlled by the Operating System and software. In order to keep system in balance, what and how much input is needed is determined by Output Specifications.

Feedback

- Feedback provides the control in a dynamic system.
- Positive feedback is routine in nature that encourages the performance of the system.
- Negative feedback is informational in nature that provides the controller with information for action.

Environment

The environment is the "supersystem" within which an organization operates.

- It is the source of external elements that strike on the system.
- It determines how a system must function. For example, vendors and competitors of organization's environment, may provide constraints that affect the actual performance of the business.

Boundaries and Interface

- A system should be defined by its boundaries. Boundaries are the limits that identify its components, processes, and interrelationship when it interfaces with another system.
- Each system has boundaries that determine its sphere of influence and control.
- The knowledge of the boundaries of a given system is crucial in determining the nature of its interface with other systems for successful design.

TYPES OF SYSTEMS

The systems can be divided into the following types:

Physical or Abstract Systems

- Physical systems are tangible entities. We can touch and feel them.
- Physical System may be static or dynamic in nature. For example, desks and chairs are the physical parts of computer center which are static. A programmed computer is a dynamic system in which programs, data, and applications can change according to the user's needs.
- Abstract systems are non-physical entities or conceptual that may be formulas, representation or model of a real system.

Open or Closed Systems

- An open system must interact with its environment. It receives inputs from and delivers outputs to the outside of the system. For example, an information system which must adapt to the changing environmental conditions.
- A closed system does not interact with its environment. It is isolated from environmental influences. A completely closed system is rare in reality.

Adaptive and Non Adaptive System

- Adaptive System responds to the change in the environment in a way to improve their performance and to survive. For example, human beings, animals.
- Non Adaptive System is the system which does not respond to the environment. For example, machines.

Permanent or Temporary System

- Permanent System persists for long time. For example, business policies.
- Temporary System is made for specified time and after that they are demolished. For example, A DJ system is set up for a program and it is dissembled after the program.

Natural and Manufactured System

- Natural systems are created by the nature. For example, Solar system, seasonal system.
- Manufactured System is the man-made system. For example, Rockets, dams, trains.

Deterministic or Probabilistic System

- Deterministic system operates in a predictable manner and the interaction between system components is known with certainty. For example, two molecules of hydrogen and one molecule of oxygen makes water.
- Probabilistic System shows uncertain behavior. The exact output is not known. For example, Weather forecasting, mail delivery.

Social, Human-Machine, Machine System

- Social System is made up of people. For example, social clubs, societies.
- In Human-Machine System, both human and machines are involved to perform a particular task. For example, Computer programming.
- Machine System is where human interference is neglected. All the tasks are performed by the machine. For example, an autonomous robot.

Man–Made Information Systems

- It is an interconnected set of information resources to manage data for particular organization, under Direct Management Control (DMC).
- This system includes hardware, software, communication, data, and application for producing information according to the need of an organization.

Man-made information systems are divided into three types:

- Formal Information System: It is based on the flow of information in the form of memos, instructions, etc., from top level to lower levels of management.
- Informal Information System: This is employee based system which solves the day to day work related problems.
- Computer Based System: This system is directly dependent on the computer for managing business applications. For example, automatic library system, railway reservation system, banking system, etc.

SYSTEM DEVELOPMENT LIFE CYCLE

An effective System Development Life Cycle (SDLC) should result in a high quality system that meets customer expectations, reaches completion within time and cost evaluations, and works effectively and efficiently in the current and planned Information Technology infrastructure .

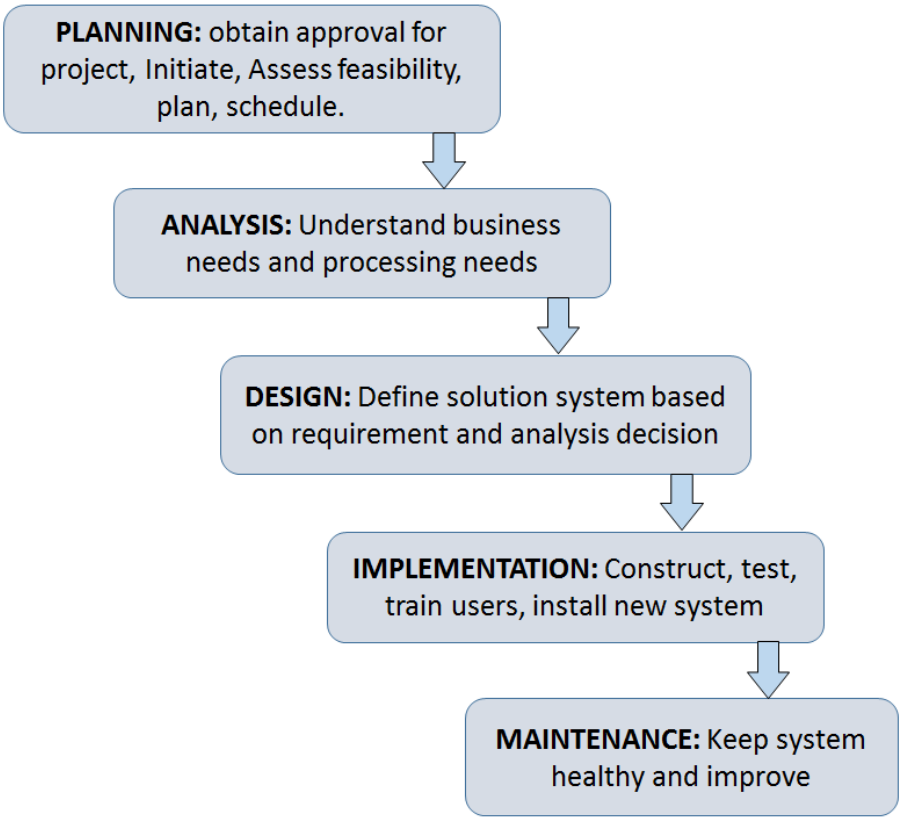
System Development Life Cycle (SDLC) is a conceptual model which includes policies and procedures for developing or altering systems throughout their life cycles.

SDLC is used by analysts to develop an information system. SDLC includes the following activities:

- requirements
- design
- implementation
- testing
- deployment
- operations
- maintenance

PHASES OF SDLC

Systems Development Life Cycle is a systematic approach which explicitly breaks down the work into phases that are required to implement either new or modified Information System.



Feasibility Study or Planning

- Define the problem and scope of existing system.
- Overview the new system and determine its objectives.
- Confirm project feasibility and produce the project Schedule.
- During this phase, threats, constraints, integration and security of system are also considered.
- A feasibility report for the entire project is created at the end of this phase.

Analysis and Specification

- Gather, analyze, and validate the information.
- Define the requirements and prototypes for new system.
- Evaluate the alternatives and prioritize the requirements.
- Examine the information needs of end-user and enhances the system goal.
- A Software Requirement Specification (SRS) document, which specifies the software, hardware, functional, and network requirements of the system is prepared at the end of this phase.

System Design

- Includes the design of application, network, databases, user interfaces, and system interfaces.
- Transform the SRS document into logical structure, which contains detailed and complete set of specifications that can be implemented in a programming language.
- Create a contingency, training, maintenance, and operation plan.
- Review the proposed design. Ensure that the final design must meet the requirements stated in SRS document.
- Finally, prepare a design document which will be used during next phases.

Implementation

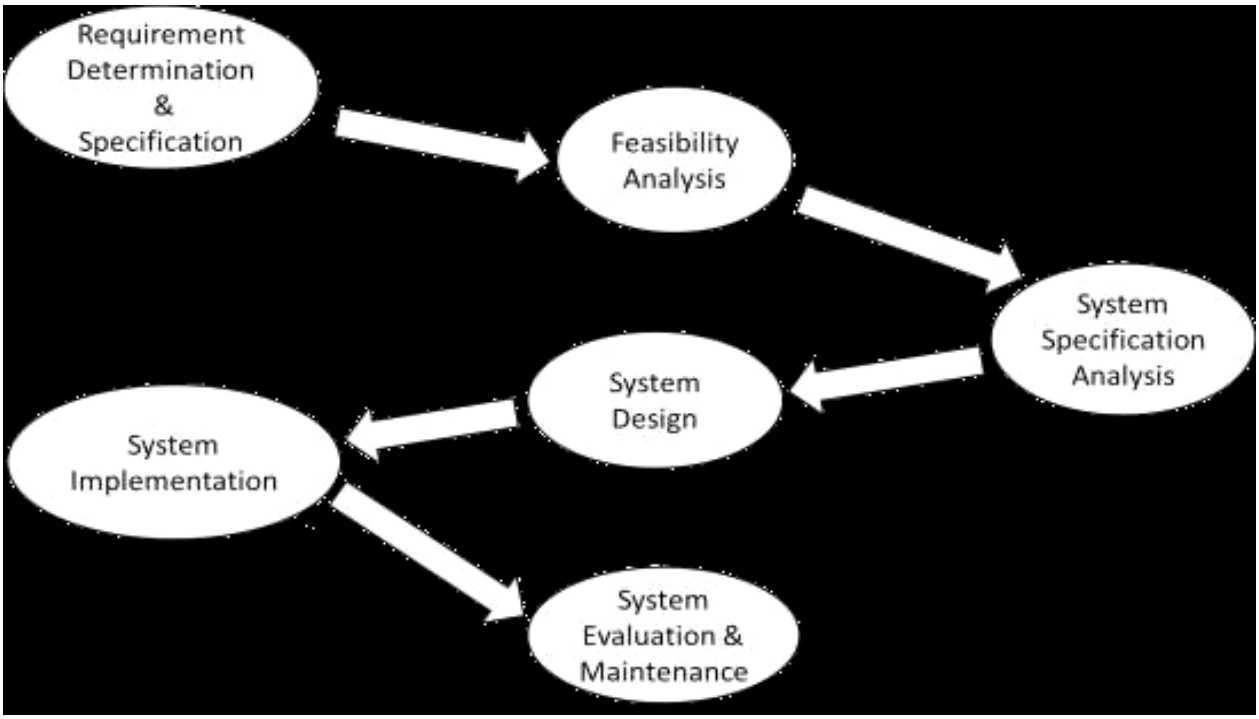
- Implement the design into source code through coding.
- Combine all the modules together into training environment that detects errors and defects.
- A test report which contains errors is prepared through test plan that includes test related tasks such as test case generation, testing criteria, and resource allocation for testing.
- Integrate the information system into its environment and install the new system.

Maintenance/Support

- Include all the activities such as phone support or physical on-site support for users that is required once the system is installing.
- Implement the changes that software might undergo over a period of time, or implement any new requirements after the software is deployed at the customer location.
- It also includes handling the residual errors and resolve any issues that may exist in the system even after the testing phase.
- Maintenance and support may be needed for a longer time for large systems and for a short time for smaller systems.

LIFE CYCLE OF SYSTEM ANALYSIS AND DESIGN

The following diagram shows the complete life cycle of the system during analysis and design phase.



SYSTEM PLANNING

WHAT IS REQUIREMENTS DETERMINATION?

A requirement is a vital feature of a new system which may include processing or capturing of data, controlling the activities of business, producing information and supporting the management.

Requirements determination involves studying the existing system and gathering details to find out what are the requirements, how it works, and where improvements should be made.

MAJOR ACTIVITIES IN REQUIREMENT DETERMINATION

Requirements Anticipation

- It predicts the characteristics of system based on previous experience which include certain problems or features and requirements for a new system.
- It can lead to analysis of areas that would otherwise go unnoticed by inexperienced analyst. But if shortcuts are taken and bias is introduced in conducting the investigation, then requirement Anticipation can be half-baked.

Requirements Investigation

- It is studying the current system and documenting its features for further analysis.
- It is at the heart of system analysis where analyst documenting and describing system features using fact-finding techniques, prototyping, and computer assisted tools.

Requirements Specifications

- It includes the analysis of data which determine the requirement specification, description of features for new system, and specifying what information requirements will be provided.
- It includes analysis of factual data, identification of essential requirements, andselection of Requirement-fulfillment strategies.

INFORMATION GATHERING TECHNIQUES

The main aim of fact finding techniques is to determine the information requirements of an organization used by analysts to prepare a precise SRS understood by user.

Ideal SRS Document should:

- be complete, Unambiguous, and Jargon-free.
- specify operational, tactical, and strategic information requirements.
- solve possible disputes between users and analyst.
- use graphical aids which simplify understanding and design.

VARIOUS INFORMATION GATHERING TECHNIQUES:

Interviewing

Systems analyst collects information from individuals or groups by interviewing. The analyst can be formal, legalistic, play politics, or be informal; as the success of an interview depends on the skill of analyst as interviewer.

It can be done in two ways:

- Unstructured interview: The system analyst conducts question-answer session to acquire basic information of the system.
- Structured interview: It has standard questions which user need to respond in either close (objective) or open (descriptive) format.

Advantages of Interviewing

- This method is frequently the best source of gathering qualitative information.
- It is useful for them, who do not communicate effectively in writing or who may not have the time to complete questionnaire.
- Information can easily be validated and cross checked immediately.
- It can handle the complex subjects.
- It is easy to discover key problem by seeking opinions.
- It bridges the gaps in the areas of misunderstandings and minimizes future problems.

Questionnaires

This method is used by analyst to gather information about various issues of system from large number of persons.

There are two types of questionnaires:

- Open-ended Questionnaires: It consists of questions that can be easily and correctly interpreted. They can explore a problem and lead to a specific direction of answer.
- Closed-ended Questionnaires: It consists of questions that are used when the systems analyst effectively lists all possible responses, which are mutually exclusive.

Advantages of questionnaires

- It is very effective in surveying interests, attitudes, feelings, and beliefs of users which are not co-located.
- It is useful in situation to know what proportion of a given group approves or disapproves of a particular feature of the proposed system.
- It is useful to determine the overall opinion before giving any specific direction to the system project.
- It is more reliable and provides high confidentiality of honest responses.
- It is appropriate for electing factual information and for statistical data collection which can be emailed and sent by post.

Review of Records, Procedures, and Forms

Review of existing records, procedures, and forms helps to seek insight into a system which describes the current system capabilities, its operations, or activities.

Advantages

- It helps user to gain some knowledge about the organization or operations by themselves before they impose upon others.
- It helps in documenting current operations within short span of time as the procedure manuals and forms describe the format and functions of present system.
- It can provide a clear understanding about the transactions that are handled in the organization, identifying input for processing, and evaluating performance.
- It can help an analyst to understand the system in terms of the operations that must be supported.
- It describes the problem, its affected parts, and the proposed solution.

Observation

This is a method of gathering information by noticing and observing the people, events, and objects. The analyst visits the organization to observe the working of current system and understands the requirements of the system.

Advantages

- It is a direct method for gleaning information.
- It is useful in situation where authenticity of data collected is in question or when complexity of certain aspects of system prevents clear explanation by end-users.
- It produces more accurate and reliable data.
- It produces all the aspect of documentation that are incomplete and outdated

Joint Application Development (JAD)

It is a new technique developed by IBM which brings owners, users, analysts, designers, and builders to define and design the system using organized and intensive workshops. JAD trained analyst act as facilitator for workshop who has some specialized skills.

Advantages of JAD

- It saves time and cost by replacing months of traditional interviews and follow-up meetings.
- It is useful in organizational culture which supports joint problem solving.
- Fosters formal relationships among multiple levels of employees.
- It can lead to development of design creatively.
- It Allows rapid development and improves ownership of information system.

Secondary Research or Background Reading

This method is widely used for information gathering by accessing the gleaned information. It includes any previously gathered information used by the marketer from any internal or external source.

Advantages

- It is more openly accessed with the availability of internet.
- It provides valuable information with low cost and time.
- It act as forerunner to primary research and aligns the focus of primary research.
- It is used by the researcher to conclude if the research is worth it as it is available with procedures used and issues in collecting them.

FEASIBILITY STUDY

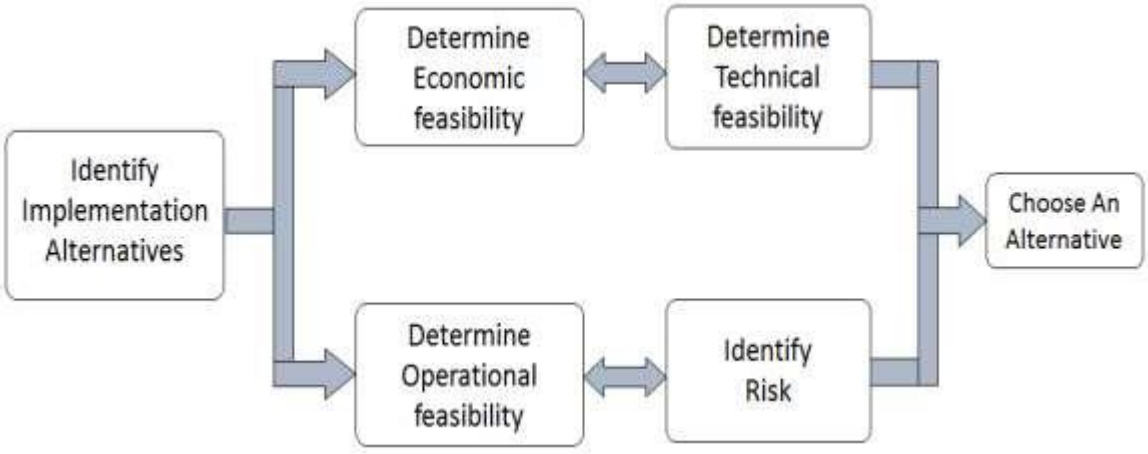
Feasibility Study can be considered as preliminary investigation that helps the management to take decision about whether study of system should be feasible for development or not.

- It identifies the possibility of improving an existing system, developing a new system, and produce refined estimates for further development of system.
- It is used to obtain the outline of the problem and decide whether feasible or appropriate solution exists or not.
- The main objective of a feasibility study is to acquire problem scope instead of solving the problem.
- The output of a feasibility study is a formal system proposal act as decision document which includes the complete nature and scope of the proposed system.

STEPS INVOLVED IN FEASIBILITY ANALYSIS

The following steps are to be followed while performing feasibility analysis:

1. Form a project team and appoint a project leader.
2. Develop system flowcharts.
3. Identify the deficiencies of current system and set goals.
4. Enumerate the alternative solution or potential candidate system to meet goals.
5. Determine the feasibility of each alternative such as technical feasibility, operational feasibility, etc.
6. Weight the performance and cost effectiveness of each candidate system.
7. Rank the other alternatives and select the best candidate system.
8. Prepare a system proposal of final project directive to management for approval.



STUCTURED DATA ANALYSIS

Analysts use various tools to understand and describe the information system. One of the ways is using structured analysis.

WHAT IS STRUCTURED ANALYSIS?

Structured Analysis is a development method that allows the analyst to understand the system and its activities in a logical way.

It is a systematic approach, which uses graphical tools that analyze and refine the objectives of an existing system and develop a new system specification which can be easily understandable by user.

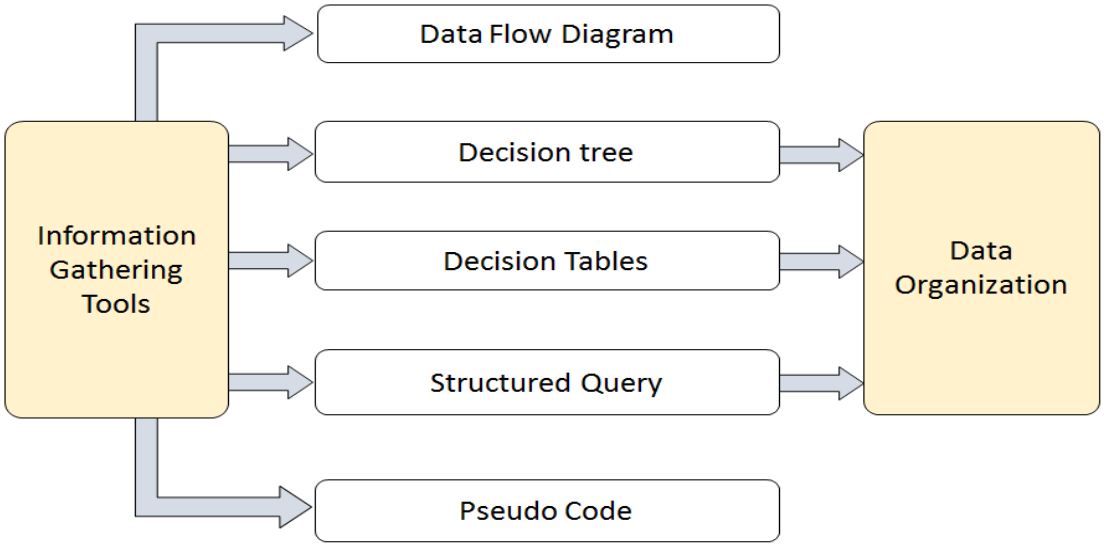
It has following attributes:

- It is graphic which specifies the presentation of application.
- It divides the processes so that it gives a clear picture of system flow.
- It is logical rather than physical i.e., the elements of system do not depend on vendor or hardware.
- It is an approach that works from high-level overviews to lower-level details.

STRUCTURED ANALYSIS TOOLS

During Structured Analysis, various tools and techniques are used for system development. They are:

- Data Flow Diagrams
- Data Dictionary
- Decision Trees
- Decision Tables
- Structured English
- Pseudocode



DATA FLOW DIAGRAMS (DFD) OR BUBBLE CHART

It is a technique developed by Larry Constantine to express the requirements of system in a graphical form.

- It shows the flow of data between various functions of system and specifies how the current system is implemented.
- It is an initial stage of design phase that functionally divides the requirement specifications down to the lowest level of detail.
- Its graphical nature makes it a good communication tool between user and analyst or analyst and system designer.
- It gives an overview of what data a system processes, what transformations are performed, what data are stored, what results are produced and where they flow.

TYPES OF DFD

DFDs are of two types: Physical DFD and Logical DFD. The following table lists the points that differentiate a physical DFD from a logical DFD.

Physical DFD

It is implementation dependent. It shows which functions are performed. It provides low level details of hardware, software, files, and people. It depicts how the current system operates and how a system will be implemented.

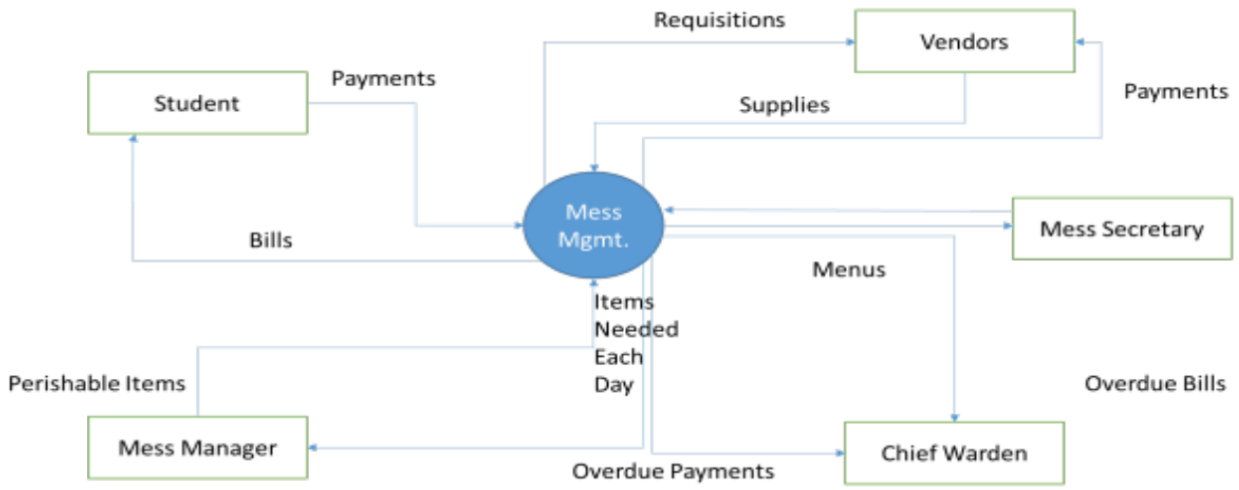
Logical DFD

It is implementation independent. It focuses only on the flow of data between processes. It explains events of systems and data required by each event. It shows how business operates; not how the system can be implemented.

CONTEXT DIAGRAM

A context diagram helps in understanding the entire system by one DFD which gives the overview of a system. It starts with mentioning major processes with little details and then goes onto giving more details of the processes with the top-down approach.

The context diagram of mess management is shown below.

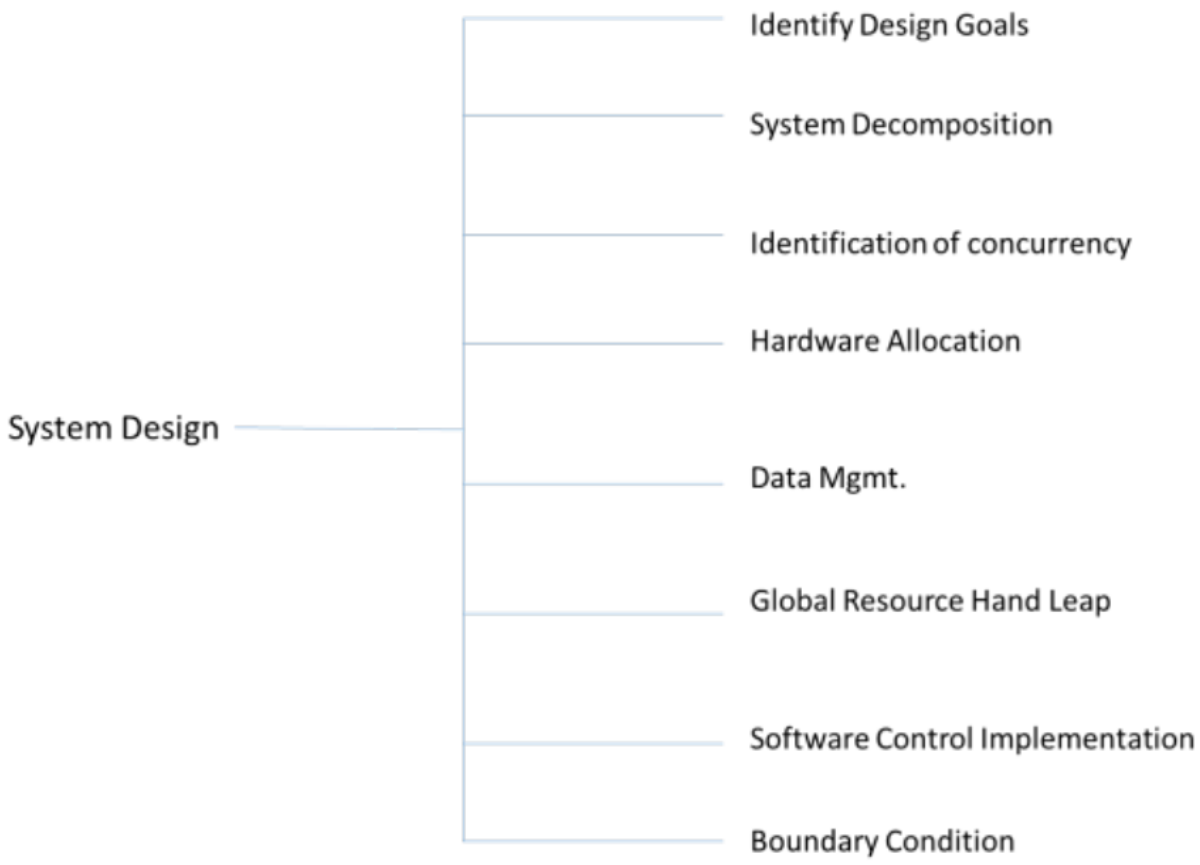


SYSTEM DESIGN

System design is the phase that bridges the gap between problem domain and the existing system in a manageable way. This phase focuses on the solution domain, i.e. "how to implement?"

It is the phase where the SRS document is converted into a format that can be implemented and decides how the system will operate.

In this phase, the complex activity of system development is divided into several smaller sub-activities, which coordinate with each other to achieve the main objective of system development.



INPUTS TO SYSTEM DESIGN

System design takes the following inputs:

- Statement of work
- Requirement determination plan
- Current situation analysis
- Proposed system requirements including a conceptual data model, modified DFDs, and Metadata (data about data).

OUTPUTS FOR SYSTEM DESIGN

System design gives the following outputs:

- Infrastructure and organizational changes for the proposed system.
- A data schema, often a relational schema.
- Metadata to define the tables/files and columns/data-items.
- A function hierarchy diagram or web page map that graphically describes the program structure.
- Actual or pseudocode for each module in the program.
- A prototype for the proposed system.

TYPES OF SYSTEM DESIGN

Logical Design

Logical design pertains to an abstract representation of the data flow, inputs, and outputs of the system. It describes the inputs (sources), outputs (destinations), databases (data stores), procedures (data flows) all in a format that meets the user requirements.

While preparing the logical design of a system, the system analyst specifies the user needs at level of detail that virtually determines the information flow into and out of the system and the required data sources. Data flow diagram, E-R diagram modeling are used.

Physical Design

Physical design relates to the actual input and output processes of the system. It focuses on how data is entered into a system, verified, processed, and displayed as output. It produces the working system by defining the design specification that specifies exactly what the candidate system does. It is concerned with user interface design, process design, and data design.

It consists of the following steps:

- Specifying the input/output media, designing the database, and specifying backup procedures.
- Planning system implementation.
- Devising a test and implementation plan, and specifying any new hardware and software.
- Updating costs, benefits, conversion dates, and system constraints.

Architectural Design

It is also known as high level design that focuses on the design of system architecture. It describes the structure and behavior of the system. It defines the structure and relationship between various modules of system development process.

Detailed Design

It follows Architectural design and focuses on development of each module.

Conceptual Data Modeling

It is representation of organizational data which includes all the major entities and relationship. System analysts develop a conceptual data model for the current system that supports the scope and requirement for the proposed system.

The main aim of conceptual data modeling is to capture as much meaning of data as possible. Most organization today uses conceptual data modeling using E-R model which uses special notation to represent as much meaning about data as possible.

Entity Relationship Model

It is a technique used in database design that helps describe the relationship between various entities of an organization.

FILE ORGANIZATION

It describes how records are stored within a file.

There are four file organization methods:

- Serial: Records are stored in chronological order (in order as they are input or occur). Examples: Recording of telephone charges, ATM transactions, Telephone queues
- Sequential: Records are stored in order based on a key field which contains a value that uniquely identifies a record. Examples: Phone directories
- Direct (relative): Each record is stored based on a physical address or location on the device. Address is calculated from the value stored in the record's key field. Randomizing routine or hashing algorithm does the conversion
- Indexed: Records can be processed both sequentially and non-sequentially using indexes.

FILE ACCESS

One can access a file using either Sequential Access or Random Access. File Access methods allow computer programs read or write records in a file.

Sequential Access

Every record on the file is processed starting with the first record until End of File (EOF) is reached. It is efficient when a large number of the records on the file need to be accessed at any given time. Data stored on a tape (sequential access) can be accessed only sequentially

Direct (Random) Access

Records are located by knowing their physical locations or addresses on the device rather than their positions relative to other records. Data stored on a CD device (direct-access) can be accessed either sequentially or randomly.

Types of Files used in an Organization System

Following are the types of files used in an organization system:

- Master file: It contains the current information for a system. For example, customer file, student file, telephone directory.
- Table file: It is a type of master file that changes infrequently and stored in a tabular format. For example, storing Zipcode.
- Transaction file: It contains the day-to-day information generated from business activities. It is used to update or process the master file. For example, Addresses of the employees.
- Temporary file: It is created and used whenever needed by a system.

- Mirror files: They are the exact duplicates of other files. Help minimize the risk of downtime in cases when the original becomes unusable. They must be modified each time the original file is changed.
- Log files: They contain copies of master and transaction records in order to chronicle any changes that are made to the master file. It facilitates auditing and provides mechanism for recovery in case of system failure.
- Archive file: Backup files that contain historical versions of other files.

DOCUMENTATION CONTROL

Documentation is a process of recording the information for any reference or operational purpose. It helps users, managers, and IT staff, who require it. It is important that prepared document must be updated on regular basis to trace the progress of the system easily.

After the implementation of system if the system is working improperly, then documentation helps the administrator to understand the flow of data in the system to correct the flaws and get the system working. Programmers or systems analysts usually create program and system documentation.

Systems analysts usually are responsible for preparing documentation to help users learn the system. In large companies, a technical support team that includes technical writers might assist in the preparation of user documentation and training materials.

Advantages

- It can reduce system downtime, cut costs, and speed up maintenance tasks.
- It provides the clear description of formal flow of present system and helps to understand the type of input data and how the output can be produced.
- It provides effective and efficient way of communication between technical and nontechnical users about system.
- It facilitates the training of new user so that he can easily understand the flow of system.
- It helps the user to solve the problems such as troubleshooting and helps the manager to take better final decisions of the organization system.
- It provides better control to the internal or external working of the system.

TYPES OF DOCUMENTATIONS

When it comes to System Design, there are following four main documentations:

- Program documentation
- System documentation
- Operations documentation
- User documentation

Program Documentation

- It describes inputs, outputs, and processing logic for all the program modules.
- The program documentation process starts in the system analysis phase and continues during implementation.
- This documentation guides programmers, who construct modules that are well supported by internal and external comments and descriptions that can be understood and maintained easily.

Operations Documentation

Operations documentation contains all the information needed for processing and distributing online and printed output. Operations documentation should be clear, concise, and available online if possible.

It includes the following information:

- Program, systems analyst, programmer, and system identification.
- Scheduling information for printed output, such as report, execution frequency, and deadlines.
- Input files, their source, output files, and their destinations.
- E-mail and report distribution lists.
- Special forms required, including online forms.
- Error and informational messages to operators and restart procedures.
- Special instructions, such as security requirements.

User Documentation

It includes instructions and information to the users who will interact with the system. For example, user manuals, help guides, and tutorials. User documentation is valuable in training users and for reference purpose. It must be clear, understandable, and readily accessible to users at all levels.

The users, system owners, analysts, and programmers, all put combined efforts to develop a user's guide.

A user documentation should include:

- A system overview that clearly describes all major system features, capabilities, and limitations.
- Description of source document content, preparation, processing, and, samples.
- Overview of menu and data entry screen options, contents, and processing instructions.
- Examples of reports that are produced regularly or available at the user's request, including samples.
- Security and audit trail information.
- Explanation of responsibility for specific input, output, or processing requirements.
- Procedures for requesting changes and reporting problems.
- Examples of exceptions and error situations.
- Frequently asked questions (FAQs).
- Explanation of how to get help and procedures for updating the user manual.

System Documentation

System documentation serves as the technical specifications for the IS and how the objectives of the IS are accomplished. Users, managers and IS owners need never reference system documentation. System documentation provides the basis for understanding the technical aspects of the IS when modifications are made.

- It describes each program within the IS and the entire IS itself.
- It describes the system's functions; the way they are implemented, each program's purpose within the entire IS with respect to the order of execution, information passed to and from programs, and overall system flow.

- It includes data dictionary entries, data flow diagrams, object models, screen layouts, source documents, and the systems request that initiated the project
- Most of the system documentation is prepared during the system analysis and system design phases.
- During systems implementation, an analyst must review system documentation to verify that it is complete, accurate, and up-to-date, and including any changes made during the implementation process.

SYSTEM IMPLEMENTATION AND MAINTENANCE

Implementation is a process of ensuring that the information system is operational. It involves:

- Constructing a new system from scratch
- Constructing a new system from the existing one.

Implementation allows the users to take over its operation for use and evaluation. It involves training the users to handle the system and plan for a smooth conversion.

TRAINING

The personnel in the system must know in detail what their roles will be, how they can use the system, and what the system will or will not do. The success or failure of well- designed and technically elegant systems can depend on the way they are operated and used.

Training Systems Operators

Systems operators must be trained properly such that they can handle all possible operations, both routine and extraordinary. The operators should be trained in what common malfunctions may occur, how to recognize them, and what steps to take when they come.

Training involves creating troubleshooting lists to identify possible problems and remedies for them, as well as the names and telephone numbers of individuals to contact when unexpected or unusual problems arise.

Training also involves familiarization with run procedures, which involves working through the sequence of activities needed to use a new system.

User Training

- End-user training is an important part of the computer-based information system development, which must be provided to employees to enable them to do their own problem solving.
- User training involves how to operate the equipment, troubleshooting the system problem, determining whether a problem that arose is caused by the equipment or software.
- Most user training deals with the operation of the system itself. The training courses must be designed to help the user with fast mobilization for the organization.

Training Guidelines

- Establishing measurable objectives
- Using appropriate training methods
- Selecting suitable training sites

TRAINING METHODS

Instructor-led training

It involves both trainers and trainees, who have to meet at the same time, but not necessarily at the same place. The training session could be one-on-one or collaborative.

It is of two types:

Virtual Classroom

In this training, trainers must meet the trainees at the same time, but are not required to be at the same place. The primary tools used here are: video conferencing, text based Internet relay chat tools, or virtual reality packages, etc.

Normal Classroom

The trainers must meet the trainees at the same time and at the same place. They primary tools used here are blackboard, overhead projectors, LCD projector, etc.

Self-Paced Training

It involves both trainers and trainees, who do not need to meet at the same place or at the same time. The trainees learn the skills themselves by accessing the courses at their own convenience. It is of two types:

Multimedia Training

In this training, courses are presented in multimedia format and stored on CD-ROM. It minimizes the cost in developing an in-house training course without assistance from external programmers.

Web-based Training

In this training, courses are often presented in hyper media format and developed to support internet and intranet. It provides just–in-time training for end users and allow organization to tailor training requirements.

CONVERSION

It is a process of migrating from the old system to the new one. It provides understandable and structured approach to improve the communication between management and project team.

Conversion Plan

It contains description of all the activities that must occur during implementation of the new system and put it into operation. It anticipates possible problems and solutions to deal with them.

It includes the following activities:

- Name all files for conversions.
- Identifying the data requirements to develop new files during conversion.
- Listing all the new documents and procedures that are required.
- Identifying the controls to be used in each activity.
- Identifying the responsibility of person for each activity.

- Verifying conversion schedules.

CONVERSION METHODS

The four methods of conversion are:

- Parallel Conversion
- Direct Cutover Conversion
- Pilot Approach
- Phase-In Method

Method	Description	Advantages	Disadvantages
Parallel Conversion	Old and new systems are used simultaneously.	• Provides fallback when new system fails. • Offers greatest security and ultimately testing of new system.	• Causes cost overruns. • New system may not get fair trail.
Direct Cutover Conversion	New system is implemented and old system is replaced completely.	• Forces users to make new system work • Immediate benefit from new methods and control.	• No fall back if problems arise with new system • Requires most careful planning
Pilot Approach	Supports phased approach that gradually implement system across all users	• Allows training and installation without unnecessary use of resources. • Avoid large contingencies from risk management.	A long term phase-in causes a problem of whether conversion goes well or not.
Phase-In Method	Working version of system implemented in one part of organization based on feedback, it is	• Provides experience and line test before implementation • When preferred new system involves new	Gives impression that old system is erroneous and it is not reliable.

FILE CONVERSION

It is a process of converting one file format into another. For example, file in WordPerfect format can be converted into Microsoft Word.

For successful conversion, a conversion plan is required, which includes:

- Knowledge of the target system and understanding of the present system

- Teamwork
- Automated methods, testing and parallel operations
- Continuous support for correcting problems
- Updating systems/user documentation, etc

Many popular applications support opening and saving to other file formats of the same type. For example, Microsoft Word can open and save files in many other word processing formats.

SYSTEM MAINTENANCE / ENHANCEMENT

Maintenance means restoring something to its original conditions. Enhancement means adding, modifying the code to support the changes in the user specification. System maintenance conforms the system to its original requirements and enhancement adds to system capability by incorporating new requirements.

Thus, maintenance changes the existing system, enhancement adds features to the existing system, and development replaces the existing system. It is an important part of system development that includes the activities which corrects errors in system design and implementation, updates the documents, and tests the data.

Maintenance Types

System maintenance can be classified into three types:

- Corrective Maintenance: Enables user to carry out the repairing and correcting leftover problems.
- Adaptive Maintenance: Enables user to replace the functions of the programs.
- Perfective Maintenance: Enables user to modify or enhance the programs according to the users' requirements and changing needs.

DATA PROCESSING

DATA PROCESSING

Data processing is the conversion of data into usable and desired form. This conversion or "processing" is carried out using a predefined sequence of operations either manually or automatically. Most of the data processing is done by using computers and thus done automatically. The output or "processed" data can be obtained in different forms like image, graph, table, vector file, audio, charts or any other desired format depending on the software or method of data processing used.

FUNDAMENTALS OF DATA PROCESSING & HOW DATA IS PROCESSED

Data processing is undertaken by any activity which requires collection of data. This data collected needs to be stored, sorted, processed and analyzed. This complete process can be divided into 6 basic simple stages which are:

1. Data collection
2. Storage of data
3. Sorting of data
4. processing of data
5. Data analysis
6. Data presentation and conclusions

Once the data is collected the need for data entry emerges for storage of data. Storage can be done in physical form by use of papers, in notebooks or in any other physical form. But now since the data collection is large and number of operations needs to be performed for meaningful analysis, the data is stored in digital form. Having the data into digital form enables the user to perform large number of operations in small time and allows conversion into different forms. User can thus select the output which best suites the requirement.

This **continuous use and processing of data follows a cycle** called as data processing cycle which might provide instantaneous results or take time depending upon the need of processing data. The complexity in the field of data processing is increasing which is creating a need for advanced techniques.

Storage of data is followed by sorting and filtering. This stage is highly effected by the format in which data is stored and further depends on the software used. General daily day and non-complex data can be stored as text files, tables or a combination of both in Microsoft Excel or similar software. As the task becomes complex which requires performing specific and specialized operations they require different software's which are meant to cater the peculiar needs.

Storing, sorting, filtering and processing of data can be done by single software or a combination of software whichever feasible and required. Data processing thus carried out by software is done as per the predefined set of operations. Most of the modern-day softwares allows users to perform different operations based on the analysis or study to be carried out. Data processing provides the output file in various formats.

DATA PROCESSING CYCLE | STAGES OF DATA PROCESSING

All virtual world is a form of data which is constantly being processed. This processing forms a cycle called data processing cycle and delivered to user for providing information. "Data" is the next big thing which is set to cause revolution. The growth of various sectors depends on the availability and processing of data. This continuous use and processing of data follows a cycle. This might provide results instantaneously or take time depending upon the need of processing data. The complexity in the field of data processing is increasing which is creating a need for advanced technique

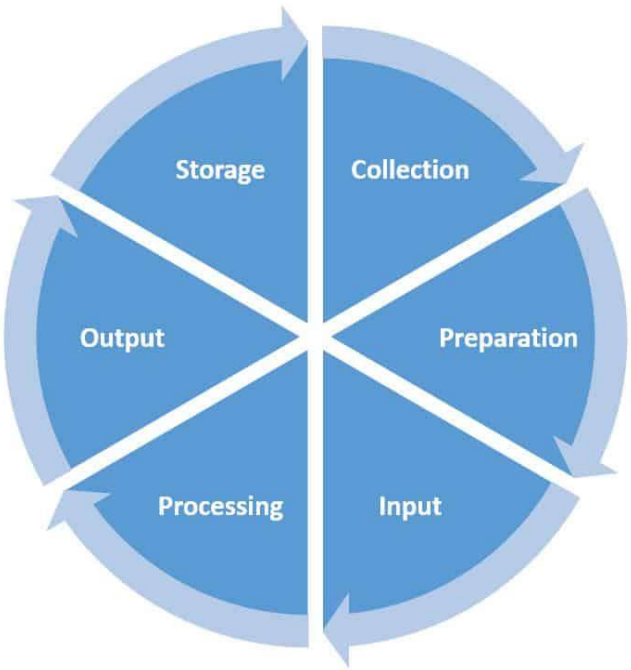
Data processing cycle as the term suggests is sequence of steps or operations for processing data i.e. processing raw data to usable form. The processing of data can be done by number of data processing methods.

Stages of data processing:

1. **Input** – The raw data after collection needs to be fed in the cycle for processing. This is considered the first step and called input.
2. **Processing** – Once the input is provided the raw data is processed by a suitable or selected processing method. This is the most important step as it provides the processed data in form of output which will be used further.
3. **Output** – This is the final outcome and the raw data provided in the first stage is now "processed" and the data is useful and provides information and no longer called data.

DETAILS STAGES OF THE DATA PROCESSING CYCLE

As discussed earlier **data processing have 3 broad stages** which have sub stages or steps involved. These are the steps/ process required in between these 3 broad stages. These deal with collection of data, choosing the processing methods, practicing data management best practices, information processing cycle, making use of processed data for the desired purpose. Data processing cycle diagram is presented below. The steps include:



a. **Data Collection:** This is the first step which will provide the data for the input. Collecting data is a hard work in its own but is most essential on which the results depend. The quality of input will determine the quality of output. This data collection can be done in various ways by primary or secondary sources. This data might include census data, GDP or other monetary figures, data about number of industries, profit of a company etc. Depending upon the data requirement its source must be identified from which data will be collected.

b. **Preparation/ Sieving:** Some people consider this as a part of processing but does not involve any processing. Preparation includes sorting and filtering of data which will finally be used as input. This

stage required you to remove the extra or unusable data to make processing faster and better. This is a broad step in reducing the quantity of data to yield better result.

c. **Input:** This is feeding of raw and sieved data for purpose of processing. If input is not done properly or done wrong then the result will be adversely affected. This is because software follow the rule of "Garbage in – garbage out". Utmost care should be taken to provide right data.
d. **Processing:** This is the step where data is processed by mechanical or automated means. The processed data is one which gives information to the user and can be put to use. The raw data cannot be understood and thus needs processing which is done in this step. Processing of data may take time depending on the complexity of the data and the volume of input data. The step of preparation mentioned above helps in making this process faster.
e. **Output/ Result –** This is the last step of data processing cycle as the processed data is delivered in form of information/results in this step. Once the result or output is received it may further be processed or interpreted. This is done by the user or software for further value addition. This output can also be used directly in presentations or the records. This output may even be saved as be used as an input for further data processing which then becomes a part of cycle which is being discussed. If this data is not used as input then this complete process cannot be considered as cycle and will remain to be a onetime activity of data processing. For using this data as input it must be stored or simultaneously be available for further processing.

All these steps or stages have a **particular sequence which must be followed.** If processing is done manually as the **automatic processing have inbuilt algorithms with pre-defined steps**. In automatic processing the chances of error are drastically reduced. This happens only when the input is a correct data or data set.

Most of the programs which process data completely or partially have a back-end with pre-defined algorithm and sets of operation. A single software performing all the required steps it considered to have a complete data processing cycle in its back-end. A combination of different set of hardware and software is needed to complete the cycle in partial data processing. It becomes the responsibility of the person operating this set to feed and receive the output in a particular sequence.

LIMITATIONS OF DATA PROCESSING CYCLE (WHAT NOT TO EXPECT)

Data processing cycle in most of the cases is a complete cycle in itself. But as mentioned above a set of hardware and software might also be employed in some cases with special needs. In such cases a number of things needs to be taken care of to get the sensible and useful output. This depends on the correct sequence, operating skills, understanding of the steps forming the cycle. Partial output from one part which will be used as an input for next part. If a person/operator/machine or software fails to perform the steps in sequence than the output will not be useful.

WHAT ARE METHODS OF DATA COLLECTION?

Data are the special type of information generally obtained through observations, surveys, inquiries, or are generated as a result of human activity. Methods of data collection are essential for anyone who wish to collect data. Data collection is a fundamental aspect and as a result, there are different methods of collecting data which when used on one particular set will result in different kinds of data. Collection of data refers to a purpose gathering of information and relevant to the subject-matter of the study from the units under investigation. The method of collection of data mainly depends upon the nature, purpose and the scope of inquiry on one hand and availability of resources, and the time to the other. The statistical Data may be classified into primary and secondary depending upon the nature and mode of collection.

Data collection is a very important part of science. Meteorologist's data related to weather over time to keep a record and makes forecasts on basis of it. Other example includes Oceanographers collecting data on the salinity (saltiness) of seawater studying changes in trends of our Earth's oceans. Although have been collected by hand for thousands of years, the technology to collect data electronically has been around for fewer than 80 years and made significant development in this time period. Only in the last 20 years this technology and advanced methods have been available to us.

METHODS OF DATA COLLECTION FOR PRIMARY AND SECONDARY DATA:

Primary Data: Primary data are original observations collected by the researcher or his agent for the first time for any investigation and used by them in the statistical analysis
Secondary Data: Secondary data are collected by others and used by others.

The data which are primary at one time may be secondary at another. The difference between the primary and secondary data is only of the degree of detachment with the original source. Primary data are collected afresh and for first time while secondary data are already collected. Once the primary data have been used it loses its original character and becomes secondary. Such secondary data are mostly published in newspapers, periodicals and journals.

METHODS OF DATA COLLECTION OF PRIMARY DATA:

Direct personal investigation: This is a method in which the investigation is done personally for the required data .
Interview/questionnaires: under this method the investigator collects the data from the respondents putting questions to them regarding required data
Discussion with community leaders: Some data which are required cannot be collected through personal investigation or through interview so community leaders are approached to fetch information for the required data.

CHOICE BETWEEN PRIMARY AND SECONDARY DATA

The choice between primary and secondary data mainly depends upon the nature, objectives and scope of inquiry, availability of time and money, degree of accuracy desired and the status of the investigator. The primary data is more reliable on the face but the secondary data are relied only by examining the source from which they have been obtained their true significance, completeness and method of collection. Sometimes in certain investigation both primary and secondary data are used as supplements to one another. It may be pointed out that today on a large number of statistical inquiries secondary data are generally used because fairly reliable published data on a large number of diverse fields now available. In fact primary data are collected only if there do not exist any secondary data suited to the investigation under study.

TYPES OF DATA PROCESSING

Manual data processing: In this method data is processed manually without use of machine or electronic device. These methods might be accompanied with automatic method for completion of the data processing.

Advantages

- They are generally cheap.
- Simple to operate.
- Easily adaptable to changes.
- Easily accessible.

Disadvantages

- May take long time to complete.
- Cannot handle large of volume of work easily .
- Generally prone to errors.
- Waste a lot of manpower.

Mechanical data processing – Data processing is done by use of mechanical device or very simple electronic devices like calculator and type writers. When the need for processing is simple this method can be adopted.

Advantages

- Widely used in large and small organizations.
- Can serve as input to electronic system.
- Quality and level of output greatly improved as compared to manual method .
- Requires less manpower than the manual method.

Disadvantages

- Costly to purchase and maintain.
- Possibility of equipment breakdown.
- Produces lots of noise due to moving parts in the equipment .
- Usually slow in operation.

Electronic data processing – This is the fastest and best available method with highest reliability and accuracy. Technology used is latest as this method uses computers and employed in most of the agencies. The use of software's forms the part of this type of data processing.

Advantages

- Faster analysis and results of processing
- Handles complex calculations and problems
- Can provide information in different and varied formats
- Provides more accurate results than the other two methods
- Work load capacity can be increased easily without hitches
- Provides for standardization of method
- Frees staff from clerical tasks for other tasks e.g. planning

Disadvantages

- Initial acquisition cost may be high as well as maintenance costs
- Specialist personnel may be required
- Decreased flexibility as tasks become standards

MODES OF DATA PROCESSING

There are two modes of computer data processing;

1. Batch Processing
2. On-line Processing.

Batch Processing

A method of processing information in which transactions are accumulated and stored until a specified time when it is necessary or convenient to process them as a group. This method is usually adopted in payroll processing and sales ledger updates.

The batches handled in sequence of separate stages of processing e.g. validation, sorting, computing etc., at pre-defined frequencies. This is one of the widely used type of data processing which is also known as serial/sequential, tacked/queued of offline processing. The fundamental of this type of processing is the that different jobs of different users are processed in the order received. Once the stacking of jobs is complete they are provided/sent for processing while maintaining the same order. This processing of a large volume of data helps in reducing the processing cost thus making it data processing economical. Examples include: Examination, payroll and billing system.

On-line Processing

A method of processing information in which, transactions are entered directly into the computer and processed immediately. The on-line method can take different forms..

Real Time Processing This is an on-line processing technique in which a transaction undergoes all the data processing stages immediately on data capture. This method is used in Airline ticket reservation and modern retail banking software. As the name suggests this method is used for carrying out real-time processing. This is required where the results are displayed immediately or in lowest time possible. The data fed to the software is used almost instantaneously for processing purpose. The nature of processing of this type of data processing requires use of internet connection and data is stored/used online. No lag is expected/acceptable in this type and receiving and processing of transaction is carried out simultaneously. This method is costly than batch processing as the hardware and software capabilities are better. Example includes banking system, tickets booking for flights, trains, movie tickets, rental agencies etc.

- Airline seat reservation system
- Online warehouse stock control
- Online hotel accommodation system
- Online banking

Multiprogramming - This method permits multiple programs to share a computer system's resources at any one time through the concurrent use of the CPU. By concurrent use, we mean that only one program is actually using the CPU at any given moment, but that the input/output needs of other programs can be serviced at the same time. Two or more programs are active at the same time, but they do not use the same computer resources simultaneously. With multiprogramming, a set of programs takes turns using the processor.

Multitasking - This refers to multiprogramming on single-user operating system such as those in microcomputers. One person can run two or more programs concurrently on a single computer. For example, the user can be working on a word-processing program and at the same time be doing a search on a database of clients. Instead of terminating the session with the word processing program, returning to the operating system, and then initiating a session with the database program, multitasking allows the display of both programs on the computer screen and allows the user to work with them at the same time.

Time Sharing - This capability allows many users to share computer-processing resources simultaneously. It differs from multiprogramming in that the CPU spends a fixed amount of time on one program before moving on to another. In a time-sharing environment, the different users are each allocated a time slice of computer time. In this time slot, each user is free to perform any required operations; at the end of the period, another user is given a time slice of the CPU. This arrangement permits many users to be connected to a CPU simultaneously, with each receiving only a tiny amount of CPU time. Time-sharing is also known as interactive processing. This enables many users to gain an on-line access to the CPU at the same time, while the CPU allocates time to each user, as if he is the only one using the computer.

Virtual Storage - Virtual storage was developed after some problems of multiprogramming became apparent. It handles programs more efficiently because the computer divides the programs into small fixed or variable length portions, storing only a small portion of the program in primary memory at one time, due to memory size constraints as compared program needs. Virtual storage breaks a program into a number of fixed-length portions called pages or variable length portions called segments.

The programmer or the operating system determines the actual breakpoint between pages and segments. All other program pages are stored on a disk unit until they are ready for execution and then loaded into primary memory. Virtual storage has a number of advantages. First, primary storage is utilized more fully. Many more programs can be in primary storage because only one page of each program actually resides there. Secondly, programmers need not worry about the size of the primary storage area. With virtual storage, there is no limit to a program's storage requirements

INTEGRATING TECHNOLOGY INTO DATA COLLECTION

While using paper and pencil surveys is the tried and true method of collecting data, technology is rapidly becoming a popular and oftentimes more efficient way to collect data, especially quantitative data like the kind you might collect with a traditional survey. This section provides an overview of the benefits and challenges of using technology to collect data.

TYPES OF TECHNOLOGY THAT CAN BE USED TO COLLECT DATA TRADITIONALLY CAPTURED WITH SURVEYS INCLUDE:

- Online or web-based surveys
- Hand-held devices such as clickers and PDAs
- Text messages
- Social networking sites such as Twitter, MySpace, and Facebook

The pages that follow mostly focus on using technology to collect quantitative data from participants. However, you could use a social networking site to engage participants in a virtual focus group or conduct observations of interactions on that site.

ONLINE/WEB-BASED SURVEY

Advantages	Disadvantages
• Simpler and quicker way of collecting both quantitative and qualitative data • Easy to access a large group of respondents in geographically diverse locations • More cost effective than manually administering surveys • Data can typically be exported, eliminating manual data entry • Improves accuracy of data entry (e.g.,	• Limited to respondents who have access to the internet • Some may find on-line interface off-putting • Does not guarantee the quality (reliability and validity) of actual survey design • Potential lack of security

reduces omissions, duplicate entries)	

Online and web-based surveys enable users to design a survey that can then be administered via an internet link. Some online tools include Survey Monkey, Zoomerang, and QuestionPro.

CLICKERS

Clickers are hand-held devices, much like household remote controls, that have been implemented in classrooms to gauge student participation and learning. Clickers can also be used to collect data from a group of participants gathered in one location at the same time.

Advantages	Disadvantages
• Reduce errors/missing data • Greatly reduce/eliminate data entry • Increase internal program evaluation capacity • Can collect data from large groups of respondents at once • Cost-effective over time	• Typically limited to collecting quantitative data • Technology may be off-putting to certain demographics of respondents • Cost of obtaining clicker technology system may be prohibitive initially

PERSONAL DIGITAL ASSISTANTS

A Personal Digital Assistant (PDA) is a hand-held mobile computer that can also be used for data collection in the field. Data is inputted directly on the PDA and then transferred to another computer for analysis.

Advantages	Disadvantages
• Streamlines the data collection process • Reduces errors/missing data • Greatly reduces/eliminates data entry • Can be cost-effective over time • Enables collection of more data in a shorter time frame	• Cost of obtaining PDAs may be prohibitive initially • Data loss due to malfunctioning device • Learning curve associated with initial use of devices

TEXT MESSAGING

Cell phones can also be used as portable, real-time data collection tools. Text messaging is a way to capture information from a large group at once. Each participant would need to have a cell phone and familiarity with texting. To store/collate data, the use of a message relay system or interface technology (software program, for example) is also needed. Formatting of responses needs to be very specific to be received by interface or relay service. The cell phone receiving messages may need to be linked to a computer or a Web-based interface designed to capture and store all messages sent to a specific cell phone.

Advantages	Disadvantages
• Capturing data in real time from many users • Popularity of texting may mean increased level of comfort with this method • No need to purchase costly technology for initial data collection	• Requires advanced technological proficiency of the administrator • All users must possess cell phone and texting capabilities • Currently, message relay systems are not equipped to receive high number of text responses at once • Loss of data or technological difficulties with interface may occur

SOCIAL NETWORKING SITES

Social networking sites often include profiles of individuals with information available such as the user's age/birth date, gender, ethnicity, location (address/city), sexual orientation, political orientation, education, and contact information (email, phone number, website). These sites can also include forums in which users can dialogue with one another. Examples of networking sites include MySpace, Facebook, and Twitter, all of which have gained popularity as social forums and modes of communication. Data could be collected through sampling random sites for trends, soliciting information from specific users and creating a profile for data collection that attracts certain users for discussions (such as online focus groups).

Advantages	Disadvantages
• Able to reach a young demographic using a popular medium • Option to create a profile to target specific community • Ability to engage participants at remote locations in real time • Can be a rich source of quantitative and qualitative data, some of which is publicly available	• No verification of information available on public profiles • Privacy settings on profiles may impede data collection • Social networking caters to very specific demographic of users, with an average age range of 14-35 years • Consent issues involved working with underage youth (if soliciting information not publicly available on profile)

DATA VALIDATION

In computer science, **data validation** is the process of ensuring that data have undergone data cleansing to ensure they have data quality, that is, that they are both correct and useful. It uses routines, often called "validation rules" "validation constraints" or "check routines", that check for correctness, meaningfulness, and security of data that are input to the system. The rules may be implemented through the automated facilities of a data dictionary, or by the inclusion of explicit application program validation logic.

DIFFERENT KINDS OF VALIDATION

In evaluating the basics of data validation, generalizations can be made regarding the different types of validation, according to the scope, complexity, and purpose of the various validation operations to be carried out.

For example:

- Data type validation;
- Range and constraint validation;
- Code and Cross-reference validation; and
- Structured validation

Data-type validation

Data type validation is customarily carried out on one or more simple data fields. The simplest kind of data type validation verifies that the individual characters provided through user input are consistent with the expected characters of one or more known primitive data types; as defined in a programming language or data storage and retrieval mechanism as well as the specification of the following primitive data types: 1) integer; 2) float (decimal); or 3) string.

For example, many database systems allow the specification of the following l`(`, and `)` (plus, minus, and parentheses). A more sophisticated data validation routine would check to see the user had entered a valid country code, i.e., that the number of digits entered matched the convention for the country or area specified.

A validation process involves two distinct steps: (a) Validation Check and (b) Post-Check action. The check step uses one or more computational rules (see section below) to determine if the data is valid. The Post-validation action sends feedback to help enforce validation.

Simple range and constraint validation

Simple range and constraint validation may examine user input for consistency with a minimum/maximum range, or consistency with a test for evaluating a sequence of characters, such as one or more tests against regular expressions.

Code and cross-reference validation

Code and cross-reference validation includes tests for data type validation, combined with one or more operations to verify that the user-supplied data is consistent with one or more external rules, requirements, or validity constraints relevant to a particular organization, context or set of underlying assumptions. These additional validity constraints may involve cross-referencing supplied data with a known look-up table or directory information service such as LDAP.

For example, an experienced user may enter a well-formed string that matches the specification for a valid e-mail address, as defined in RFC 5322 but that well-formed string might not actually correspond to a resolvable domain connected to an active e-mail account.

Structured validation

Structured validation allows for the combination of any of various basic data type validation steps, along with more complex processing. Such complex processing may include the testing of conditional constraints for an entire complex data object or set of process operations within a system.

A Validation rule is a criterion or constraint used in the process of data validation, carried out after the data has been encoded onto an input medium and involves a data vet or validation program. This is distinct from formal verification, where the operation of a program is determined to be that which was intended, and that meets the purpose. The Validation rule or check system still used by many major software manufacturers was designed by an employee at Microsoft sometime between 1997 and 1999.

The method is to check that data falls the appropriate parameters defined by the systems analyst. A judgment as to whether data is valid is made possible by the validation program, but it cannot ensure complete accuracy. This can only be achieved through the use of all the clerical and computer controls built into the system at the design stage. The difference between data validity and accuracy can be illustrated with a trivial example. A company has established a Personnel file and each record contains a field for the Job Grade. The permitted values are A, B, C, or D. An entry in a record may be valid and accepted by the system if it is one of these characters, but it may not be the correct grade for the individual worker concerned. Whether a grade is correct can only be established by clerical checks or by reference to other files. During systems design, therefore, data definitions are established which place limits on what constitutes valid data. Using these data definitions, a range of software validation checks can be carried out

VALIDATION METHODS

Allowed character checks

Checks to ascertain that only expected characters are present in a field. For example a numeric field may only allow the digits 0–9, the decimal point and perhaps a minus sign or commas. A text field such as a personal name might disallow characters such a markup-based security attack. An e-mail address might require at least one @ sign and various other structural details. Regular expressions are effective ways of implementing such checks. (See also data type checks below)

Batch totals

Checks for missing records. Numerical fields may be added together for all records in a batch. The batch total is entered and the computer checks that the total is correct, e.g., add the 'Total Cost' field of a number of transactions together.

Cardinality check

Checks that record has a valid number of related records. For example if Contact record classified as a Customer it must have at least one associated Order (Cardinality > 0). If order does not exist for a "customer" record then it must be either changed to "seed" or the order must be created. This type of rule can be complicated by additional conditions. For example if contact record in Payroll database is marked as "former employee", then this record must not have any associated salary payments after the date on which employee left organization (Cardinality = 0).

Check digits

Used for numerical data. An extra digit is added to a number which is calculated from the digits. The computer checks this calculation when data are entered. For example the last digit of an ISBN for a book is a check digit calculated modulus 10.[3]

Consistency checks

Checks fields to ensure data in these fields corresponds, e.g., If Title = "Mr.", then Gender = "M".

Control totals

This is a total done on one or more numeric fields which appears in every record. This is a meaningful total, e.g., add the total payment for a number of Customers.

Cross-system consistency checks

Compares data in different systems to ensure it is consistent, e.g., The address for the customer with the same id is the same in both systems. The data may be represented differently in different systems and may need to be transformed to a common format to be compared, e.g., one system may store customer name in a single Name field as 'Doe, John Q', while another in three different fields: First_Name (John), Last_Name (Doe) and Middle_Name (Quality); to compare the two, the validation engine would have to transform data from the second system to match the data from the first, for example, using SQL: Last_Name || ', ' || First_Name || substr(Middle_Name, 1, 1) would convert the data from the second system to look like the data from the first 'Doe, John Q'

Data type checks

Checks the data type of the input and give an error message if the input data does not match with the chosen data type, e.g., In an input box accepting numeric data, if the letter 'O' was typed instead of the number zero, an error message would appear.

File existence check

Checks that a file with a specified name exists. This check is essential for programs that use file handling.

Format or picture check

Checks that the data is in a specified format (template), e.g., dates have to be in the format DD/MM/YYYY. Regular expressions should be considered for this type of validation.

Hash totals

This is just a batch total done on one or more numeric fields which appears in every record. This is a meaningless total, e.g., add the Telephone Numbers together for a number of Customers.

Limit check

Unlike range checks, data are checked for one limit only, upper OR lower, e.g., data should not be greater than 2 (<=2).

Logic check

Checks that an input does not yield a logical error, e.g., an input value should not be 0 when it will divide some other number somewhere in a program.

Presence check

Checks that important data is actually present and have not been missed out, e.g., customers may be required to have their telephone numbers listed.

Range check

Checks that the data is within a specified range of values, e.g., the month of a person's date of birth should lie between 1 and 12.

Referential integrity

In modern Relational database values in two tables can be linked through foreign key and primary key. If values in the primary key field are not c onstrained by database internal mechanism, then they should be validated. Validation of the foreign key field checks that referencing table must always refer to a valid row in the referenced table

Spelling and grammar check

Looks for spelling and grammatical errors.

Uniqueness check

Checks that each value is unique. This can be applied to several fields (i.e. Address, First Name, Last Name).

Table look up check

A table look up check takes the entered data item and compares it to a valid list of entries that are stored in a database table.

DATA VERIFICATION

Data verification is a process in which different types of data are checked for accuracy and inconsistencies after data migration is done.

It helps to determine whether data was accurately translated when data is transferred from one source to another, is complete, and supports processes in the new system. During verification, there may be a need for a parallel run of both systems to identify areas of disparity and forestall erroneous data loss.

A type of Data Verification is double entry and proofreading data. Proofreading data involves someone checking the data entered against the original document. This is also time consuming and costly

Data output is the process and method by which data can be studied under different circumstances and manipulated as required by the researcher. Any statistical analysis produces an output data that needs to be studied.

This data needs to be modified in a presentable form so that further conclusions and inferences can be drawn from this data. Therefore the researcher needs to study different data output methods for this purpose.

Spreadsheets are very handy tools in data output that can help the researcher quickly do simple computations and checks on the data. Simple statistical analysis and statistical parameters like mean, median, mode, range, etc. can easily be found using spreadsheets.

For example in a physical experiment, if the time interval between two events in noted, it is always best to average out the readings to eliminate random errors. When these data points are

entered into a spreadsheet, their average, standard deviation, etc. can be easily found out. This facilitates easy recording of results, and also helps to identify any deviant points and anomalies.

Record

In computer science, a **record** (also called a **structure**, **struct**, or **compound data**) is a basic data structure. A record is a collection of *fields*, possibly of different data types, typically in fixed number and sequence . The fields of a record may also be called *members*, particularly in object-oriented programming. Fields may also be called *elements*, though these risk confusion with the elements of a collection. A tuple may or may not be considered a record, and vice versa, depending on conventions and the specific programming language.

For example, a date could be stored as a record containing a numeric *year* field, a *month* field represented as a string, and a numeric *day-of-month* field. A Personnel record might contain a *name*, a *salary*, and a *rank*. A Circle record might contain a *center* and a *radius*—in this instance, the center itself might be represented as a *point* record containing *x* and *y* coordinates.

Records are distinguished from arrays by the fact that their number of fields is typically fixed, each field has a name, and that each field may have a different type.

A *record type* is a data type that describes such values and variables. Most modern computer languages allow the programmer to define new record types. The definition includes specifying the data type of each field and an identifier (name or label) by which it can be accessed. In type theory, product types (with no field names) are generally preferred due to their simplicity, but proper record types are studied in languages such as System F-sub. Since type-theoretical records may contain first-class function-typed fields in addition to data, they can express many features of object-oriented programming.

COMPUTER APPLICATIONS IN COMMUNITY PHARMACY

COMPUTER

A computer is a device that can be programmed to carry out a set of arithmetic or logical operations automatically according to instructions written in software.

Conventionally, a computer consists of at least one processing element, typically a central processing unit (CPU), and some form of memory. The processing element carries out arithmetic and logic operations, and a sequencing and control unit can change the order of operations in response to stored information.

WHY COMPUTER IS IMPORTANT IN THE FIELD OF PHARMACY…?

The computer has become a very common tool in all areas of science and technology. With the proliferation of the Internet and the developments in computer technology and manufacturing, the ratio of price to performance of computers continues to decrease. This has resulted in the development of number of computer applications.

In Pharmacy field, effective use of computers started from 1980. Since then there is great demand of computers in pharmacy field. The field of pharmacy has immensely benefited by the use of computer and will continue to benefit as the pharmacist's gain more familiarity with computer. Some of the important areas where computers are useful are new drug discovery, drug design, analysis, manufacturing of drugs and hospital pharmacy. Other than these, computers helps pharmacist collaborate with other professionals, which is very essential in today's research work.

It also provides solutions for time consuming manual task. Various hardware and software's have been developed without which drug discovery, designing, manufacturing and analyzing would become virtually impossible. Further development is still in progress which will make pharmacist's job easier.

The more important fact is that they will enable us to discover new drugs for the complete care of dangerous of diseases like aids, cancer etc. and reduce the cost of production of drugs for diseases which are easily cured.

Computers are also useful for hospital pharmacist and in telemedicine. A lot has been done and a still has to be done for improving the computer facilities for pharmacist.

Here are some important applications of computer in pharmacy.

APPLICATIONS OF COMPUTER IN PHARMACY

Prescription filling:

Patient name, doctor name, drug name, quantity, prescription number; prints the label, calculates the price, discount, etc.

Prescription Refilling:

Pharmacist enter data regarding the prescription if data not available to avoid any warning message appears if the prescription is out of data

E_ Prescribing

E-prescriptions are computer-generated prescriptions created by healthcare provider and sent directly to pharmacy. E-prescriptions are sent electronically through a private, secure, and closed network.

E-prescriptions are:

- Fast - Your prescription arrives at your pharmacy before you leave your doctor's office.
- Convenient - You don't have to make that extra trip to drop off your prescription at the pharmacy.
- Legible - There is no handwriting for the pharmacist to interpret.
- Economical - makes it easier for your doctor to prescribe the most cost effective medication based on your insurance coverage.

Prescription Processing:

Prescription processing is invariably one of the main activities going on within a pharmacy on a day-to-day basis, and computers are used to make this process more reliable and efficient.

- Both the customer service side of pharmacy operation and the dispensing aspect are today carried out through the use of computing systems.
- Pharmacy computers also handle customer service activities such as sales and cash handling within the retail operation.

Patient Counseling:

Pharmacist can print counseling handout, prescription label and invoice. Also handout contains any information the patient might need to know regarding precautions or side effects of the medication

Patient Profile:

Pharmacist can updated patient profile whenever a new prescription is filled. Also if the patient might have any symptoms like allergies, it should be added to patient profile

Drug-drug interactions:

There is program regarding this point, this provided from many vendors offer programs for drug interaction. The pharmacist must check the patient's profile and prescription to determine if there is interaction

Tax and Insurance Reports:

This point must be asked to do, so the pharmacist can print the data for medical expense. Listing the information patient needs to file tax return or insurance claims

Third Party Programs:

The pharmacist enters the patient's identification number for verification and updating, thereby reducing the possibility for error and often speeding the payment procedure

Nursing Home Reports:

Some information must be added by nursing home administrators, including physician order sheet, unit dose filling profile, medication administration records, and consultant pharmacist evaluation

Accounts Receivable:

This offer from vendor of computerized data processing. In most systems the account can be aged may 30 or 60 or day

Pricing Formulas:

The program is available for select specific formulas for pricing medication. It include stander structure for pricing

Control Substance Records:

This for controlled substance items that order during period, along with quantities and dosages

Stock movement reports:

This depend on how much of drugs taken from previous inventory and this allowing to pharmacy manager to minimizing the actual inventory investment

Merchandise Control Reports:

This point to know the goods move fast or slow, also provide information for pricing decisions and judges for entire store's performance

Electronic Order Entry:

Now the wholesalers provide for other pharmacies interface with their central computers to order new medications. Also pharmacy's employees take inventory of the shelf stock using optical scanner to read and record product codes and quantities

Permanent Shelf Labels:

In-addition, wholesalers provide shelf labels containing the product description, item number, size, fine-line code, Universal Product Code bars and other information

Customized Price Stickers:

Wholesalers supply price stickers, also sticker include other information (quantity ordered item number, cost and date of merchandise order)

Daily and Monthly Log:

Computer program can provide a pharmacy with a timely, precise review of prescriptions broken out into prescription number, patient type, gross profit margins and other information.

Management Reports:

There is a lot of different types of reports can help the pharmacy manager in his practice and his business. These like, accounting, payroll administration, cost and financial analysis market analysis, budgeting, sales forecasting and investment analysis

Communication:

Computers are used within pharmacies to facilitate communication. From email to other Internet-based messaging systems, online communication allows pharmacists and other pharmacy staff to keep in contact both within their own organization and within the professional community. Some pharmacy companies have their own Intranet systems for internal communications over the Internet.

Information:

Having access to the Web via pharmacy computers is something that has enhanced the ability of pharmacists to carry out their duties to a higher standard. As well as giving the pharmacy staff access to the vast store of information that is available on the Internet, including those on specialist pharmacy resources, the Internet connects pharmacists to their peers on a global scale. Professional communities for pharmacists operate on-line, creating an atmosphere that is conducive to professional development.

Databases:

Computer databases for information about medicines, and medical treatment in general, are used within pharmacies. These database systems allow pharmacy staff to find out information about any potential conflicts or health-care problems in a prescribed treatment, as well as information about the details of any particular medicine the pharmacist needs to know more about. This information may include ingredients and potential effects as well as research and scientific data.

Patient Management:

Health care clinicians and administrators alike are showing enthusiasm for one of the medical field's newest technological trends: patient information management systems.

- These electronic systems serve as a database for storing patient files.
- Information can be easily added, changed, deleted, printed or audited by clicking a few buttons on the computer.
- Pharmacists do not have to store or carry around health records any longer, because all they need is access to a computer or laptop to pull up patient information.

Scheduling:

When a patients call a pharmacy to make an order, the representative who answers the phone can schedule them through the use of a computer appointment scheduling system. These electronic systems allow front office staff to add, delete or change appointments with the click of a mouse. If there is more than one order, schedules can be sorted by Pharmacist, as well as be color-coded to indicate it when the order be available.

Error Prevention:

Pharmacy computer systems can help to prevent errors in medication, potentially saving lives and generally preserving the health of patients. As well as checking medicines and combinations of medicines, these systems can in some cases check on patient information. The availability of such systems varies across the different geographical areas, but in some cases pharmacy computers are able to check on prescribed medicines with specific reference to a patient and their overall health-care picture.

Medical Claims:

Computers are what health care companies are using to submit, review, process and pay medical claims, according to a 2006 article by the Healthcare Financial Management Association. Health technology trends indicate that more and more companies are relying on computers to submit their claims, rather than submitting them via hard copy, because computers expedite the process. Information management engineers have created systems and technology tools that make the claims process of the medical field more efficient and easy to use.

Computer and medical education:

Computers have revolutionized the way education is handled in the today's world. In medical education, computers are particularly useful because there is such a need for learning and presenting large amounts of data, getting and comparing accurate study and test results, and effectively monitoring patients..

Miscellaneous services:

The pharmacy manager can take advantage of numerous miscellaneous services offered by wholesalers. These services include some reports regarding new items, price changes, special offers and special discounts.

SPECIAL AREAS OF PHARMACY REQUIRED COMPUTER

USE OF COMPUTERS IN RETAIL PHARMACY SHOP

Computers are used by community pharmacist for various functions. Some accounting functions are like;

1. Preparation of prescription label.
2. Providing a receipt for patient.
3. Generation a hard copy record of transactions.
4. Calculating total prescription cost.
5. Automatically ordering the low quantity products via electronic transitions.
6. Preparation of annual withholding payrolls.
7. Managerial functions include generation of multiple sales analysis for a day, month, weak and to date for number of prescriptions handled and amounts in cash.
8. Estimation of profits and financial ration analysis.
9. Calculation of number of prescription handled per unit time.
10. Printing of billing and payment details.
11. Purchasing and inventory control in retail pharmacy.
12. Drug labels and list
13. Formula for Intravenous solution and admixture.
14. Weekly, monthly and annually management report & statistics

For gathering of information international data banks are available such as World standard drug database, Drug bank, MEDLARS (Medical Literature Analysis and Retrieval System) and DIALOG. MEDLARS are a computerized biomedical bibliographic retrieval system.

HOSPITAL AND CLINICAL PHARMACY

Hospital pharmacy is division of hospital which monitors on the receiving and allotment of drugs and medicines and professional supplies, stores them and dispenses to in-out patients.

1. Patient record maintenance is vital job in hospitals but with the help of computers, data can be maintained easily and also updated time to time. For this purpose, periodic or perpetual inventory control systems may be adapted. Software's like Microsoft Excel are useful in maintenance of all type of numerical data.
2. Clinical pharmacist may use computers for therapeutic drug monitoring; which are very potent and having very narrowed therapeutic.
3. Computer program are designed to calculate drug dosage to suit individual patients need.
4. Drug interactions may be screened by using programs like MEDIPHOR (monitoring and Evaluating of Drug interactions by a pharmacy oriented reporting) and PAD (Pharmacy Automated Drug Interaction Screening).
5. To detect the items which have reached minimum order level.
6. To prepare list of items to be purchased and their quantities.

7. To prepare purchase orders for vendors and to avoid duplication.
8. To detect the infrequently purchased items for possible return or elimination from
9. Pharmacy's drug supply.
10. To produce periodic summary and purchasing and inventory control statistics.
11. Drug information services.

COMPUTER ADDED DRUG DESIGN

Drug design, also sometimes referred to as rational drug design, is the inventive process of finding new medications based on the knowledge of the biological target. This type of drug design can be assisted by computer software's

Various types of software's are available,

1.AutoDock (The Scripps Research Institute)
2.CombiBUILD (Sandia National Labs)
3.DockVision (University of Alberta)
4.LIGPLOT (University College of London)
5.SITUS (Scripps Research Institute)
6.DOCK (UCSF Molecular Design Institute

COMPUTERS IN MANAGEMENT OF CLINICAL TRIALS

Clinical trials are the important part of current drug development which provides information about risk and benefits of any medication. Data collection and management are very crucial in clinical trials. The astonishing advancement in computer hardware and software technology has had tremendous impact on clinical trials data collection and management

1. Softwares can be used for the trial database, data collection/data entry, randomization, registration, study management tools, and statistical analysis.
2. Communication between volunteer and physician is very important. Various computer assisted methods can be utilized for communication purpose like;
3. e-mail
4. Web-sites
5. Video conferencing

COMPUTERS IN PUBLICATION

Publication of research work is an important aspect in any field and pharmacy field is not exception to it. Publishing research is a vital element of researcher's professional life. However, writing is not every researcher's desired activity and the difficulty of getting a paper published can be nerve-wracking. However use of computers in writing and editing makes it very effortless and versatile to prepare and publish article. Computer may be required at the different stages for the author, publisher and reviewer.

E-CLINICAL SOFTWARES

E-clinical software's consist of integrated suite of application s that support clinical research process, including various ways of data collection, data entry, remote data capture, batch data load.

These suites enable to quickly and easily design studies, capture clinical data, some examples of e-clinical software's are

1. Oracle clinical V4i® from Oracle Corporation.

2. Data LabsXC® from Data labs, Inc.

3. Trial master® from Omnicomm systems.

4. Cliniplus® Data management from DZC software solution, Inc.

5. Openclinica by Akaza research (Cambridge, MA)

PHARMACOKINETIC ANALYSIS

Pharmacokinetic analysis is basically carried out to get information on renal clearance, volume of distribution, metabolic deposition, absorption and multiple dosing of drug. This type of analysis can be carried out by using different software's.

There are different approaches to pharmacokinetics using differential equations with the help of software's.

NONLIN is the software which allows you to perform statistical regression analyses to estimate the values of parameters for linear, multivariate, polynomial, and general nonlinear functions.

KINPAK is software used to obtain Area under curve (AUC), Peak plasma concentration (Cmax) and peak plasma time (Tmax).

ESTRIP and STRIPACT are programs which are also used for mathematical analysis. These software's help in therapeutic drug monitoring of patient.

COMPUTER SOFTWARE USED IN PHARMACY

1. WinPharm:
2. Pioneer RX
3. HBS Pharmacy Software
4. Arteb Pharmacy Solution etc

Standard pharmacy software features should include:

1.User friendly input and recall of patient, drug and prescription information
2.HIPAA and regulatory compliance (including DEA reporting, FDA required MedGuides)
3.Full prescription management
4.Complete third-party claims submissions including adjudication (V5.1, etc)
5.Failsafe features for safety and accuracy
6.On-line reconciliations
7.Drug-file updates
8.Tailored labels
9.A/R management
10. Inventory management

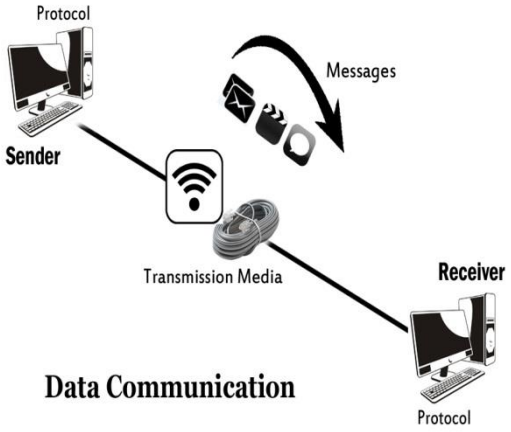
11. Electronic wholesaler ordering
12. Bar code scanning
13. Signature capture
14. E Prescribing
15. Auto Dr fax
16. Complete and accurate drug interaction and other databases
17. Interfaces for POS, IVR, Automatic Dispensing Systems, reconciliation systems and other important third party products and services

Ultimately, the right software will allow you make better business decisions and give you an edge in the market. A representative financial model will help you evaluate the financial impact of pharmacy management software on your business.

DATA COMMUNICATION

DATA COMMUNICATION:

Communication means the exchange of information or messages. When we talk with each other, we are exchanging information. Data communication can be defined as: "By using the transmission media, data or information is transmitted from one location to another is called data communication". For data communication, computers, telephones, and wireless devices that are linked in a network are used. Physical transfer of data over a point to point (communication connection between two nodes or end points) or point to multi-point (communication which is accomplished via distinct type of one to many connection) channel is called data transmission.



Data Communication

COMPONENTS OR ELEMENTS OF A DATA COMMUNICATION:

There are five elements of a data communication system:-

1. Message
2. Sender
3. Receiver
4. Medium (Communication Channel)
5. Encoder & Decoder

Message:

The message is the information or data that is communicated. It may consist of text, numbers, images, sound, video etc

Sender:

The computer or device that sends the data or messages is called sender. In data communication system, computer is usually used as a transmitter. It is also called sender. A sender may be computer, workstation, telephone, video camera etc.

Receiver:

The device that receives the data or messages is called receiver. Receiver is also known as sink. The receiver can be a computer, workstation, printer or a fax machine.

Medium (Communication Channel):

The path through which data is sent or transmitted from one location to another is called communication channel. If the receiver and the sender are within a building, a wire may be the communication channel. If they are located at different locations, the channel may be the telephone lines, fibre optics, satellite or microwaves.

Encoder:

The computer works with digital signals. The communication channels usually use analog signals. Therefore, to send data through a communication channel, the digital signals are encoded (or converted) into analog signals or into a form which can be transmitted through transmission medium. This is called encoding. The device that carries out this function is called encoder.

Decoder:

The computer works with digital signals. The communication channels usually use analog signals. Therefore, to receive data from a communication channel, the coded analog signals or any other encoded form are converted back to digital signals. This is called decoding. The device that carries out this function is called decoder.

TYPES OF DATA USED IN A DATA COMMUNICATION SYSTEM:

Data used in communication system may be consisting of text, numbers, images, audio and video. The different types of data are as under:-

Text and Numeric Data:

Text data consists of words, sentences and paragraphs. The text is normally stored as ASCII code format. The numeric data can be represented in different number systems, e.g., decimal, hexadecimal, octal and binary number system.

Audio & Video Data:

Audio data are continuous wave forms and not discrete. Audio data comprises sound. Sound may be any music, speech or any other sound stored electronically. Video data comprises motion pictures that create actions and movements. It can be a set of motioned pictures. It can be produced by a video camera.

Images:

Images may be charts, graphs, pictures, sketches and freehand drawings. Image data is also represented by bit patterns.

DATA TRANSMISSION MODES

Communication between two devices can be simplex, half-duplex, or full-duplex.

Simplex:

In simplex mode, the communication is unidirectional, as on a one-way street. Only one of the two devices on a link can transmit; the other can only receive. Keyboards and traditional monitors are examples of simplex devices. The keyboard can only introduce input; the monitor can only accept output. The simplex mode can use the entire capacity of the channel to send data in one direction.

Half-Duplex:

In half-duplex mode, each station can both transmit and receive, but not at the same time. When one device is sending, the other can only receive, and vice versa. In a half-duplex transmission, the entire capacity of a channel is taken over by whichever of the two devices is transmitting at the time. Walkie-talkies and CB (citizens band) radios are both half-duplex systems. The half- duplex mode is used in cases where there is no need for communication in both directions at the same time; the entire capacity of the channel can be utilized for each direction.

Full-Duplex:

In full-duplex both stations can transmit and receive simultaneously. The full-duplex mode is like a two way street with traffic flowing in both directions at the same time. In full-duplex mode, signals going in one direction share the capacity of the link: with signals going in the other direction. One common example of full-duplex communication is the telephone network. When two people are communicating by a telephone line, both can talk and listen at the same time. The full-duplex mode is used when communication in both directions is required all the time. The capacity of the channel, however, must be divided between the two directions.

NETWORK

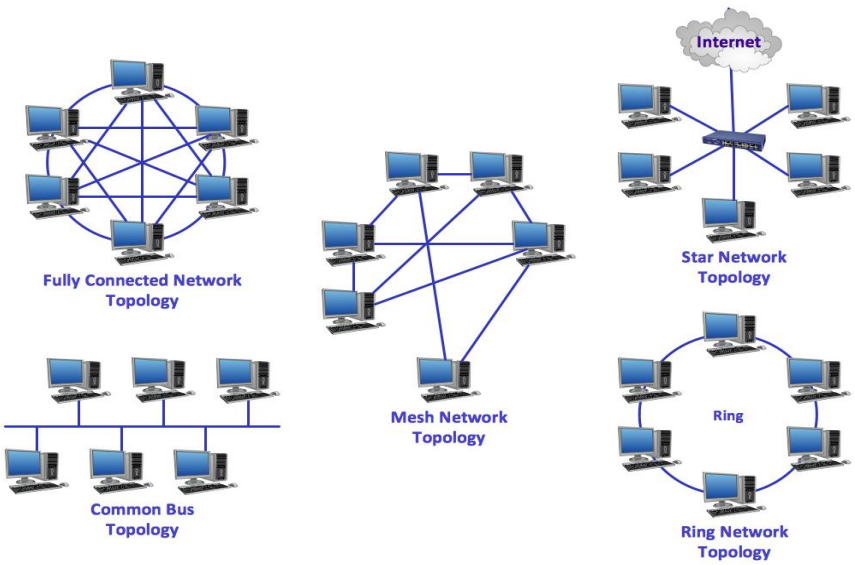
A network is a set of technologies—including hardware, software, and media— that can be used to connect computers together, enabling them to communicate, exchange information, and share resources in real time.

Networks allow many users to access shared data and programs almost instantly. When data and programs are stored on a network and are shared, individual users can substantially reduce the need for programs on their own computers. Networks open up new ways to communicate, such as e-mail and instant messaging.

Data communications usually take place over media {such as cables or wireless links) that arc specifically set up (or the network, and thus are known as dedicated media. The alternative to using dedicated media is to use the telephone system—called the plain old telephone system (POTS)—for data communications. This option is possible because the telephone system is really just a giant network owned by the telephone companies.

TOPOLOGY:

The term physical topology refers to the way in which a network is laid out physically. One or more devices connect to a link; two or more links form a topology. The topology of a network is the geometric representation of the relationship of all the links and linking devices (usually called nodes) to one another. There are four basic topologies possible: mesh, star, bus, and ring.



Mesh:

In a mesh topology, every device has a dedicated point-to-point link to every other device. The term dedicated means that the link carries traffic only between the two devices it connects. To find the number of physical links in a fully connected mesh network with n nodes, we first consider that each

node must be connected to every other node. Node 1 must be connected to n - I nodes, node 2 must be connected to n – 1 nodes, and finally node n must be connected to n - 1 nodes. We need n(n - 1) physical links. However, if each physical link allows communication in both directions (duplex mode), we can divide the number of links by 2. In other words, we can say that in a mesh topology, we need n(n -1) /2 duplex-mode links. To accommodate that many links, every device on the network must have n – 1 input/output ports to be connected to the other n - 1 stations.

Advantages:

1. The use of dedicated links guarantees that each connection can carry its own data load, thus eliminating the traffic problems that can occur when links must be shared by multiple devices.
2. A mesh topology is robust. If one link becomes unusable, it does not incapacitate the entire system. There is the advantage of privacy or security. When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages.
3. There is the advantage of privacy or security. When every message travels along a dedicated line, only the intended recipient sees it. Physical boundaries prevent other users from gaining access to messages
4. Point-to-point links make fault identification and fault isolation easy. Traffic can be routed to avoid links with suspected problems. This facility enables the network manager to discover the precise location of the fault and aids in finding its cause and solution.

Disadvantages:

1. Disadvantage of a mesh are related to the amount of cabling because every device must be connected to every other device.
2. Installation and reconnection are difficult.
3. The sheer bulk of the wiring can be greater than the available space (in walls, ceilings, or floors) can accommodate.
4. The hardware required to connect each link (I/O ports and cable) can be prohibitively expensive.

Star Topology:

In a star topology, each device has a dedicated point-to-point link only to a central controller, usually called a hub. The devices are not directly linked to one another. Unlike a mesh topology, a star topology does not allow direct traffic between devices. The controller acts as an exchange: If one device wants to send data to another, it sends the data to the controller, which then relays the data to the other connected device .

Advantages:

1. A star topology is less expensive than a mesh topology. In a star, each device needs only one link and one I/O port to connect it to any number of others.
2. Easy to install and reconfigure.
3. Far less cabling needs to be housed, and additions, moves, and deletions involve only one connection: between that device and the hub.

4. Other advantage include robustness. If one link fails, only that link is affected. All other links remain active. This factor also lends itself to easy fault identification and fault isolation. As long as the hub is working, it can be used to monitor link problems and bypass defective links.

Disadvantages:

One big disadvantage of a star topology is the dependency of the whole topology on one single point, the hub. If the hub goes down, the whole system is dead. Although a star requires far less cable than a mesh, each node must be linked to a central hub. For this reason, often more cabling is required in a star than in some other topologies (such as ring or bus).

BUS:

A bus topology is multipoint. One long cable acts as a backbone to link all the devices in a network. Nodes are connected to the bus cable by drop lines and taps. A drop line is a connection running between the device and the main cable. A tap is a connector that either splices into the main cable or punctures the sheathing of a cable to create a contact with the metallic core. As a signal travels along the backbone, some of its energy is transformed into heat.

Therefore, it becomes weaker and weaker as it travels farther and farther. For this reason there is a limit on the number of taps a bus can support and on the distance between those taps.

Advantages:

Advantages of a bus topology include ease of installation. Backbone cable can be laid along the most efficient path, then connected to the nodes by drop lines of various lengths. In this way, a bus uses less cabling than mesh or star topologies. In a star, for example, four network devices in the same room require four lengths of cable reaching all the way to the hub. In a bus, this redundancy is eliminated. Only the backbone cable stretches through the entire facility. Each drop line has to reach only as far as the nearest point on the backbone.

Disadvantages:

Disadvantages include difficult reconnection and fault isolation. A bus is usually designed to be optimally efficient at installation. It can therefore be difficult to add new devices. Signal reflection at the taps can cause degradation in quality. This degradation can be controlled by limiting the number and spacing of devices connected to a given length of cable. Adding new devices may therefore require modification or replacement of the backbone. In addition, a fault or break in the bus cable stops all transmission, even between devices on the same side of the problem. The damaged area reflects signals back in the direction of origin, creating noise in both directions.

RING:

In a ring topology, each device has a dedicated point-to-point connection with only the two devices on either side of it. A signal is passed along the ring in one direction, from device to device, until it reaches its destination. Each device in the ring incorporates a repeater. When a device receives a signal intended for another device, its repeater regenerates the bits and passes them along.

Advantages:

A ring is relatively easy to install and reconfigure. Each device is linked to only its immediate neighbors (either physically or logically). To add or delete a device requires changing only two connections. The only constraints are media and traffic considerations (maximum ring length and number of devices). In addition, fault isolation is simplified. Generally in a ring, a signal is circulating at all times. If one device does not receive a signal within a specified period, it can issue an alarm. The alarm alerts the network operator to the problem and its location.

Disadvantages:

Unidirectional traffic can be a disadvantage. In a simple ring, a break in the ring (such as a disabled station) can disable the entire network. This weakness can be solved by using a dual ring or a switch capable of closing off the break. Ring topology was prevalent when IBM introduced its local-area network Token Ring. Today, the need for higher-speed LANs has made this topology less popular.

TYPES OF NETWORKS

There are two main types of networks: local area networks (LANs) and wide area networks (WANs).

Local Area Networks (LANs)

A local area network (LAN) is a data communication system consisting of several devices such as computers and printers. This type of network contains computers that are relatively near each other and are physically connected using cables, infrared links, or wireless media. A LAN can consist of just two or three PCs connected together to share resources, or it can include hundreds of computers of different kinds. Any network that exists within a single building, or even a group of adjacent buildings, is considered a LAN. A LAN is not a system that connects to the public environment (such as the Internet) using phone or data lines.

It is often helpful to connect separate LANs together so they can communicate and exchange data. In a large company, for example, two departments located on the same floor of a building may have their own separate LANs, but if the departments need to share data, then they can create a link between the two LANs.

Wide Area Networks (WANs)

Typically, a wide area network (WAN) is two or more LANs connected together, generally across a wide geographical area. For example, a company may have its corporate headquarters and manufacturing plant in one day and its marketing office in another. Each site needs resources, data, and programs locally, but it also needs to share data with the other sites. To accomplish this feat of data communication, the company can attach devices that connect over public utilities to create a WAN. (Note, however, that a WAN does not have to include any LAN systems. For example, two distant mainframe computers can communicate through a WAN, even though neither is part of a local area network.)

These remote LANs are connected through a telecommunication network (a phone company) or via the Internet through an Internet service provider (ISP) that contracts with the telecommunication networks to gain access to the Internet's backbone.

HYBRID NETWORKS

Between the LAN and WAN structures, you will find hybrid networks such as campus area networks (CANs) and metropolitan area networks (MANs). In addition, a new form of network type is emerging called home area networks (HANs).

The need to access corporate Web sites has created two classifications known as intranets and extranets. The following sections introduce these networks.

Campus Area Networks (CANs)

A campus area network (CAN) follows the same principles as a local area net- work, only on a larger and more diversified scale. With a CAN, different campus offices and organizations can be linked together. For example, in a typical university setting, a bursar's office might be linked to a registrar's office. In this manner once a student has paid his or her tuition fees to the bursar, this information is transmitted to the registrar's system so the student can enroll for classes. Some university departments or organizations might be linked to the CAN even though they already have their own separate LANs.

Metropolitan Area Networks (MANs)

The metropolitan area network (MAN) b a large-scale network that connects multiple corporate LANs together.

MANs usually are not owned by a single organization; their communication devices and equipment are usually maintained by a group or single network provider that sells its networking services to corporate customers. MANs of- ten take the role of a high-speed network that allows for the sharing of regional resources. MANs also can provide a shared connections to other networks using a WAN link.

Home Area Networks (HANs)

A home area network (HAN) is a network contained within a user's home that connects a person's digital devices, from multiple computers and their peripheral devices, such as a printer to telephones, VCRs, DVDs, televisions, video games, home security systems, "smart" appliances, fax machines, and other digital de- and WAVs vices that are wired into the network.

GATEWAY

A gateway is a network node that connects two networks using different protocols together. While a bridge is used to join two similar types of networks, a gateway is used to join two dissimilar networks. The most common gateway is a router that connects a home or enterprise network to the internet. In most IP-based networks, the only traffic that doesn't go through at least one gateway is traffic flowing among nodes on the same local area network (LAN) segment -- for example, computers connected to the same switch.

Gateways can take several forms and perform a variety of tasks.

These include:

Web application firewall - filters traffic to and from a web server and look at application-layer data.

API, SOA or XML gateway - manages traffic flowing into and out of a service, microservices-oriented architecture or an XML-based web service.

IoT gateway - aggregates sensor data, translates between sensor protocols, processes sensor data before sending it onward and more.

Cloud storage gateway - translates storage requests with various cloud storage service API calls.

Media gateway - converts data from the format required for one type of network to the format required for another.
Amazon API Gateway - allows a developer to connect non-AWS applications to AWS back-end resources.

VoIP trunk gateway - facilitates the use of plain old telephone service (POTS) equipment, such as landline phones and fax machines, with a voice over IP (VoIP) network.

Email security gateway - prevents the transmission of emails that break company policy or will transfer information with malicious intent.

E-MAIL/INTERNET CONCEPTS

The Internet

The Internet is not the world wide web - nor is it web browsing. This is a common misconception. Often people will say such things as, "OK, I am finished with email. Now I want to use the Internet." Email is something you can do over the Internet - as is web browsing with Netscape or Internet Explorer.
In other words, the Internet is a collection of computers, networked together throughout the world, and communicating with each other through a common language called TCP/IP (Transfer Control Protocol/Internet Protocol). The **Internet** is the global system of interconnected computer networks that use the Internet protocol suite (TCP/IP) to link devices worldwide. It is a *network of networks* that consists of private, public, academic, business, and government networks of local to global scope, linked by a broad array of electronic, wireless, and optical networking technologies. The Internet carries a vast range of information resources and services, such as the inter-linked hypertext documents and applications of the World Wide Web(WWW), electronic mail, telephony, and file sharing.

The origins of the Internet date back to research commissioned by the United States Federal Government in the 1960s to build robust, fault-tolerant communication via computer networks. The linking of commercial networks and enterprises in the early 1990s marked the beginning of the transition to the modern Internet, and generated rapid growth as institutional, personal, and mobile computers were connected to the network. By the late 2000s, its services and technologies had been incorporated into virtually every aspect of human lives.

Most traditional communications media, including telephony, radio, television, paper mail and newspapers are being reshaped, redefined, or even bypassed by the Internet, giving birth to new services such as email, Internet telephony, Internet television, online music, digital newspapers, and video streaming websites. Newspaper, book, and other print publishing are adapting

to website technology, or are reshaped into blogging, web feeds and online news aggregators. The Internet has enabled and accelerated new forms of personal interactions through instant messaging, Internet forums, and social networking. Online shopping has grown exponentially both for major retailers and small businesses and entrepreneurs, as it enables firms to extend their "brick and mortar" presence to serve a larger market or even sell goods and services entirely online. Business-to-business and financial services on the Internet affect supply chains across entire industries.

The Internet has no centralized governance in either technological implementation or policies for access and usage; each constituent network sets its own policies. Only the overreaching definitions of the two principal name spaces in the Internet, the Internet Protocol address (IP address) space and the Domain Name System (DNS), are directed by a maintainer organization, the Internet Corporation for Assigned Names and Numbers (ICANN). The technical underpinning and standardization of the core protocols is an activity of the Internet Engineering Task Force (IETF), a non-profit organization of loosely affiliated international participants that anyone may associate with by contributing technical expertise.
The various programs you use recognize TCP/IP, and if everything is set up correctly, and the computer is correctly connected to the Internet, all sorts of things can be done, including:

1. Web Crossing
2. email
3. the World Wide Web (WWW)
4. FTP
5. Internet news
6. All sorts of other services

E-MAIL

E-mail or 'electronic mail' allows you to send and receive messages from anyone in the world with an e-mail address. To access e-mail you need an e-mail account, the e-mail address of the person you would like to send a message to, and internet access. To access e-mail you must go to your email provider (i.e. gmail.com, yahoo.com, Hotmail.com etc.) by writing that URL into the address bar.

Email first entered limited use in the 1960s and by the mid-1970s had taken the form now recognized as email. Email operates across computer networks, which today is primarily the Internet. Some early email systems required the author and the recipient to both be online at the same time, in common with instant messaging. Today's email systems are based on a store-and-forward model. Email servers accept, forward, deliver, and store messages. Neither the users nor their computers are required to be online simultaneously; they need to connect only briefly, typically to a mail server or a webmail interface, for as long as it takes to send or receive messages.

Originally an ASCII text-only communications medium, Internet email was extended by Multipurpose Internet Mail Extensions (MIME) to carry text in other character sets and multimedia content attachments. International email, with internationalized email addresses using UTF-8, has been standardized, but as of 2017 it has not been widely adopted.

The history of modern Internet email services reaches back to the early ARPANET, with standards for encoding email messages published as early as 1973 (RFC 561). An email message sent in the early 1970s looks very similar to a basic email sent today. Email had an important role in creating

the Internet and the conversion from ARPANET to the Internet in the early 1980s produced the core of the current services.

INFORMATION RETRIEVAL & STORAGE

INTRODUCTION

Information Retrieval: ***"the purposeful searching for information in a system, of whatever kind, in which information – whether in the form of documents, or their surrogates, or factual material ('information itself'), are stored and represented."*** **or**

Systematic process of collecting and cataloging data so that they can be located and displayed on request. Computers and data processing techniques have made possible to access the high-speed and large amounts of information for government, commercial, and academic purposes.

A branch of computer or library science relating to storage, locating, searching and selecting, upon demand , relevant data on a given subject. (Encyclopedia of Medical concept).

Basic concept of information storage It can refer to a place like a storage room where paper records are kept. It can also refer to a storage device such as a computer hard disk, CD, DVD, or similar device which can hold data. Types of Information storage media Storage keeps data and information for use in the future. Common storage mediums are:

5. Hard Drive
6. Floppy Disk
7. CD&DVD
8. USB Flash Drive

Basic concept of Information Retrieval ***"An information retrieval system is an information system, that is, a system used to store items of information that need to be processed, searched, retrieved, and disseminated to various user populations"*** (Salton, 1983)

COMPONENTS OF IR INFORMATION RETRIEVAL

Major Components of IR Information retrieval can be divided into several major constitutes which include:

1. Database
2. Search mechanism
3. Language
4. Interface

Database:

A system whose base, whose key concepts, is simply a particular way of handling data & its objective is to record and maintain information.

Search mechanism:

Information organized systematically that can be searched and retrieved when a corresponding search mechanism is provided. Search procedures can be categorized as basic or

advance search procedure. Capacity of search mechanism determines what retrieval techniques will be available to users and how information stored in databases can be retrieved.

Language:

Information relies on language when being processed, transferred or communicated. Language can be identified as natural language and controlled vocabulary.

Natural Language: Natural language concerned with the interaction between computer and human interaction.

Controlled Vocabulary: Controlled vocabularies are structured hierarchies of terms used to categorize images. Any user-created terms assigned to images by users, such as tags, folksonomies, and keywords

Such vocabularies are typically created and maintained by a particular institution of authority Up-to-date, new terms are immediately available

1. No immediate up to date
2. Words of author liable to be misconstrued
3. Structure Data
4. Incompatibility a barrier to easy exchange
5. Words of author used.
6. Unstructured data
7. Easier exchange of material between databases

Interface:

Interface regularly considered whether or not an information retrieval system is user friendly. Quality of interface checked by interaction mode. Determines the ultimate success of a system for information retrieval

RETRIEVAL TECHNIQUES

Major retrieval techniques are:

1. Basic Retrieval Techniques
2. Advanced Retrieval Techniques

Basic Retrieval Techniques

1. Boolean Searching
2. Case sensitivity searching
3. Truncation
4. Proximity searching
5. Range searching

Boolean Searching: Logical operations are also known as Boolean Logic. When Boolean logic is applied to information retrieval, the three operators, called Boolean operators. The AND operate for narrowing down a search. The OR operates for broadening a search. The NOT operator for excluding unwanted results

Case sensitivity searching: Text sometimes exhibits case sensitivity; that is, words can differ in meaning based on differing use of uppercase and lowercase letters. Words with capital letters do not always have the same meaning when written with lowercase letters. For example, Bill is the first name of former U.S. president William Clinton, who could sign a bill. The opposite term of "case-sensitive" is "case-insensitive" For example, Google searches are generally case-insensitive and Gmail is case-sensitive by default.

Truncation: Truncation allows a search to be conducted for all the different forms of a word having the same common roots. Used symbol (Question mark? , asterisk* and pound sign #) for truncation purpose. A number of different options are available for truncation like Left truncation, Right truncation and middle truncation. Left truncation retrievals all the words having the same characteristics at the right hand part, for example, *hyl will retrieval words such as "methyl" and "ethyl". Right truncation, for example the term of Network as a query results in retrieving documents on networks and networking. Similarly middle truncation retrieval the words having the same characteristics at the left hand and right hand part, for example, "Colo*r" will retrieval both the term "colour" and "color".

Proximity searching: A proximity search allows you to specify how close two (or more) words must be to each other in order to register a match. There are three types of proximity searches:

1. Word proximity
2. Sentence proximity
3. Paragraph proximity

Range searching: It is most useful with numerical information. The following options are usually available for range searching - greater than (>) less than (<) equal to (=) not equal to (/= or o) greater than equal to (>=) less than or equal to (<=). Example of Range Searching To search for documents or items that contain numbers within a range, type your search term and the range of numbers separated by two periods (".."). For example, to search for pencils that cost between $1.50 and $2.50, type the following:

ADVANCED RETRIEVAL TECHNIQUES

1. Fuzzy searching
2. Query expansion
3. Multiple databases searching

Fuzzy searching: It is designed to find out terms that are spelled incorrectly at data entry and query point. For example the term computer could be misspelled as compter, compiter, or comyter. Optical Character Recognition (OCR) or compressed texts could also result in erroneous results. Fuzzy searching is designed for detection and correction of spelling errors that result from OCR and text compression.

Query expansion: Query expansion is a retrieval technique that allows the end user to improve retrieval performance by revising search queries based on results already retrieved Start Submit Query Conduct Search Present search result NO Query Expansion YES Satisfied? END

Multiple Database Searching: It means searching more than one IR systems. The need for searching multiple databases seems threefold.

1. First, searching in single IR system may not get what the user is looking for.
2. Secondly, multiple databases searching can serve as a selection tool if the user is not sure which systems would be the best choice for a given query .
3. Third, result obtained from multiple databases searching can suggest or indicate suitable systems for the user to conduct further searches. Examples: EBSCOhost, ProQuest.

Today computers provide us with powerful tools for information handling - for collection, organisation, classification, retrieval and distribution. Computers have been used since the late 1960s for the storage of large databases such as library catalogues and bibliographic references. Development of optical storage media such as CD-ROM has given us the possibility of storing large quantities of text, graphics, pictures, and sound at a low cost. These new optical memories can function as distributed stores for encyclopaedias, databases, books etc. This has stimulated the development of local information systems.

TYPES OF DATABASES

There are a number of types of databases:

1. *Library catalogues* - catalogues covering the holdings (books, reports, journals conference proceedings, etc.) of one or more library.
2. *Bibliographic databases* containing bibliographic references, with or without abstracts.
3. *Reference databases* (in addition to those mentioned under 1. and 2.), for example, current research projects, handbooks, encyclopedias, product suppliers, etc.
4. *Factual databases* or data banks containing information, often in numerical form, which can be used directly, e.g. chemical structures, tables, terminology.
5. *Full-text databases* which contain the complete version of the text of given publications.

INFORMATION RETRIEVAL SYSTEMS

1. Online systems
2. CD-ROM systems
3. OPAC
4. Web information Retrieval Systems

Computerized library catalogues

Computerized library catalogues were first introduced during the late 1960s. The online catalogue, known as the Online Public Access Catalogue, or OPAC, has gradually become more user friendly with the use of menus and simple commands. Access for users is now often in the form of a Web (World Wide Web) interface. Computerised library catalogues usually form an integral part of an automated library system, which includes circulation routines as well as acquisition processing. Computerised library catalogues contain the details of books, conference publications, reports, periodical titles, etc. Note: OPACs do not, as a rule, contain details of individual journal articles.

The computerised library catalogues allow you to:

- check to see if a certain book or journal is available at the library or
- see which books are available on a specific subject
- see whether or not a book is currently available or out on loan.

Online information retrieval from databases

Online information retrieval from databases is the acquisition of information from a distant computer via a terminal or PC, involving an interactive dialogue between enquirer and computer. The computer handles a number of databases stored in electronic form, consisting of references to journal articles, conference papers, reports, books etc, which the Information Retrieval Service (IRS) or 'host' makes available to interested parties, such as university libraries, on a commercial basis.

CD-ROMs and WWW interfaces

CD-ROMs and WWW interfaces have been designed for end-users. They are relatively user-friendly and the search software is (more-or-less) self-explanatory. Today, CD-ROMs often are mounted on a server so in reality the user will not be able to notice any differences between using online databases or a CD-ROM

Computer-based information retrieval

Computer-based information retrieval involves an interactive dialogue between the enquirer and database. The computer matches any input search terms against its files, and then displays any resulting matches. These can then be printed out or downloaded by the searcher. Searches can be carried out directly by end-users or by information specialists acting as intermediaries.

Information is stored in the form of a structured database on a host computer and is available online to users by means of communication networks such as the Internet. Computerised information retrieval or online searching is carried out in the form of a dialogue in real time between the user at his/her computer terminal or personal computer (PC) and the various databases stored on a host computer. The various groups involved in computerised information retrieval are the database producers, the host vendors or system operators, institutions providing terminals or PCs, intermediaries and end-users.

EVALUATION OF INFORMATION RETRIEVAL SYSTEMS

Lancaster states that we can evaluate an information retrieval system by considering the following three issues.

1. How well the system is satisfying its objectives, how well it is satisfying the demands placed upon it
2. How efficiently it is satisfying its objectives and finally
3. Whether the system justifies its existence

Evaluation Measures for Information Retrieval Recall and Precision Measure of whether or not a particular item is retrieved or the extent to which the retrieval of wanted items occurs • The performance of a system is often measured by recall ratio, which denotes the percentage of relevant items retrieved in a given situation. Number of relevant items retrieve Recall= × 100 Total number of relevant items in the collection

Steps for Evaluation

1. Designing the scope of evaluation
2. Designing the evaluation program
3. Execution of the evaluation
4. Analysis and interpretation of results
5. Modifying the system in the light of the evaluation results

FUTURE TRENDS IN ONLINE INFORMATION RETRIEVAL SYSTEMS

1. A great increase in the number of information services that can be accessed from around the world.
2. Specialized systems will be more "user oriented," easily accessible.
3. They should be oriented to natural language rather than controlled vocabularies.
4. Computer aided instruction should be incorporated into systems.
5. Future of on-line systems must require less effort to use.
6. They should adapt to the user rather than expecting the user to adapt to them.

THE ADVANTAGES OF COMPUTERIZED INFORMATION RETRIEVAL

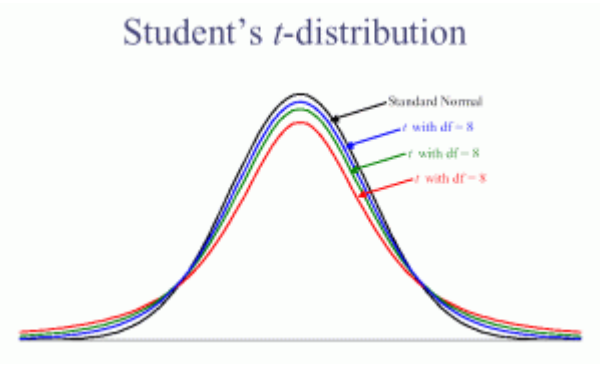
Computerized information retrieval has several advantages over manual methods for literature searching:

1. You save time.
2. Information stored in a database is more current than in the corresponding printed publication.
3. You can search for information in several subject areas during the same search; for example, information on environmental pollution can be searched for in databases covering biology, engineering and chemistry.
4. You can carry out a more detailed search with the help of the computer than by manual methods. In the printed publications, it is usually only possible to search under subject headings or authors, whereas the computerized system permits many more search entries such as institution, title of journal, classification code, keywords or descriptors and words included in the title/abstract. Every unit of information which has been stored in the computer is potentially searchable.

DATA ANALYSIS

T-TEST

The t test (also called Student's T Test) compares two averages (means) and tells you if they are different from each other. The t test also tells you how significant the differences are; In other words it lets you know if those differences could have happened by chance.



A very simple example:

Let's say you have a cold and you try a naturopathic remedy. Your cold lasts a couple of days. The next time you have a cold, you buy an over-the-counter pharmaceutical and the cold lasts a week. You survey your friends and they all tell you that their colds were of a shorter duration (an average of 3 days) when they took the homeopathic remedy. What you *really* want to know is, are these results repeatable? A t test can tell you by comparing the means of the two groups and letting you know the probability of those results happening by chance.

T Score

The t score is a ratio between the **difference between two groups and the difference within the groups**. The larger the t score, the more difference there is between groups. The smaller the t score, the more similarity there is between groups. A t score of 3 means that the groups are three times as different *from* each other as they are within each other. When you run a t test, the bigger the t-value, the more likely it is that the results are repeatable.

- A large t-score tells you that the groups are different.
- A small t-score tells you that the groups are similar.

Calculating the Statistic / Test Types

There are **three main types of t-test:**

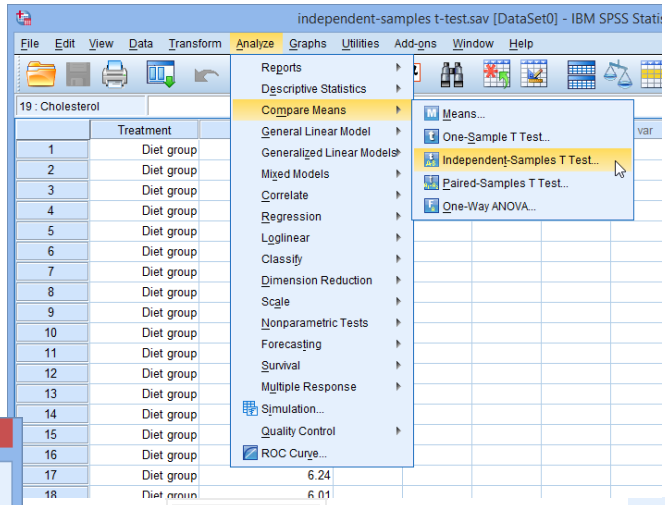
- An Independent Samples t-test compares the means for two groups.
- A Paired sample t-test compares means from the same group at different times (say, one year apart).
- A One sample t-test tests the mean of a single group against a known mean.

SETUP IN SPSS STATISTICS

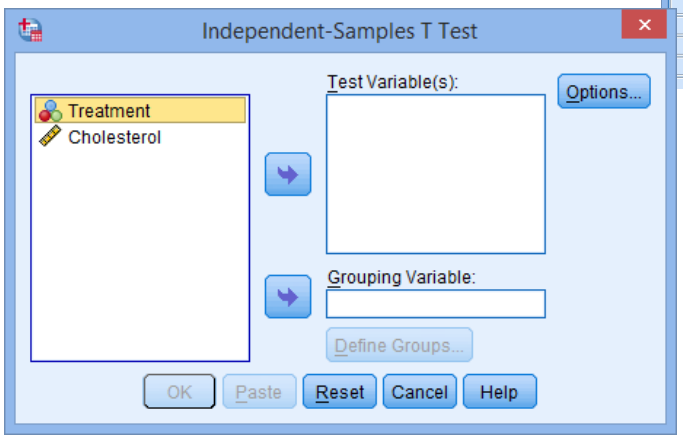
In SPSS Statistics, we separated the groups for analysis by creating a grouping variable called Treatment (i.e., the independent variable), and gave the "diet group" a value of **"1"** and the "exercise group" a value of **"2"** (i.e., the two groups of the independent variable). Cholesterol concentrations were entered under the variable name Cholesterol (i.e., the dependent variable).

Test Procedure in SPSS Statistics

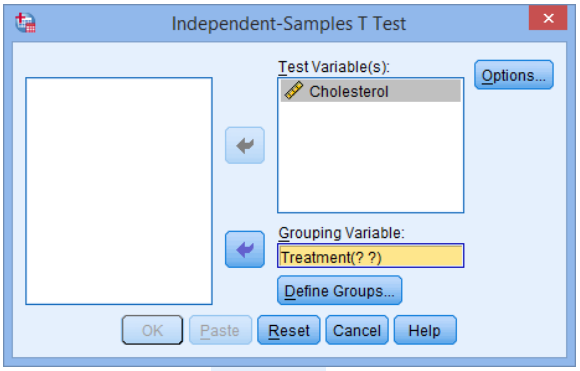
- Click **Analyze** > **Compare Means** > **Independent-Samples T Test...** on the top menu, as shown:



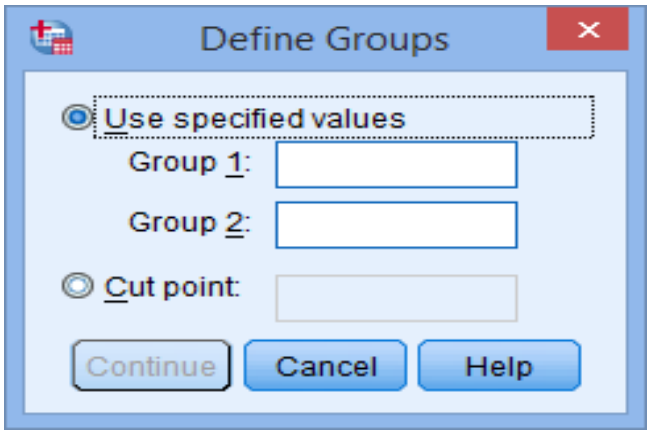
- You will be presented with the **Independent-Samples T Test** dialogue box, as shown below:



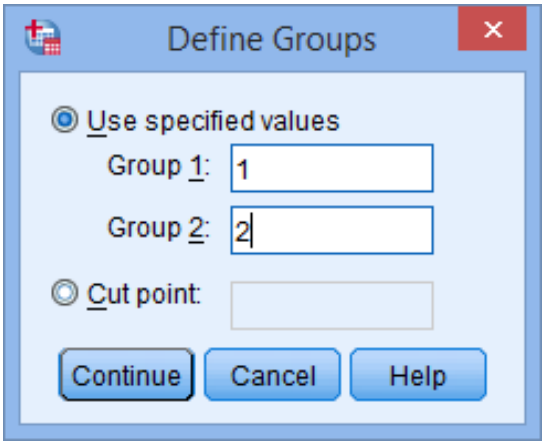
- Transfer the dependent variable, Cholesterol, into the Test Variable(s): box, and transfer the independent variable, Treatment, into the Grouping Variable: box, by highlighting the relevant variables and pressing the  buttons. You will end up with the following screen:



- You then need to define the groups (treatments). Click on the Define Groups... button. You will be presented with the **Define Groups**dialogue box, as shown below:



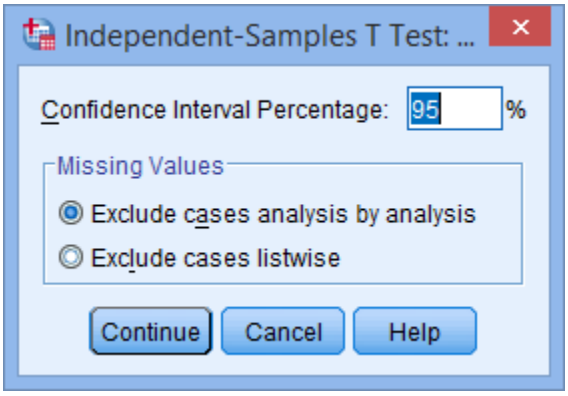
Enter "**1**" into the Group 1: box and enter "**2**" into the Group 2: box. Remember that we labelled the **Diet Treatment** group as **1** and the **Exercise Treatment** group as **2**.



Click the Continue button.

If you need to change the confidence level limits or change how to exclude cases, click the Options... button. You will be presented with the following:

Click the Continue button. You will be returned to the **Independent-Samples T Test** dialogue box. Click the OK button.



This table provides the actual results from the independent t-test.

Independent Samples Test

			Cholesterol Concentration	
			Equal variances assumed	Equal variances not assumed
Levene's Test for Equality of Variances	F		.314	
	Sig.		.579	
t-test for Equality of Means	t		2.428	2.428
	df		38	34.886
	Sig. (2-tailed)		.020	.021
	Mean Difference		.35000	.35000
	Std. Error Difference		.14418	.14418
	95% Confidence Interval of the Difference	Lower	.05813	.05727
		Upper	.64187	.64273

CHI SQUARE

The chi-square test for independence, also called Pearson's chi-square test or the chi-square test of association, is used to discover if there is a relationship between two categorical variables.

Example

Educators are always looking for novel ways in which to teach statistics to undergraduates as part of a non-statistics degree course (e.g., psychology). With current technology, it is possible to present how-to guides for statistical programs online instead of in a book. However, different people learn in different ways. An educator would like to know whether gender (male/female) is associated with the preferred type of learning medium (online vs. books). Therefore, we have two nominal variables: Gender (male/female) and Preferred Learning Medium (online/books).

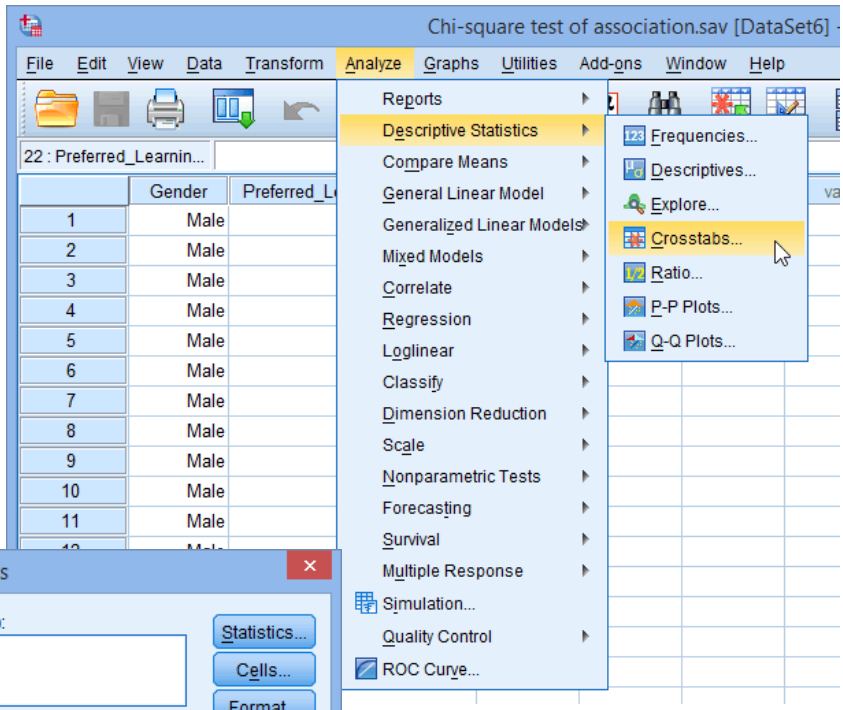
SETUP IN SPSS STATISTICS

In SPSS Statistics, we created two variables so that we could enter our data: Gender and Preferred_Learning_Medium

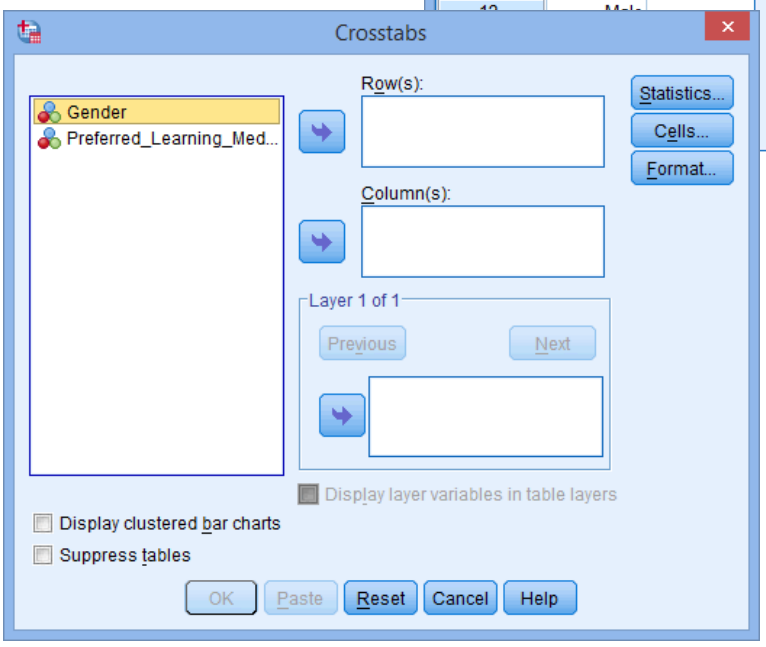
Test Procedure in SPSS Statistics

The 13 steps below show you how to analyse your data using a chi-square test for independence in SPSS Statistics. At the end of these 13 steps, we show you how to interpret the results from your chi-square test for independence.

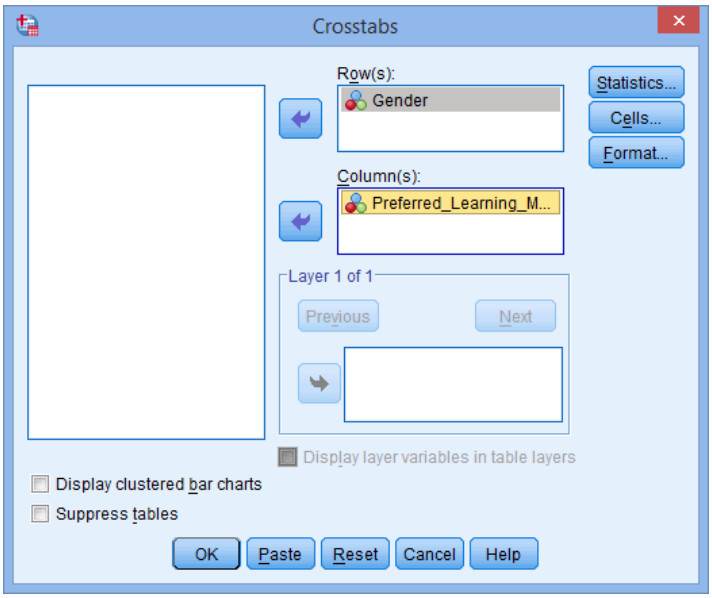
- Click **Analyze** > **Descriptives Statistics** > **Crosstabs...** on the top menu, as shown below:



You will be presented with the following **Crosstabs** dialogue box:

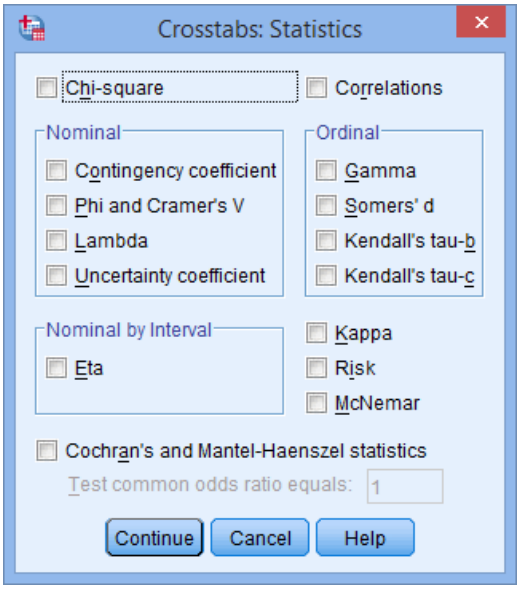


Transfer one of the variables into the Row(s): box and the other variable into the Column(s): box. In our example, we will transfer the Gender variable into the Row(s): box and Preferred_Learning_Medium in to the Column(s): box. There are two ways to do this. You can either: (1) highlight the variable with your mouse and then use the relevant buttons to transfer the variables; or (2) drag-and-drop the variables. How do you know which variable goes in the row or column box? There is no right or wrong way. It will depend on how you want to present your data.

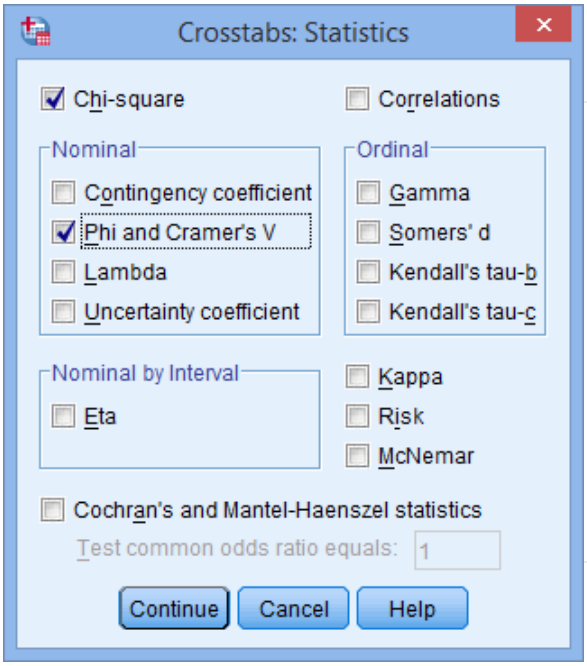


If you want to display clustered bar charts (recommended), make sure that Display clustered bar charts checkbox is ticked.

You will end up with a screen similar to the one below: Click on the Statistics... button. You will be presented with the following **Crosstabs: Statistics** dialogue box:

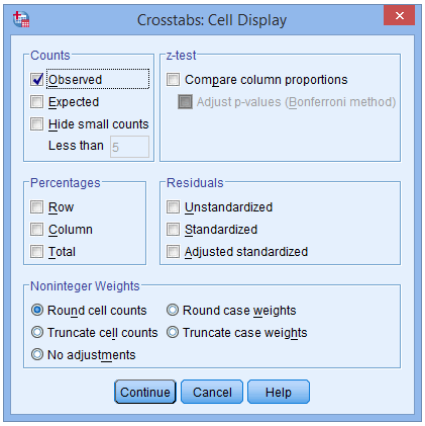


- Select the Chi-square and Phi and Cramer's V options, as shown below:

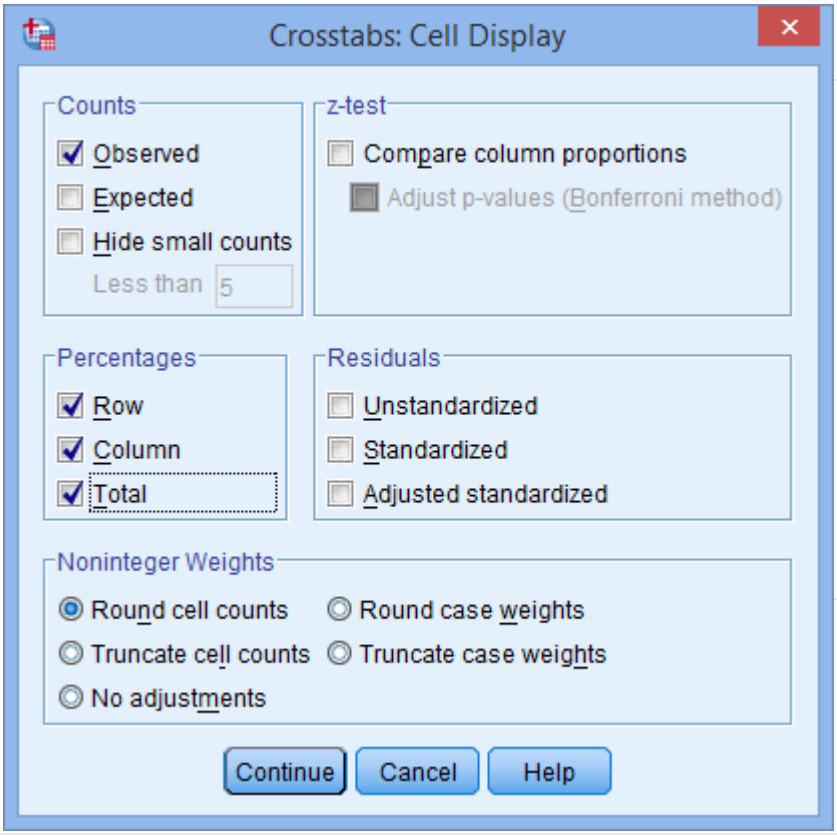


- Click the Continue button.

Click the Cells... button. You will be presented with the following **Crosstabs: Cell Display** dialogue box:



Select Observed from the –Counts– area, and Row, Column and Total from the –Percentages– area, as shown below:

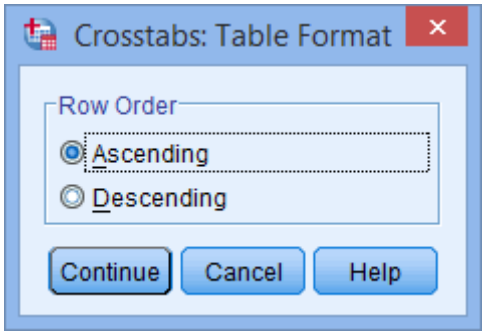


Click the Continue button.

Click the Format... button.

Note: This next option is only really useful if you have more than two categories in one of your variables, but we will show it here in case you have. If you don't, you can skip to STEP 12.

You will be presented with the following:



This option allows you to change the order of the values to either ascending or descending.

- Once you have made your choice, click the Continue button.
- Click the OK button to generate your output.

Output

You will be presented with some tables in the Output Viewer under the title "Crosstabs". The tables of note are presented below:

The Crosstabulation Table (Gender*Preferred Learning Medium Crosstabulation)

Gender * Preferred Learning Medium Crosstabulation

			Preferred Learning Medium		
			Books	Online	Total
Gender	Male	Count	16	24	40
		% within Gender	40.0%	60.0%	100.0%
		% within Preferred Learning Medium	55.2%	47.1%	50.0%
		% of Total	20.0%	30.0%	50.0%
	Female	Count	13	27	40
		% within Gender	32.5%	67.5%	100.0%
		% within Preferred Learning Medium	44.8%	52.9%	50.0%
		% of Total	16.3%	33.8%	50.0%
Total		Count	29	51	80
		% within Gender	36.3%	63.8%	100.0%
		% within Preferred Learning Medium	100.0%	100.0%	100.0%
		% of Total	36.3%	63.8%	100.0%

This table allows us to understand that both males and females prefer to learn using online materials versus books.

The Chi-Square Tests Table

When reading this table we are interested in the results of the "**Pearson Chi-Square**" row. We can see here that $\chi(1) = 0.487$, $p = .485$. This tells us that there is no statistically significant association between Gender and Preferred Learning Medium; that is, both Males and Females equally prefer online learning versus books.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.487[a]	1	.485		
Continuity Correction[b]	.216	1	.642		
Likelihood Ratio	.487	1	.485		
Fisher's Exact Test				.642	.321
Linear-by-Linear Association	.481	1	.488		
N of Valid Cases	80				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 14.50.

b. Computed only for a 2x2 table

The Symmetric Measures Table

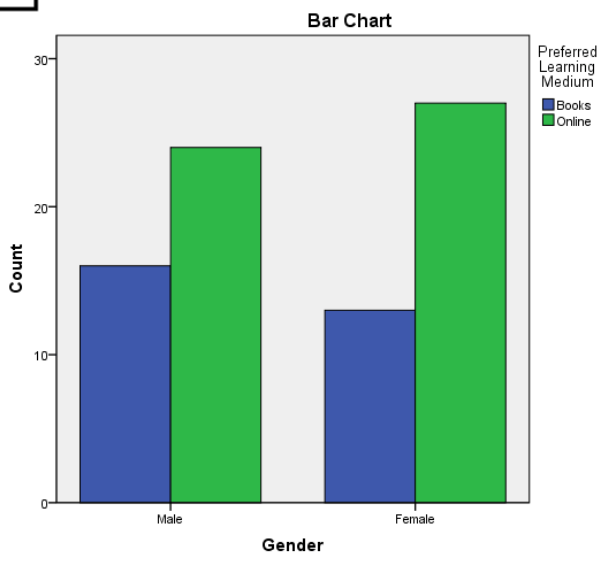
Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	.078	.485
	Cramer's V	.078	.485
N of Valid Cases		80	

Phi and Cramer's V are both tests of the strength of association. We can see that the strength of association between the variables is very weak.

Bar chart

It can be easier to visualize data than read tables. The clustered bar chart option allows a relevant graph to be produced that highlights the group categories and the frequency of counts in these groups.



ANOVA

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups). For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels amongst students, dividing students into three independent groups (e.g., low, medium and high-stressed students). Also, it is important to realize that the one-way ANOVA is an **omnibus** test statistic and cannot tell you which specific groups were statistically significantly different from each other; it only tells you that at least two groups were different. Since you may have three, four, five or more groups in your study design, determining which of these groups differ from each other is important.

Example

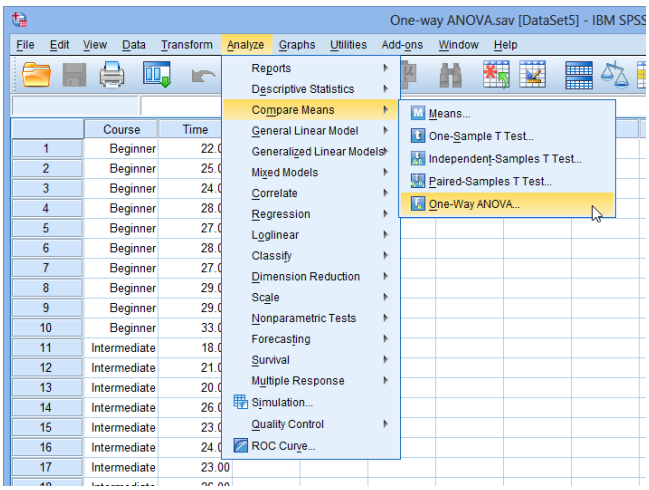
A manager wants to raise the productivity at his company by increasing the speed at which his employees can use a particular spreadsheet program. As he does not have the skills in-house, he employs an external agency which provides training in this spreadsheet program. They offer 3 courses: a beginner, intermediate and advanced course. He is unsure which course is needed for the type of work they do at his company, so he sends 10 employees on the beginner course, 10 on the intermediate and 10 on the advanced course. When they all return from the training, he gives them a problem to solve using the spreadsheet program, and times how long it takes them to complete the problem. He then compares the three courses (beginner, intermediate, advanced) to see if there are any differences in the average time it took to complete the problem.

SETUP IN SPSS STATISTICS

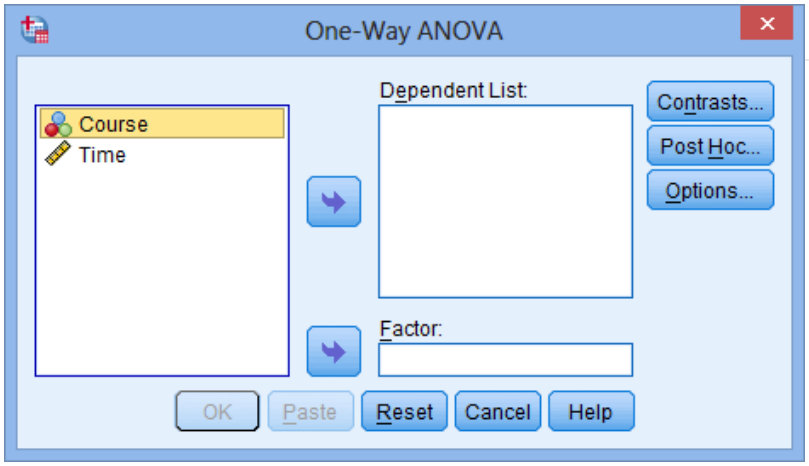
In SPSS Statistics, we separated the groups for analysis by creating a grouping variable called Course (i.e., the independent variable), and gave the beginners course a value of "1", the intermediate course a value of "2" and the advanced course a value of "3". Time to complete the set problem was entered under the variable name Time (i.e., the dependent variable).

Test Procedure in SPSS Statistics

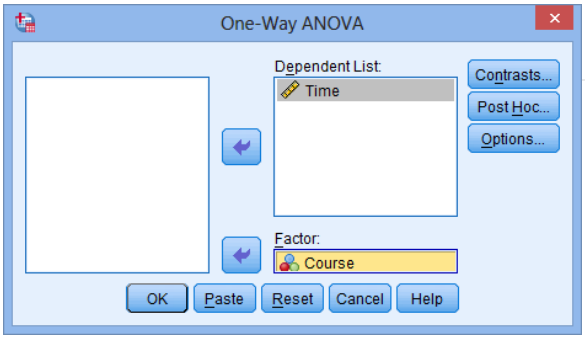
- Click **Analyze** > **Compare Means** > **One-Way ANOVA...** on the top menu, as shown below.



You will be presented with the **One-Way ANOVA** dialogue box:

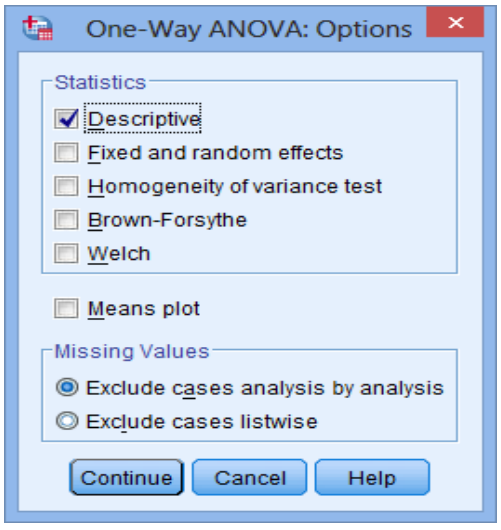


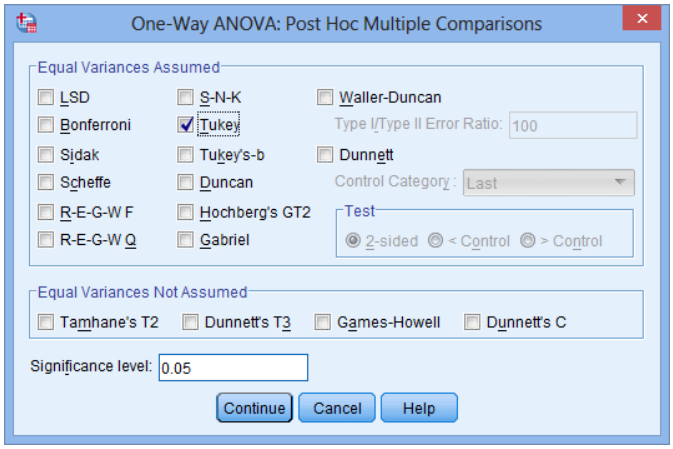
Transfer the dependent variable, Time, into the Dependent List: box and the independent variable, Course, into the Factor:box using the appropriate buttons (or drag-and-drop the variables into the boxes), as shown below:



Click the Post Hoc... button. Tick the Tukey checkbox as shown below:Click the Continue button.

- Click the Options... button. Tick the Descriptive checkbox in the –Statistics– area, as shown below:





- Click the Continue button.
- Click the OK button.

Descriptives Table

The descriptives table (see below) provides some very useful descriptive statistics, including the mean, standard deviation and 95% confidence intervals for the dependent variable (Time) for each

separate group (Beginners, Intermediate and Advanced), as well as when all groups are combined (Total). These figures are useful when you need to describe your data.

Descriptives

Time

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Beginner	10	27.2000	3.04777	.96379	25.0198	29.3802	22.00	33.00
Intermediate	10	23.6000	3.30656	1.04563	21.2346	25.9654	18.00	29.00
Advanced	10	23.4000	3.23866	1.02415	21.0832	25.7168	18.00	29.00
Total	30	24.7333	3.56161	.65026	23.4034	26.0633	18.00	33.00

ANOVA Table

This is the table that shows the output of the ANOVA analysis and whether there is a statistically significant difference between our group means. We can see that the significance value is 0.021 (i.e., $p = .021$), which is below 0.05. and, therefore, there is a statistically significant difference in the mean length of time to complete the spreadsheet problem between the different courses taken. This is great to know, but we do not know which of the specific groups differed. Luckily, we can find this out in the **Multiple Comparisons** table which contains the results of the Tukey post hoc test.

ANOVA

Time

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	91.467	2	45.733	4.467	.021
Within Groups	276.400	27	10.237		
Total	367.867	29			

Multiple Comparisons Table

From the results so far, we know that there are statistically significant differences between the groups as a whole. The table below, **Multiple Comparisons**, shows which groups differed from each other. The Tukey post hoc test is generally the preferred test for conducting post hoc tests on a one-way ANOVA, but there are many others. We can see from the table below that there is a statistically significant difference in time to complete the problem between the group that took the beginner course and the intermediate course ($p = 0.046$), as well as between the beginner course and advanced course ($p = 0.034$). However, there were no differences between the groups that took the intermediate and advanced course ($p = 0.989$).

Multiple Comparisons

Dependent Variable: Time

Tukey HSD

(I) Course	(J) Course	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Beginner	Intermediate	3.60000*	1.43088	.046	.0523	7.1477
	Advanced	3.80000*	1.43088	.034	.2523	7.3477
Intermediate	Beginner	-3.60000*	1.43088	.046	-7.1477	-.0523
	Advanced	.20000	1.43088	.989	-3.3477	3.7477
Advanced	Beginner	-3.80000*	1.43088	.034	-7.3477	-.2523
	Intermediate	-.20000	1.43088	.989	-3.7477	3.3477

*. The mean difference is significant at the 0.05 level.

SHORTKEYS

BASIC SHORTCUT KEYS

Alt+F	File menu options in current program
Alt + E	Edit options in current program
F1	Universal help (for all programs)
Ctrl +A	Select all text
Ctrl + X	Cut selected item
Shift + Del	Cut selected item
Ctrl + C	Copy selected item
Ctrl + Ins	Copy selected item
Ctrl + V	Paste
Shift + Ins	Paste
Home	Go to beginning of current line
Ctrl + Home	Go to beginning of document
End	Go to end of current line
Ctrl + End	Go to end of document
Shift + Home	Highlight from current position to beginning of line
Shift + End	Highlight from current position to end of line
Ctrl + f	Move one word to the left at a time
Ctrl + g	Move one word to the right at a time.

MICROSOFT WINDOWS SHORTCUT KEYS

Alt + Tab	Switch between open applications
Alt + Shift + Tab	Switch backwards between open applications
Alt + Print Screen	Create screen shot for current program
Ctrl + Alt + Del	Reboot/Windows ® task manager
Ctrl + Esc	Bring up start menu

Alt + Esc	Switch between applications on taskbar
F2	Rename selected icon
F3	Start find from desktop
F4	Open the drive selection when browsing
F5	Refresh contents
Alt + F4	Close current open program
Ctrl + F4	Close window in program
Ctrl + Plus Key	automatically adjust widths of all columns in Windows Explorer
Alt + Enter	Open properties window of selected icon or program
Shift + F10	Simulate right-click on selected item
Shift + Del	Delete programs/files permanently
Holding Shift During Bootup	Boot safe mode or bypass system files
Holding Shift During Bootup	When putting in an audio CD, will prevent CD Player from playing.

WINKEY SHORTCUTS

WINKEY+ D	Bring desktop to the top of other windows
WINKEY+ M	Minimize all windows
WINKEY+ SHIFT + M	Undo the minimize done by WINKEY+ M and WINKEY+ D
WINKEY+ E	Open Microsoft Explorer
WINKEY+ Tab	Cycle through open programs on taskbar
WINKEY+ F	Display the Windows ® Search/Find feature
WINKEY+CTRL+ F	Display the search for computers window
WINKEY+ F1	Display the Microsoft ® Windows ® help
WINKEY+ R	Open the run window
WINKEY+Pause /Break	Open the system properties window
WINKEY+ U	Open utility manager

WINKEY+ L	Lock the computer (Windows XP ® & later)

WORD SHORTCUT KEYS

Ctrl + A	Select all contents of the page
Ctrl + B	Bold highlighted selection
Ctrl + C	Copy selected text
Ctrl + X	Cut selected text
Ctrl + N	Open new/blank document
Ctrl + O	Open options
Ctrl + P	Open the print window
Ctrl + F	Open find box
Ctrl + I	Italicize highlighted selection
Ctrl + K	Insert link
Ctrl + U	Underline highlighted selection
Ctrl + V	Paste
Ctrl + Y	Redo the last action performed
Ctrl + Z	Undo last action
Ctrl + G	Find and replace options
Ctrl + H	Find and replace options
Ctrl + J	Justify paragraph alignment
Ctrl + L	Align selected text or line to the left
Ctrl + Q	Align selected paragraph to the left
Ctrl + E	Align selected text or line to the center
Ctrl + R	Align selected text or line to the right
Ctrl + M	Indent the paragraph
Ctrl + T	Hanging indent
Ctrl + D	Font options
Ctrl + Shift + F	Change the font

Ctrl + Shift + >	Increase selected font +1
Ctrl +]	Increase selected font +1
Ctrl + Shift + <	Decrease selected font -1
Ctrl + [	Decrease selected font -1
Ctrl + Shift + *	View or hide non printing characters
Ctrl + f	Move one word to the left
Ctrl + g	Move one word to the right
Ctrl + h	Move to beginning of the line or paragraph
Ctrl + i	Move to the end of the paragraph
Ctrl + Del	Delete word to right of cursor
Ctrl + Backspace	Delete word to left of cursor
Ctrl + End	Move cursor to end of document
Ctrl + Home	Move cursor to beginning of document
Ctrl + Space	Reset highlighted text to default font
Ctrl + 1	Single-space lines
Ctrl + 2	Double-space lines
Ctrl + 5	1.5-line spacing
Ctrl + Alt + 1	Change text to heading 1
Ctrl + Alt + 2	Change text to heading 2
Ctrl + Alt + 3	Change text to heading 3
F1	Open help
Shift + F3	Change case of selected text
Shift + Insert	Paste
F4	Repeat last action performed (Word 2000 +)
F7	Spell check selected text and/or document
Shift + F7	Activate the thesaurus
F12	Save as

Ctrl + S	Save
Shift + F12	Save
Alt + Shift + D	Insert the current date
Alt + Shift + T	Insert the current time
Ctrl + W	Close document

EXCEL SHORTCUT KEYS

F2	Edit the selected cell
F5	Go to a specific cell
F7	Spell check selected text and/or document
F11	Create chart
Ctrl + Shift + ;	Enter the current time
Ctrl + ;	Enter the current date
Alt + Shift + F1	Insert new worksheet
Shift + F3	Open the Excel ® formula window
Shift + F5	Bring up search box
Ctrl + A	Select all contents of worksheet
Ctrl + B	Bold highlighted selection
Ctrl + I	Italicize highlighted selection
Ctrl + C	Copy selected text
Ctrl + V	Paste
Ctrl + D	Fill
Ctrl + K	Insert link
Ctrl + F	Open find and replace options
Ctrl + G	Open go-to options
Ctrl + H	Open find and replace options
Ctrl + U	Underline highlighted selection
Ctrl + Y	Underline selected text

Ctrl + 5	Strikethrough highlighted selection
Ctrl + O	Open options
Ctrl + N	Open new document
Ctrl + P	Open print dialog box
Ctrl + S	Save
Ctrl + Z	Undo last action
Ctrl + F9	Minimize current window
Ctrl + F10	Maximize currently selected window
Ctrl + F6	Switch between open workbooks/windows
Ctrl + Page up	
& Page Down	Move between Excel ® worksheets in the same document
Ctrl + Tab	Move between two or more open Excel ® files
Alt + =	Create formula to sum all of above cells
Ctrl + ‘	Insert value of above cell into current cell
Ctrl + Shift + !	Format number in comma format
Ctrl + Shift + $	Format number in currency format
Ctrl + Shift + #	Format number in date format
Ctrl + Shift + %	Format number in percentage format
Ctrl + Shift + ^	Format number in scientific format
Ctrl + Shift + @	Format number in time format
Ctrl + g	Move to next section of text
Ctrl + Space	Select entire column
Shift +	Space Select entire row
Ctrl + W	Close document

OUTLOOK SHORTCUT KEYS

Alt + S	Send the email
Ctrl + C	Copy selected text

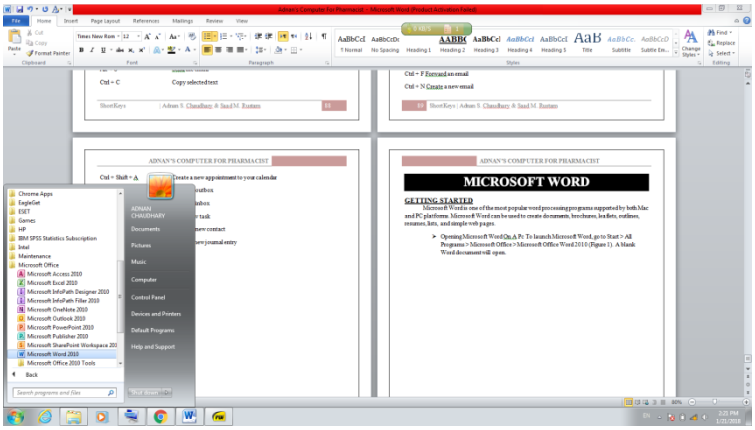
Ctrl + X	Cut selected text
Ctrl + P	Open print dialog box
Ctrl + K	Complete name/email typed in address bar
Ctrl + B	Bold highlighted selection
Ctrl + I	Italicize highlighted selection
Ctrl + U	Underline highlighted selection
Ctrl + R	Reply to an email
Ctrl + F	Forward an email
Ctrl + N	Create a new email
Ctrl + Shift + A	Create a new appointment to your calendar
Ctrl + Shift + O	Open the outbox
Ctrl + Shift + I	Open the inbox
Ctrl + Shift + K	Add a new task
Ctrl + Shift + C	Create a new contact
Ctrl + Shift+ J	Create a new journal entry

MICROSOFT WORD

GETTING STARTED

Microsoft Word is one of the most popular word processing programs supported by both Mac and PC platforms. Microsoft Word can be used to create documents, brochures, leaflets, outlines, resumes, lists, and simple web pages.

- Opening Microsoft Word On A Pc To launch Microsoft Word, go to Start > All Programs > Microsoft Office > Microsoft Office Word 2010 (Figure 1). A blank Word document will open.



- Computers crash and documents are lost all the time, so it is best to save often. Saving Initially Before you begin to type, you should save your document. To do this, go to File Tab (Figure 2) > Save As. Microsoft Word will ask you to choose a location and then browse to a folder to save it in. After selecting these a dialogue box (figure 3) will open and you can name the file. Once you have specified a name and a place for your new file, press the Save button. Note: If you want to save your document on a Mac and then open it on a PC you must specify a file extension (i.e. .doc). Usually your computer will do this for you, but if it does not you must do this process while in Save As. Once you have titled your document, you can give it a file extension by clicking in the Format box. Click Microsoft Word Document for the correct file extension and make sure Append File Extension is checked.

AUTHOR'S

Your Feedback Highly Appreciated. We will add your name as Editor if you highlight mistakes or missing topic in these notes

Adnan Sarwar Chaudhary

deenasaan@gmail.com

03041038728

Saad Muhammad Rustam

Saad.m.rustam@gmail.com

03034565605