



RGPVNOTES.IN

Subject Name: **Web Engineering**

Subject Code: **CS-7003**

Semester: **7th**



LIKE & FOLLOW US ON FACEBOOK

facebook.com/rgpvnotes.in

CS-7003 Web Engineering

UNIT-I

Introduction:

Web engineering is way of developing and organizing knowledge about Web application development and applying that knowledge to develop Web applications, or to address new requirements or challenges. It is also a way of managing the complexity and diversity of Web applications. A Web-based system is a living system. It is like a garden — it continues to evolve, change, and grow. A sound infrastructure must be in place to support the growth of a Web-based system in a controlled, but flexible and consistent manner. Web engineering helps to create an infrastructure that will allow evolution and maintenance of a Web system and that will support creativity.

History-Within a short period, the Internet and World Wide Web have become ubiquitous, surpassing all other technological developments in our history. They have also grown rapidly in their scope and extent of use, significantly affecting all aspects of our lives. Industries such as manufacturing, travel and hospitality, banking, education, and government are Web-enabled to improve and enhance their operations. E-commerce has expanded quickly, cutting across national boundaries. Even traditional legacy information and database systems have migrated to the Web. Advances in wireless technologies and Web-enabled appliances are triggering a new wave of mobile Web applications. As a result, we increasingly depend on a range of Web applications.

Evolution of Web Applications:

The Web has become closely ingrained with our life and work in just a few years. The web today is a growing universe of interlinked web pages and web apps, teeming with videos, photos, and interactive content. What the average user does not see is the interplay of web technologies and browsers that makes all this possible. Over time, web technologies have evolved to give web developers the ability to create new generations of useful and immersive web experiences. Today's web is a result of the ongoing efforts of an open web community that helps define these web technologies, like HTML5, CSS3 and WebGL and ensure that they are support in all web browsers.

Some of the Contributions to the evolution and growth of the Web are-

- **Media:** - Integration of different types of media such as data, text, graphics, images, audio and video, and their presentation (animation, 3D visualization); different types of interaction and channels of communications (one-to-one, one-to-many, many-to-one, and many-to-many).
- **Information science:** - information organization, presentation, indexing, retrieval, aggregation, and management; and collaborative and distributed content creation.
- **Information and communication technology and networking:**- Efficient and cost-effective storage, retrieval, processing, and presentation of information; infrastructures that facilitate transfer and sharing of data and information; wired and wireless Internet communication; and personalized and context-aware Web applications.

Need for Web Engineering:

The web has changed into an environment employed for the delivery of many different types of applications. Such applications range from small-scale information-dissemination-like applications, typically developed by writers and artists, to large-scale commercial, enterprise-planning and scheduling, collaborative-work applications. In today's world many Industries such as Construction, Education, and Hospitality, manufacturing, banking, government and business-utilized web based applications to improve and increase their operations.

Categories of Web Applications: We can categorize web applications as follows:

Document centric web application- Document centric web sites are static html documents stored on web server that sent directly to the client on request. The web pages are manual updated with the help of respective tools. These applications are static, simple, stable and take less time to respond.

Interactive web application- Interactive web application is offer by CGI, HTML Forms. It includes radio buttons, selection menus, forms etc. These applications are simple and fast. In this kind of application, the web pages and links are generating according to user input.

Transactional web application- These kinds of web applications have facility of modification by user. These applications are more interactive and support structured queries from database. The database system handles data consistently and efficiently.

Workflow based web application- These kinds of web applications are capable of handing the workflow among companies, private authorities or public authorities. Web services are included for interoperability. It is robust, reliable and flexible to handle workflow with autonomy of companies. B2B e-commerce solutions are best example of such applications.

Collaborative web application- These kinds of application are mainly use as group applications where group communications are important part. Chatrooms, online forums, e-learning websites or websites where information are shared with option of editing like Wikipedia.

Portal-oriented web application- This kind of web applications are those where single access point is there to separate different sources of information and services. Search engines, community portals etc. are best examples of portal oriented application.

Knowledge-based web application- This kind of application is use for providing knowledge for both human and machine. The knowledge management is based on semantic web technologies. Mining the web, linking and reusing knowledge are a few examples.

Characteristics of Web Applications- The following WebApp characteristics drive the process:

Immediacy- Web-based applications have an immediacy that is not found in any other type of software. That is, the time to market for a complete Web site can be a matter of a few days or weeks. Developers must use methods for planning, analysis, design, implementation, and testing that have been adapted to the compressed time schedules required for WebApp development.

Security- Because WebApps are available via network access, it is difficult, if not impossible, to limit the population of end-users who may access the application. In order to protect sensitive content and provide secure modes of data transmission, strong security measures must implemented throughout the infrastructure that supports a WebApp and within the application itself.

Aesthetics- An undeniable part of the appeal of a WebApp is its look and feel. When an application has been design to market or sell products or ideas, aesthetics may have as much to do with success as technical design.

Web Engineering Models-

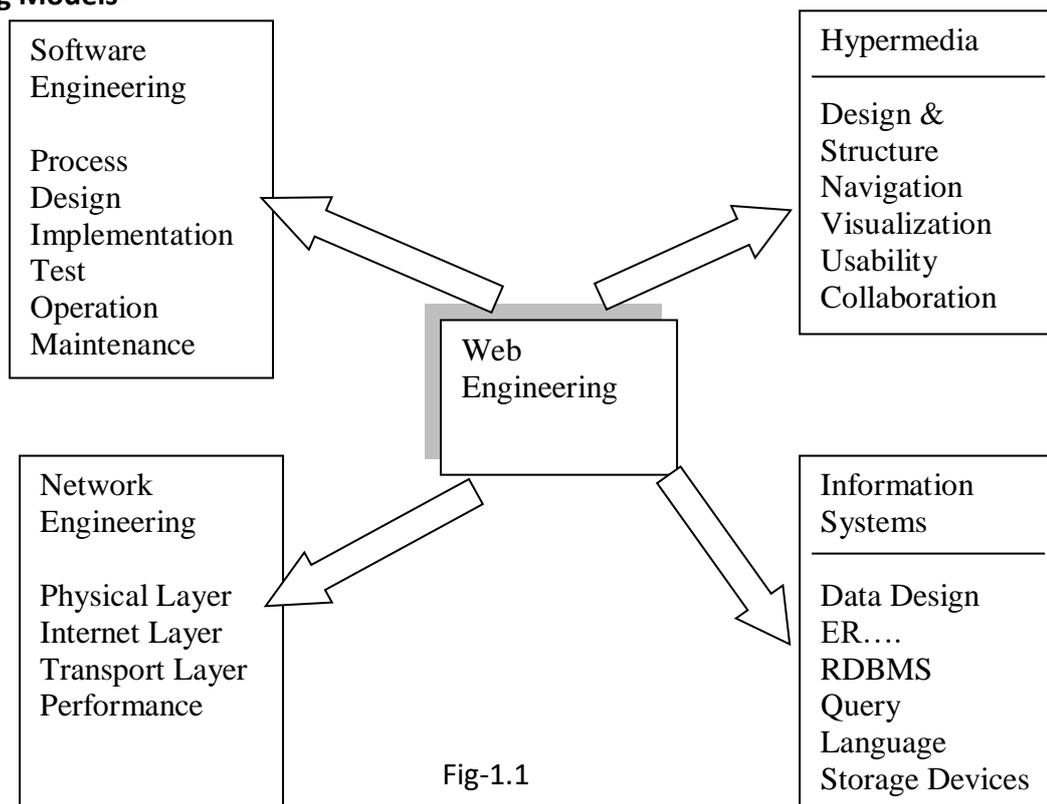


Fig-1.1

Software Engineering v/s Web Engineering-

Software engineering is using software to solve a problem, i.e. applying the science of making software to solve a problem such as making a program or game. Engineering is just solving any problem using science. A software engineer creates software through the practice of software engineering. Software is the product; software engineering is the process to create the product. Software engineering is a branch to make software this is a continuous task while software system is a product, which is made by the software engineering. Software engineering based on computer science, information science and discrete mathematics. In software engineering, two main concerns are cost of development and reliability measured by the no. of errors per thousand lines of source code.

In Web Engineering, Web development is the process of developing the software of web sites, and especially the software unique to each one. That may or may not include the web server and browser, usually includes any web-app frameworks but no other kinds, and generally does not include the languages or other such tools used, but certainly the unique code that runs in order to fulfill any functionality. Software development can be a part of web development, but web development is not always so. When you need a website or a web segment to run on it, you are looking for website development services.

World Wide Web: Stands for "World Wide Web." It is important to know that this is not a synonym for the Internet. The World Wide Web, or just "the Web," as ordinary people call it, is a subset of the Internet. The Web consists of pages that can be accessed using a Web browser. The Internet is the actual network of networks where all the information resides. Things like Telnet, FTP, Internet gaming, Internet Relay Chat (IRC), and e-mail are all part of the Internet, but are not part of the World Wide Web. The Hyper-Text Transfer Protocol (HTTP) is the method used to transfer Web pages to your computer. With hypertext, a word or phrase can contain a link to another Web site.

Berners-Lee developed three essential technologies for WWW:

1. Universal Document Identifier (UDI)
2. Hypertext Mark-up Language (HTML)
3. Hypertext Transfer Protocol (HTTP)

Introduction to TCP/IP-TCP/IP is made up of two acronyms, TCP, for Transmission Control Protocol, and IP, for Internet Protocol. TCP handles packet flow between systems and IP handles the routing of packets. However, we will expound further on that simplistic answer. All modern networks are now designed using a layered approach. Each layer presents a predefined interface to the layer above it. By doing so, a modular design can be developed to minimize problems in the development of new applications or in adding new interfaces.

The ISO/OSI protocol with seven layers is the usual reference model. Since TCP/IP designed before the ISO model was developed, it has four layers; however, the differences between the two are mostly minor **OSI**

Protocol Stack

- | | |
|-----------------|---|
| 7. Application | -- End user services such as email. |
| 6. Presentation | -- Data problems and data compression |
| 5. Session | -- Authentication and authorization |
| 4. Transport | -- Guarantee end-to-end delivery of packets |
| 3. Network | -- Packet routing |
| 2. Data Link | -- Transmit and receive packets |
| 1. Physical | -- The cable or physical connection itself. |

Software Components of TCP/IP-

Application Layer-Some of the applications we will cover are SMTP (mail), Telnet, FTP, Rlogin, NFS, NIS, and LPD Transport Layer The transport uses two protocols, UDP and TCP. UDP, which stands for User Datagram Protocol, does not guarantee packet delivery and applications, which use this, must provide their own means of verifying delivery. TCP does guarantee delivery of packets to the applications, which use it. Network Layer The network layer is concerned with packet routing and used low-level protocols such as ICMP, IP, and IGMP. In addition, routing protocols such as RIP, OSPF, and EGP will discuss.

Link Layer-The link layer is concerned with the actual transmittal of packets as well as IP to Ethernet address translation. This layer is concerned with Arp, the device driver, and RARP.

WAP-In 1997, several companies organized an industry group called the WAP Forum. This group produces the WAP specification, a (long and detailed) series of technical documents that define standards for implementing wireless network applications. Hundreds of industry firms have given strong backing to the WAP Forum, so the technology should become widely adopted, and it is already well-hyped.

These three protocols can think of as "glue layers" in WAP:

- Wireless Transaction Protocol (WTP)
- Wireless Transaction Layer Security (WTLS)
- Wireless Datagram Protocol (WDP)

WTP provides transaction-level services for both reliable and unreliable transports. It prevents duplicate copies of packets from received by a destination, and it supports retransmission, if necessary, in cases where packets dropped. In this respect, WTP is analogous to TCP. However, WTP also differs from TCP. WTP is essentially a pared-down TCP that squeezes some extra performance from the network. WTLS provides authentication and encryption functionality analogous to Secure Sockets Layer (SSL) in Web networking. Like SSL, WTLS is optional and used only when the content server requires it. WDP implements an abstraction layer to lower-level network protocols; it performs functions similar to UDP. WDP is the bottom layer of the WAP stack, but it does not implement physical or data link capability. To build a complete network service, the WAP stack must implemented on some low-level legacy interface not technically part of the model. These interfaces, called bearer services or bearers, can be IP-based or non-IP based. WAP specifies architecture based on layers that follow the OSI model closely

DNS:-Domain Name Server as the name suggests is a distributed naming system for computer, services and resources connected to the internet. The main function is that it translates the domain name to the numerical IP address needed for the location of computers. The domain name system distributes the responsibility of assigning domain names to Authorities name servers. Authorities name servers are assigned to control their particular domain. They can also assign other authorities name servers for their sub domain. This method helps the DNS distribution fault free and avoids the need of the central register to be regularly consulted and updated.

E-MAIL:- Electronic mail (e mail) is one of the use of the Electronic mail (e mail) is one of the use of the World Wide Web, according to most businesses, improves productivity. Traditional methods of sending mail within an office environment are inefficient, as it normally requires an individual requesting a secretary to type the letter. A faster method, and more secure method of sending information is to use electronic mail where by a computer user can exchange messages with other computer users (or groups of users) via a communications network. Electronic mail is one of the most popular uses of the Internet. For example, a memo with 100 words will sent in a fraction of a second. Other types of data can also be sent with mail message such as images, sound, and so on.

The main standards that relate to the protocols of email transmission and reception are:

Simple Mail Transfer Protocol (SMTP) - which is use with the TCP/IP protocol suite? It has traditionally been limited to the text based electronic messages.

Multipurpose Internet Mail Extension (MIME) - Which allows the transmission and reception of mail that contains various types of data, such as speech, images, and motion video? It is a newer standard than STMP and uses much of its basic protocol.

TELNET:-Telnet is a network protocol used on the Internet or local area networks to provide a bidirectional interactive text-oriented communication facility using a virtual terminal connection. User data is interspersed in-band with Telnet control information in an 8-bit byte oriented data connection over the Transmission Control Protocol (TCP).Telnet is a client-server protocol, based on a reliable connection-oriented transport. Typically, this protocol is use to establish a connection to Transmission Control Protocol (TCP) port number 23, where a Telnet server application is listening. Telnet, however, predates TCP/IP and originally run over Network Control Program (NCP) protocols.

TELNET Overview-TELNET is a general protocol, meant to support logging in from almost any type of terminal to almost any type of computer. It allows a user at one site to establish a TCP connection to a login server or terminal server at another site. A TELNET server generally listens on TCP Port 23.

HTTP:- HTTP is a set of standards that allow users of the World Wide Web to exchange information found on web pages. When wanting to access any web page enter http:// in front of the web address, which tells the browser to communicate over HTTP. For example, the full URL for Computer Hope is http://www.computerhope.com. Today's modern browsers no longer require HTTP in front of the URL since it is the default method of communication. However, it still used in browsers because of the need to access other protocols such as FTP through the browser. Below are a few of the major facts on HTTP.

- The term HTTP was coin by Ted Nelson.
- HTTP commonly utilizes port 80, 8008, or 8080.
- HTTP/0.9 was the first version of the HTTP and was introducing in 1991.
- HTTP/1.0 is specify in RFC 1945 and introduced in 1996.

HTTPS- Short for Hypertext Transfer Protocol over Secure, HTTPS is a secure method of accessing or sending information across a web page. All data sent over HTTPS is encrypted before it is sent, this prevents anyone from understanding that information if intercepted. Because data is encrypt over HTTPS, it is slower than HTTP, which is why HTTPS is only use when requiring login information or with pages that contain sensitive in format ion such as an online bank web page.

File Transfer Protocol: – FTP is an open protocol standard that is widely used to transport and receive large files. It can also use to send configuration files and software updates for network switches and routers. FTP uses ports for communications and uses encryption to protect the information being receive and sent.

FTP is, for web developers, a way of moving information from the computer you are working on to the server where a website is host. If you want to install WordPress on a web server, for example, you are going to need FTP to copy the files over.

It is also occasionally use as a way to share files. One person may upload a file to an FTP server, and then share a link to it with another person. This sort of usage has become less common in the age of easy-to-use cloud services, but some people prefer to have their files hosted on a home server, and use FTP to enable that.

How It Works:-TCP and IP are the two major protocols that keep the internet running smoothly. TCP manages data transfer while IP directs traffic to internet addresses. FTP is an underling of TCP and shuttles files back and forth between FTP server and FTP client. Because FTP requires that two ports be open--the server's and the client's--it facilitates the exchange of large files of information. First, you as client make a TCP control connection to the FTP server's port 21, which will remain open during the transfer process. In response, the FTP server opens a second connection that is the data connection from the server's port 20 to your computer. Using the standard active mode of FTP, your computer communicates the port number where it will stand by to receive information from the controller and the IP address-- internet location-- from which or to which you want files to be transferred. If you are using a public--or anonymous--FTP server, you will not need proprietary sign- in information to make a file transfer, but you may asked to enter your email address. If you are using a private FTP server, however, you must sign in with a user name and password to initiate the exchange of data.

Browser and search engines:

Web Browser- A web browser (commonly referred to as a browser) is a software application for retrieving, presenting and traversing information resources on the World Wide Web. A Uniform Resource Identifier (URI/URL) that may be a web page, image identifies an information resource, video or other piece of content. Hyperlinks present in resources enable users easily to navigate their browsers to related resources. The most popular web browsers that are used today are Mozilla Firefox, Google Chrome, Microsoft Internet Explorer, Apple Safari and the Opera browser. These browsers are free and available for download and use.

Search engines -A **web search engine** is a software system that is design to search for information on the World Wide Web. The search results are generally present in a line of results often referred to as search engine results pages. The information may be a mix of web pages, images, and other types of files. Some search engines also mine data available in databases or open directories. Unlike web directories, which are maintain only by human editors, search engines also maintain real-time information by running an algorithm on a web crawler.

Search strategy-A search strategy is a structured organization of terms used to search a database. The search strategy shows how these terms combine in order to retrieve the best results. Different databases work in different ways, so you need to adapt your search strategy for each of the databases you use.

Directories search engines and Meta search engines-

Directories, by contrast, are human-powered. Site owners submit their sites to directories (sometimes for a fee, sometimes free) and human editors determine the value of site, whether it should include in directory. Directory visitors can search the director or drill down to the appropriate category. Although directories have fallen out of fashion (even Yahoo's directory is now hidden at Yahoo.com under the "more" tab), there are still benefits to being listed there. Being list in an important directory helps your search engine visibility because it counts as an incoming link, which is one of the variables in the search engine's algorithm.

Metasearch engines do not build their own index. They piggyback on actual search engines, and make the search engines do the crawling work. Metasearch takes advantage of the indexes created by the bigger search engines, and usually returns results from a number of true search engines like Google, Bing. Some metasearch engines will also include other third-party sources directly in their web search results.

Working of the search engines-

Search engines use programs, often referred to as spiders, robots or crawlers, to search the internet. Those programs used by search engines build an index of internet. However most search engines will provide tips on how to improve your page ranking, the exact algorithms used are well guarded and change frequently to avoid misuse. However, by following a few well-accepted search engine optimization (SEO) techniques you can ensure that your website is well index and remains high in the rankings.

Web server -**Web server** is a computer where the web content is stored. Web server is use to host the web sites but there exists other web servers also such as gaming, storage, FTP, email etc.

Web Server Working

Web server respond to the client request in either of the following two ways:

- Sending the file to the client associated with the requested URL.
- Generating response by invoking a script and communicating with database

Multi-processing-In this single process (parent process) initiates several single-threaded child processes and distribute incoming requests to these child processes. Each of the child processes is responsible for handling single request. It is the responsibility of parent process to monitor the load and decide if processes should be kill or forked.

Multi-threaded-Unlike Multi-process, it creates multiple single-threaded processes.

Hybrid-It is combination of above two approaches. In this approach, multiple processes created and each process initiates multiple threads. Each of the threads handles one connection.

Web servers' features:

- Create one or more websites
- Configure log file settings, including where the log files are saved.
- Configure website/directory security.
- Create an FTP site. An FTP site allows users to transfer files to and from the site.
- Create virtual directories, and map them to physical directories

web caching- A web cache (or HTTP cache) is an information technology for the temporary storage (caching) of web documents, such as HTML pages and images, to reduce server lag. A web cache system stores copies of documents passing through it; subsequent requests may be satisfied from the cache if certain condition is meeting. A web cache system can refer either to an appliance, or to a computer program.

Advantage of web caching

Web caching reduces the number of requests made to the server. Due to which less bandwidth is consume and web server load is reduce. It also helps users to visit a web page if web server is down.

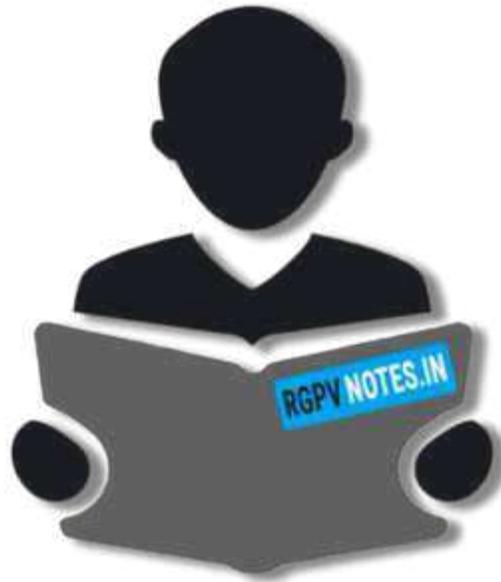
Case Study of IIS

Internet Information Services (IIS) on Windows Server 2012 is NUMA-aware and provides the optimal configuration for the IT administrators. The following section describes how **IIS 8.0** takes advantage of NUMA hardware to provide optimal performance.

IIS supports following two ways of partitioning the workload:

1. Run multiple worker processes in one application. If you are using this mode, by default, the application pool is configuring to run in a single worker process. For maximum performance, you should consider running the same number of worker processes, as there are NUMA nodes, so that there is 1:1 affinity between the worker processes and NUMA nodes. This can be done by setting "Maximum Worker Processes" App Pool setting to 0. When this setting is configuring, IIS will determine how many NUMA nodes are available on the hardware and will start the same number of worker processes.
2. Run multiple applications pools in single workload/site. In this configuration, the workload/site is dividing into multiple application pools. For example, the site may contain several applications that are configured to run in separate application pools. Effectively, this configuration results in running multiple IIS worker processes for the workload/site and IIS intelligently distributes process affinity for maximum performance. Depending upon the workload, administrator partitions the workload into multiple worker processes. Once a workload is correctly partitioned, IIS 8.0 identifies the most optimal NUMA node when the IIS worker process is about to start. By default, IIS picks the NUMA node with the most available memory. IIS has the knowledge of the memory consumption by each NUMA node and uses this information to "load balance" the IIS worker processes. This option is different from Windows default of round robin and specially designed for IIS workload.

Apache-The Apache HTTP Server, commonly referred to as Apache, is a web server application notable for playing a key role in the initial growth of the World Wide Web. Originally based on the NCSA HTTP server, development of Apache began in early 1995 after work on the NCSA code stalled. Apache quickly overtook NCSA HTTP as the dominant HTTP server, and has remained the most popular HTTP server in use since April 1996. In 2009, it became the first web server software to serve more than 100 million websites. Apache is developed and maintained by an open community of developers under the auspices of the Apache Software Foundation.



RGPVNOTES.IN

We hope you find these notes useful.

You can get previous year question papers at
<https://qp.rgpvnotes.in> .

If you have any queries or you want to submit your
study notes please write us at
rgpvnotes.in@gmail.com



LIKE & FOLLOW US ON FACEBOOK
facebook.com/rgpvnotes.in