UNIT-1
Web Intelligence

THINKING AND INTELLIGENT WEB APPLICATIONS

The meaning of the term ―thinking‖ must be provided in the context of intelligent applications on the World Wide Web as it is frequently loosely defined and ambiguously applied.

In general, thinking can be a complex process that uses concepts, their interrelationships, and inference or deduction, to produce new knowledge. However, thinking is often used to describe such disparate acts as memory recall, arithmetic calculations, creating stories, decision making, puzzle solving, and so on.

A person or an individual is considered as an intelligent if he posses qualities like accurate memory recall, the ability to apply valid and correct logic, and the capability to expand their knowledge through learning and deduction.

The term ―intelligence‖ can be applied to nonhuman entities as we do in the field of Artificial Intelligence (AI). But frequently we mean something somewhat different than in the case of human intelligence. For example, a person who performs difficult arithmetic calculations quickly and accurately would be considered as intelligent. But, a computer that could perform the same calculations faster and with greater accuracy would not be considered as an intelligent.

Human thinking involves complicated interactions within the biological components of the brain, and that the process of learning is also an important element of human intelligence.

Software applications perform tasks that are sufficiently complex and human-like that the term ―intelligent‖ may be appropriate. But, Artificial Intelligent (AI) is the science of machines simulating intelligent behavior. The concept of intelligent application on the World Wide Web takes the advantages of AI technologies in order to enhance applications and make them to behave in more intelligent ways.

Here, question arises regarding Web intelligence or intelligent software applications on the World Wide Web. The World Wide Web can be described as an interconnected network of networks. The present day Web consists not only of the interconnected networks, servers, and clients, but also the multimedia hypertext representation of vast quantities of information distributed over an immense global collection of electronic devices with software services being provided over the Web.The current Web consists of static data representations that are designed for direct human access and use.

THE INFORMATION AGE:

We are accustomed to living in a world that is rapidly changing. This is true in all aspects of our society and culture, but is especially true in the field of information technology. It is common to observe such rapid change and comment simply that ―things change.‖

Over the past decades, humanbeings have experienced two global revolutionary changes: the Agricultural Revolution and the Industrial Revolution. Each revolutionary change not only enhanced the access of humanresources but also freed the individuals to achieve higher level cultural and social goals.

In addition, over the past half century, the technological inventions of the Information Age may in fact be of such scope as to represent a third revolutionary change i.e., **the Information Revolution**.

The issue that the ―rapidly changing world of the Information Age be considered a global revolutionary change on the scale of earlier revolutions‖ can be solved by comparing the changes associated with the Agricultural Revolution with the Industrial Revolution.

Before the agriculture revolution, human beings move to warmer regions in the winter season and back to colder regions in the summer seasons. Human beings were able to migrate to all locations on the earth as they have the flexibility of human species and the capability to create adaptable human cultures.

Human beings lived to stay permanently in a single location as soon as they discovered the possibility of cultivating crops. The major implementation of a non migratory life style is that small portion of land could be exploited intensively for long periods of time. Another implication is the agricultural communities concentrated the activities into one or two cucle periods associated with growing and harvesting the crops. This new life style allowed individuals to save their resources and spend on their other activities. In additions it created a great focus on the primary necessity of planting, nurturing and harvesting the crops. The individual become very conscious of time. A part from these, they become reliant on the following:

1. Special skills and knowledge associated with agricultural production.
2. Storage and protection of food supplies.
3. Distribution of products within the community to ensure adequate substenance.
4. Sufficient seed for the near life cycle‘s planning. This life style is different from hunter-gatherer life styles.

The agricultural revolution slowly moved across villages and regions introducing land cultivation as well as a new way of life.

During agricultural revolution human and animal muscle were used to produce the energy required to run the economy. As soon as the French revolution came into existence millions of horses and oxen produced the power required to run the economy.

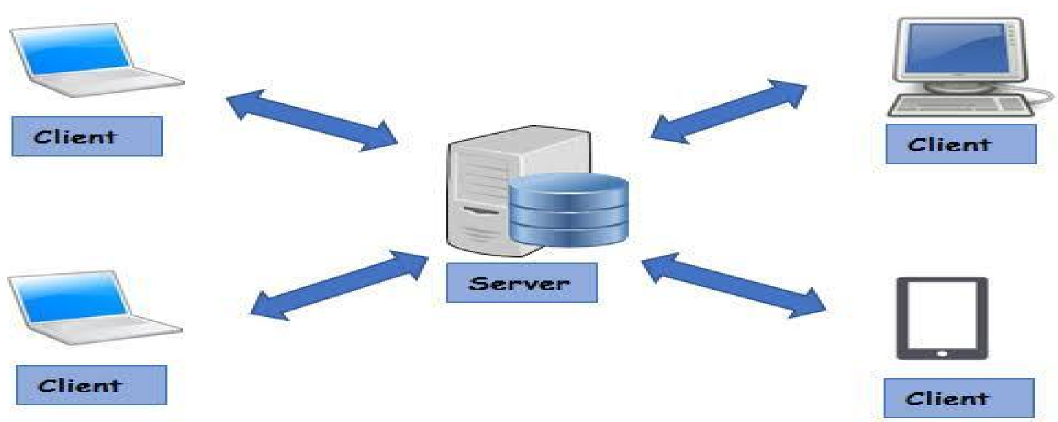
The World Wide Web (WWW):

The World Wide Web (WWW) or the Web is a repository of information spread all over the world and linked together. The WWW has a unique combination of flexibility, portability and user friendly features that distinguish if forms other services provided by the internet.

The WWW project was initiated by CERN (European laboratory for particle physics) to create a system to handle distributed resources necessary for Scientific Research. The WWW today is a distributed client-server service, in which a client using a browser can be access a service using a server. However, the service provided is distributed over many locations called Websites.

The web consists of many web pages that incorporate text, graphics, sound, animation and other multimedia components. These web pages are connected to one another by hypertext. In a hypertext environment the information is stored using the concept of pointers. WWW uses a concept of HTTP which allows communicate between a web browser and web server. The web page can be created by using a HTML. This language has some commands while which are used to inform the web browser about the way of displaying the text, graphics and multimediafiles. HTML also has some commands through which we can give links to the webpages.

The WWW today is a distributed client-server, in which a client using a web browser can access a service using a server.



Distributed client-server Architecture

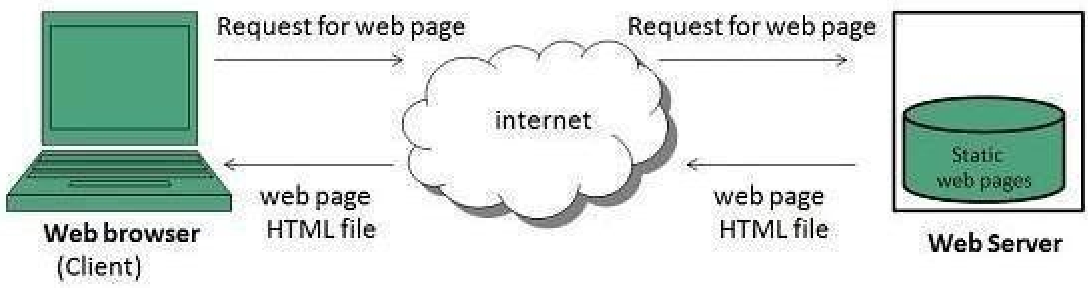
Working of a web:

Web page is a document available on World Wide Web. Web Pages are stored on web server and can be viewed using a web browser.

WWW works on client- server approach. Following steps explains how the web works:

1. User enters the URL (say, **http://www.diet.ac.in**) of the web page in the address bar of web browser.
2. Then browser requests the Domain Name Server for the IP address corresponding to www.diet.ac.in.

3. After receiving IP address, browser sends the request for web page to the web server using HTTP protocol which specifies the way the browser and web server communicates.
4. Then web server receives request using HTTP protocol and checks its search for the requested web page. If found it returns it back to the web browser and close the HTTP connection.
5. Now the web browser receives the web page, It interprets it and display the contents of web page in web browser‘s window.



ARPANET

Licklider, a psychologist and computer scientist put out the idea in 1960 of a network of computers connected together by "wide-band communication lines" through which they could share data and information storage.

Licklider was hired as the head of computer research by the Defense Advanced Research Projects Agency (DARPA), and his small idea took off.

The first ARPANET link was made on October 29, 1969, between the University of California and the Stanford Research Institute. Only two letters were sent before the system crashed, but that was all the encouragement the computer researchers needed. The **ARPANET** became a high-speed digital postoffice as people used it to collaborate on research projects. It was a distributed system of ―many-to-many‖ connections.

Robert Kahn of DARPA and Vinton Cerf of Stanford University worked together on a solution, and in 1977, the **internet protocol suite** was used to seamlessly link three different networks.

Transmission Control Protocol/Internet Protocol (**TCP/IP**), a suite of network communications protocols used to connect hosts on the Internet was developed to connect separate networks into a **"network of networks"** (e.g., the Internet). These protocols specified the framework for a few basic services that everyone would need (file transfer, electronic mail, and remote logon) across a very large number of client and server systems. Several computers linked in a local network can use TCP/IP (along with other protocols) within the local network just as they can use the protocols to provide services throughout the Internet.

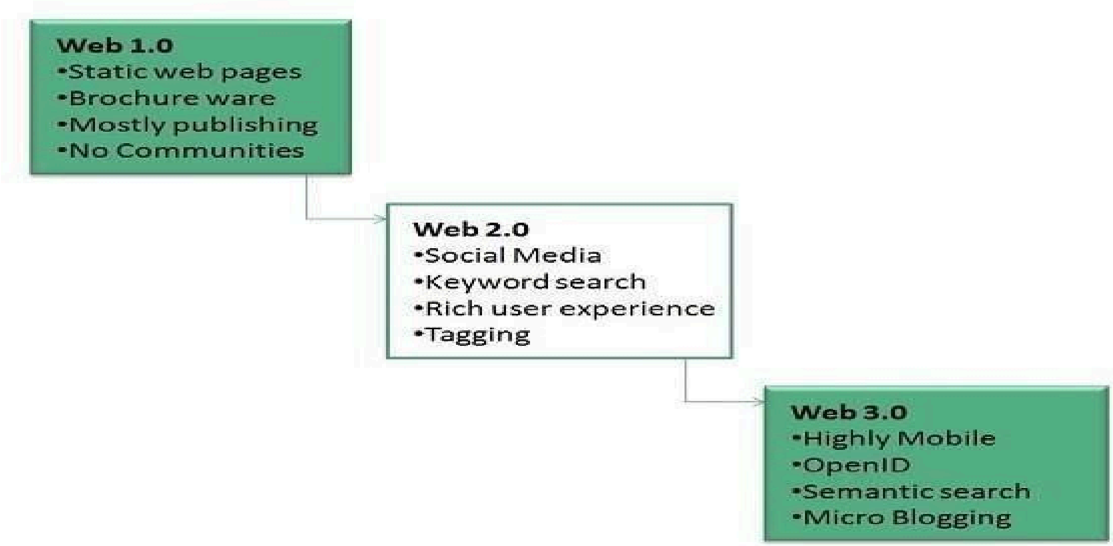
The mid-1980s marked a boom in the personal computer and superminicomputer industries. The combination of inexpensive desktop machines and powerful, network-ready servers allowed many companies to join the Internet for the first time.Corporations began to use the Internet to communicate with each other and with their customers.

By 1990, the ARPANET was decommissioned, leaving only the vast network-of-networks called the Internet with over 300,000 hosts. The stage was set for the final step to move beyond the Internet, as **three major events** and forces converged, accelerating the development of information technology.

These three events were the introduction of the World Wide Web, the widespread availability of the graphical browser, and the unleashing of commercialization.

In 1991, **World Wide Web** was created by **Timothy Berners Lee** in 1989 at **CERN** in **Geneva.** World Wide Web came into existence as a proposal by him, to allow researchers to work together effectively and efficiently at **CERN.** Eventually it became **World Wide Web.**

The following diagram briefly defines evolution of World Wide Web:



The Web combined words, pictures, and sounds on Internet pages and programmers saw the potential for publishing information in a way that could be as easy as using a word processor, but with the richness of multimedia.

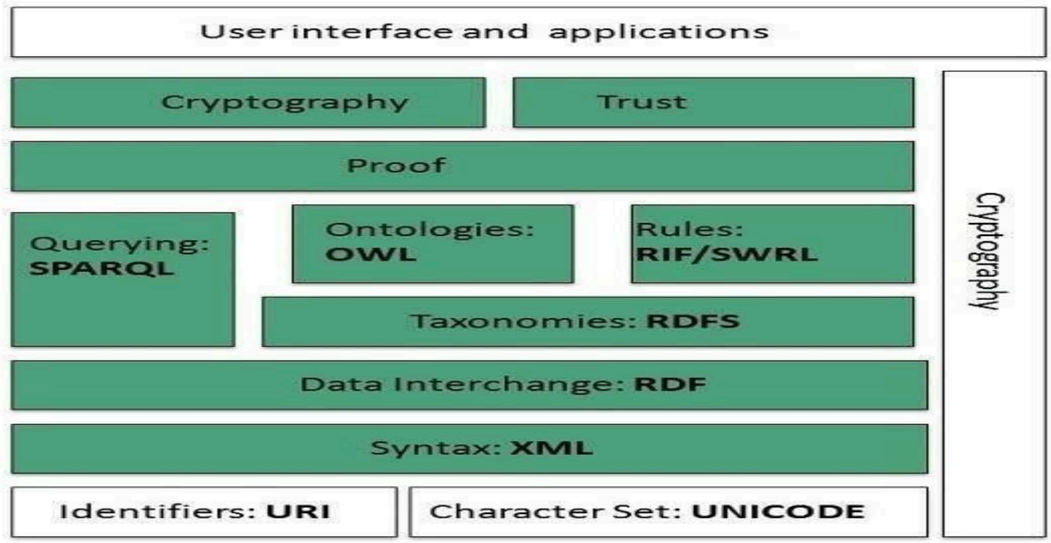
Berners-Lee and his collaborators laid the groundwork for the open standardsof theWeb. Their efforts included the Hypertext Transfer Protocol (HTTP) linking Web documents, the Hypertext Markup Language (HTML) for formatting Web documents, and the Universal Resource Locator (URL) system for addressing Web documents.

The primary language for formatting Web pages is HTML. With HTML the author describes what a page should look like, what types of fonts to use, what color the text should be, where paragraph marks come, and many more aspects of the document. All HTML documents are created by using tags.

In 1993, Marc Andreesen and a group of student programmers at NCSA (the National Center for Supercomputing Applications located on the campus of University of Illinois at Urbana Champaign) developed a graphical browser for the World Wide Web called Mosaic, which he later reinvented commercially as Netscape Navigator.

WWW Architecture

WWW architecture is divided into several layers as shown in the following diagram:



IDENTIFIERS AND CHARACTER SET

Uniform Resource Identifier (URI) is used to uniquely identify resources on the web and **UNICODE** makes it possible to built web pages that can be read and write in human languages.

SYNTAX

XML (Extensible Markup Language) helps to define common syntax in semantic web.

DATA INTERCHANGE

Resource Description Framework (RDF) framework helps in defining core representation of data for web. RDF represents data about resource in graph form.

TAXONOMIES

RDF Schema (RDFS) allows more standardized description of **taxonomies**and other **ontological** constructs.

ONTOLOGIES

Web Ontology Language (OWL) offers more constructs over RDFS. It comes in following three versions:

- OWL Lite for taxonomies and simple constraints.
- OWL DL for full description logic support.
- OWL for more syntactic freedom of RDF

RULES

RIF and **SWRL** offers rules beyond the constructs that are available from **RDFs** and **OWL.** Simple Protocol and **RDF Query Language (SPARQL)** is SQL like language used for querying RDF data and OWL Ontologies.

PROOF

All semantic and rules that are executed at layers below Proof and their result will be used to prove deductions.

CRYPTOGRAPHY

Cryptography means such as digital signature for verification of the origin of sources is used.

USER INTERFACE AND APPLICATIONS

On the top of layer User interface and Applications layer is built for user interaction.

LIMITATIONS OF TODAY'S WEB:

1. The web of today still relies on HTML, which is responsibility for describing how information is to be displayed and laid out on a web.
2. The web today donot have the ability of machine understanding and processing of web-based information.
3. The web is characterized by textual data augmented web services as it involves human assistance and relies on the intevoperation and inefficient exchange of the two competing propritery server frameworks.
4. The web is characterized by textual data augmented by pictorial and audio-visual addition.
5. The web todau is limited to manual keyboard searches as HTML do not have the ability to exploit by information retrieval techniques.
6. Web browsers are limited to access existing informationin a standard form.
7. On web, development of complex networks with meaningful content is difficult.
8. Today‘s web is restricted to search, database, support, intelligent, business logic, automation, security and trust.

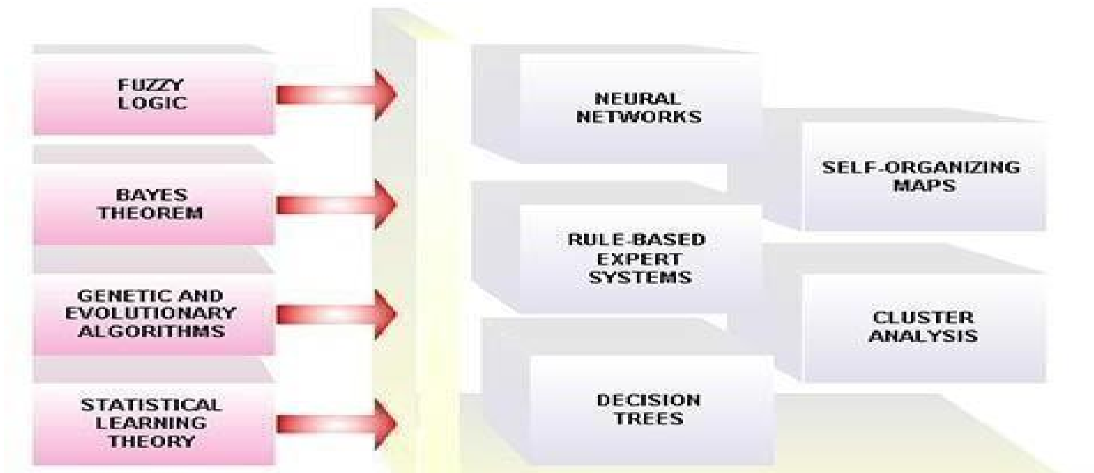
THE NEXT GENERATION WEB

A new Web architecture called the Semantic Web offers users the ability to work on shared knowledge by constructing new meaningful representations on the Web. Semantic Web research has developed from the traditions of AI and ontology languages. It offers automated processing through machine-understandable metadata.

Semantic Web agents could utilize metadata, ontologies, and logic to carry out its tasks. Agents are pieces of software that work autonomously and proactively on the Web to perform certain tasks. In most cases, agents will simply collect and organize information. Agents on the Semantic Web will receive some tasks to perform and seek information from Web resources, while communicating with other Web agents, in order to fulfill its task.

MACHINE INTELLIGENCE **(Also called artificial or computational intelligence)**

Combines a wide variety of advanced technologies to give **machines the ability to learn, adapt, make decisions, and display behaviors not explicitly programmed into their original capabilities**. Some of machine intelligence capabilities, such as **neural networks, expert systems, and self-organizing maps, are plug-in components** – they learn and manage processes at a very high level. Other capabilities, such as fuzzy logic, Bayes Theorem, and genetic algorithms are building blocks – they often provide advanced reasoning and analysis capabilities that are used by other machine reasoning components.
Machine Intelligence capabilities add powerful analytical, self-tuning, self-healing, and adaptive behavior to client applications. They also comprise the core technologies for many of advanced data mining and knowledge discovery services.



ARTIFICIAL INTELLIGENCE

Artificial intelligence (**AI**) is the intelligence of machines and the branch of computer science that aims to create it. AI textbooks define the field as "**the study and design of intelligent agents"** where an **intelligent agent is a system that perceives its environment and takes actions that maximize its chances of success**. John McCarthy, who coined the term in 1955, defines it as **"the science and engineering of making intelligent machines."**

Intelligent agent:

Programs, used extensively on the Web, that perform tasks such as **retrieving and delivering information and automating repetition** More than 50 companies are currently developing intelligent agent software or services, including **Firefly** and **WiseWire.**

Agents are designed to make computing easier. Currently they are used as Web browsers, news retrieval mechanisms, and shopping assistants. By specifying certain parameters, agents will "search" the Internet and return the results directly back to your PC.

Branches of AI

Here's a list, but some branches are surely missing, because no-one has identified them yet.

Logical AI

What a program knows about the world in general the facts of the specific situation in which it must act, and its goals are all represented by sentences of some mathematical logical language.

The program decides what to do by inferring that certain actions are appropriate for achieving its goals.

Search

AI programs often examine large numbers of possibilities, e.g. moves in a chess game or inferences by a theorem proving program. Discoveries are continually made about how to do this more efficiently in various domains.

Pattern recognition

When a program makes observations of some kind, it is often programmed to compare what it sees with a pattern. For example, a vision program may try to match a pattern of eyes and a nose in a scene in order to find a face.

Representation

Facts about the world have to be represented in some way. Usually languages of mathematical logic are used.

Inference

From some facts, others can be inferred. Mathematical logical deduction is adequate for some purposes, but new methods of *non-monotonic* inference have been added to logic since the 1970s. The simplest kind of non-monotonic reasoning is default reasoning in which a conclusion is to be inferred by default, but the conclusion can be withdrawn if there is evidence to the contrary

Common sense knowledge and reasoning

This is the area in which AI is farthest from human-level, in spite of the fact that it has been an active research area since the 1950s. While there has been considerable progress, e.g. in developing systems of *non-monotonic reasoning* and theories of action, yet more new ideas are needed.

Learning from experience

Programs do that. Programs can only learn what facts or behaviors their formalisms can represent, and unfortunately learning systems are almost all based on very limited abilities to represent information.

Planning

Planning programs start with general facts about the world (especially facts about the effects of actions), facts about the particular situation and a statement of a goal. From these, they generate a strategy for achieving the goal. In the most common cases, the strategy is just a sequence of actions.

Epistemology

This is a study of the kinds of knowledge that are required for solving problems in the world.

Ontology

Ontology is the study of the kinds of things that exist. In AI, the programs and sentences deal with various kinds of objects, and we study what these kinds are and what their basic properties are. Emphasis on ontology begins in the 1990s.

Genetic programming

Genetic programming is a technique for getting programs to solve a task by mating random Lisp programs and selecting fittest in millions of generations.

Applications of AI

Game playing

You can buy machines that can play master level chess for a few hundred dollars. There is some AI in them, but they play well against people mainly through brute force computation--looking at hundreds of thousands of positions. To beat a world champion by brute force and known reliable heuristics requires being able to look at 200 million positions per second.

Speech recognition

In the 1990s, computer speech recognition reached a practical level for limited purposes. Thus United Airlines has replaced its keyboard tree for flight information by a system using speech recognition of flight numbers and city names. It is quite convenient. On the the other hand, while it is possible to instruct some computers using speech, most users have gone back to the keyboard and the mouse as still more convenient.

Understanding natural language

Just getting a sequence of words into a computer is not enough. Parsing sentences is not enough either. The computer has to be provided with an understanding of the domain the text is about, and this is presently possible only for very limited domains.

Computer vision

The world is composed of three-dimensional objects, but the inputs to the human eye and computers' TV cameras are two dimensional. Some useful programs can work solely in two dimensions, but full computer vision requires partial three-dimensional information that is not just a set of two-dimensional views. At present there are only limited ways of representing three-dimensional information directly, and they are not as good as what humans evidently use.

Expert systems

A "knowledge engineer" interviews experts in a certain domain and tries to embody their knowledge in a computer program for carrying out some task. How well this works depends on whether the intellectual mechanisms required for the task are within the present state of AI. When this turned out not to be so, there were many disappointing results.An expert system is a computer program that is designed to solve complex problems and to provide decision-making ability like a human expert. It performs this by extracting knowledge from its knowledge base using the reasoning and inference rules according to the user queries.

Heuristic classification

One of the most feasible kinds of expert system given the present knowledge of AI is to put some information in one of a fixed set of categories using several sources of information. An example is advising whether to accept a proposed credit card purchase. Information is available about the owner of the credit card, his record of payment and also about the item he is buying and about the establishment from which he is buying it (e.g., about whether there have been previous credit card frauds at this establishment).

ONTOLOGY

Ontologies are considered one of the pillars of the *Semantic Web*, although they do not have a universally accepted definition. A (Semantic Web) **vocabulary** can be considered as a special form of (usually light-weight) ontology.

"Ontology is a formal specification of a shared conceptualization"

In the context of computer & information sciences, ontology defines a set of representational primitives with which to model a domain of knowledge or discourse. The representational primitives are typically classes (or sets), attributes (or properties), and relationships (or relations among class members).

The definitions of the representational primitives include information about their meaning and constraints on their logically consistent application. In the context of database systems, ontology can be viewed as a level of abstraction of data models, analogous to hierarchical and relational models, but intended for modeling knowledge about individuals, their attributes, and their relationships to other individuals.

Ontologies are typically specified in languages that allow abstraction away from data structures and implementation strategies;

In practice, the languages of ontologies are closer in expressive power to first-order logic than languages used to model databases. For this reason, ontologies are said to be at the ―semantic‖ level, whereas database schema are models of data at the ―logical‖ or ―physical‖ level. Due to their independence from lower level data models, ontologies are used for integrating heterogeneous databases, enabling interoperability among disparate systems, and specifying interfaces to independent, knowledge-based services.

In the technology stack of the Semantic Web standards, ontologies are called out as an explicit layer. There are now standard languages and a variety of commercial and open source tools for creating and working with ontologies.

Ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modeling a domain.

The specification takes the form of the definitions of representational vocabulary (classes, relations, and so forth), which provide meanings for the vocabulary and formal constraints on its coherent use.

KEY APPLICATIONS

Ontologies are part of the W3C standards stack for the Semantic Web, in which they are used to specify standard conceptual vocabularies in which to exchange data among systems, provide services for answering queries, publish reusable knowledge bases, and offer services to facilitate interoperability across multiple, heterogeneous systems and databases.

The key role of ontologies with respect to database systems is to specify a data modeling representation at a level of abstraction above specific database designs (logical or physical), so that data can be exported, translated, queried, and unified across independently developed systems and services. Successful applications to date include database interoperability, cross database search, and the integration of web services.

INFERENCE ENGINE

Inference means **A conclusion reached on the basis of evidence and reasoning**.

In computer science, and specifically the branches of knowledge engineering and artificial intelligence, an **inference engine** is a **“computer program that tries to derive answers from a knowledge base”**. It is the ***"brain"*** that expert systems use to reason about the information in the knowledge base for the ultimate purpose of formulating new conclusions. Inference engines are considered to be a special case of reasoning engines, which can use more general methods of reasoning.

Architecture

The separation of inference engines as a distinct software component stems from the typical production system architecture. This architecture relies on a data store,

1. An **interpreter**. The interpreter executes the chosen agenda items by applying the corresponding base rules.
2. A **scheduler**. The scheduler maintains control over the agenda by estimating the effects of applying inference rules in light of item priorities or other criteria on the agenda.
3. A **consistency enforcer.** The consistency enforcer attempts to maintain a consistent representation of the emerging solution.

Logic:

In logic, a **rule of inference**, **inference rule**, or **transformation rule** is the act of drawing a conclusion based on the form of premises interpreted as a function which takes premises, analyses their syntax, and returns a conclusion (or conclusions). For example, the rule of inference modus ponens takes two premises, one in the form of "If p then q" and another in the form of "p" and returns the conclusion "q". Popular rules of inference include modus ponens, modus tollens from propositional logic and contraposition.

Expert System

In artificial intelligence, an **expert system** is a **computer system that emulates the decision-making ability of a human expert**. Expert systems are designed to solve complex problems by reasoning about knowledge, like an expert, and not by following the procedure of a developer as is the case in conventional programming.

SOFTWARE AGENT

In computer science, a **software agent** is a **software program that acts for a user or other program in a relationship of agency**, which derives from the Latin *agere* (to do): an agreement to act on one's behalf.

The basic attributes of a software agent are that

Agents are **not strictly invoked for a task, but activate themselves,**

Agents may **reside in wait status on a host, perceiving context,**

Agents may **get to run status on a host upon starting conditions,**

Agents do not **require interaction of user,** Agents may **invoke other tasks including communication.**

Various authors have proposed different definitions of agents; these commonly include concepts such as

- ✔ ***Persistence*** (code is not executed on demand but runs continuously and decides for itself when it should perform some activity)
- ✔ ***Autonomy*** (agents have capabilities of task selection, prioritization, goal-directed behavior, decision-making without human intervention)
- ✔ ***Social ability*** (agents are able to engage other components through some sort of communication and coordination, they may collaborate on a task)
- ✔ ***Reactivity*** (agents perceive the context in which they operate and react to it appropriately).

Distinguishing agents from programs

Related and derived concepts include ***Intelligent agents*** (in particular exhibiting some aspect of Artificial Intelligence, such as learning and reasoning), ***autonomous agents*** (capable of modifying the way in which they achieve their objectives), ***distributed agents*** (being executed on physically distinct computers), ***multi-agent systems*** (distributed agents that do not have the capabilities to achieve an objective alone and thus must communicate), and ***mobile agents*** (agents that can relocate their execution onto different processors).

Examples of intelligent software agents

Haag (2006) suggests that there are only four essential types of intelligent software agents:

1. Buyer agents or shopping bots
2. User or personal agents
3. Monitoring-and-surveillance agents
4. Data Mining agents

Buyer agents (shopping bots)

Buyer agents travel around a network (i.e. the internet) retrieving information about goods and services. These agents, also known as 'shopping bots', work very efficiently for commodity products such as CDs, books, electronic components, and other one-size-fits-all products.

User agents (personal agents)

User agents, or personal agents, are **intelligent agents that take action on your behalf**. In this category belong those intelligent agents that already perform, or will shortly perform, the following tasks:

- ✔ ***Check your e-mail***, sort it according to the user's order of preference, and alert you when important emails arrive.
- ✔ ***Play computer games*** as your opponent or patrol game areas for you.
- ✔ ***Assemble customized news reports*** for you.
- ✔ ***Find information*** for you on the subject of your choice.
- ✔ ***Fill out forms*** on the Web automatically for you, storing your information for future reference
- ✔ ***Scan Web pages*** looking for and highlighting text that constitutes the "important" part of the information there
- ✔ ***"Discuss" topics*** with you ranging from your deepest fears to sports
- ✔ ***Facilitate with online job search duties*** by scanning known job boards and sending the resume to opportunities who meet the desired criteria.
- ✔ ***Profile synchronization*** across heterogeneous social networks

Monitoring-and-surveillance (predictive) agents

Monitoring and Surveillance Agents are **used to observe and report on equipment, usually computer systems.** The agents may keep track of company inventory levels, observe competitors' prices and relay them back to the company, watch stock manipulation by insider trading and rumors, etc.
For example, **NASA's Jet Propulsion Laboratory** has an agent that monitors inventory, planning, and scheduling equipment ordering to keep costs down, as well as food storage facilities. These agents usually monitor complex computer networks that can keep track of the configuration of each computer connected to the network.

Data mining agents

This agent uses **information technology to find trends and patterns in an abundance of information from many different sources**. The user can sort through this information in order to find whatever information they are seeking.

A data mining agent operates in a data warehouse discovering information. A 'data warehouse' brings together information from lots of different sources. "Data mining" is the process of looking through the data warehouse to find information that you can use to take action, such as ways to increase sales or keep customers who are considering defecting.

'Classification' is one of the most common types of data mining, which finds patterns in information and categorizes them into different classes.

TIM BERNERS-LEE WWW:

When Tim Berners-Lee was developing the key elements of the World Wide Web, he showed great insight in providing Hypertext Markup Language (HTML) as a simple easy-to-use Web development language.

The continuing evolution of the Web into a resource with intelligent features, however, presents many new challenges. The solution of the World Wide Web Consortium (W3C) is to provide a new Web architecture that uses additional layers of markup languages that can directly apply logic. However, the addition of ontologies, logic, and rule systems for markup languages means consideration of extremely difficult mathematic and logic consequences, such as paradox, recursion, undecidability, and computational complexity on a global scale.

The impact of adding formal logic to Web architecture and present the new markup languages leading to the future Web architecture: the Semantic Web. It concludes with a presentation of complexity theory and rulebased inference engines followed by a discussion of what is solvable on the Web.

THE WORLD WIDE WEB

By 1991, three major events converged to accelerate the development of the Information Revolution. These three events were the introduction of the World Wide Web, the widespread availability of the graphical browser, and the unleashing of commercialization on the Internet. The essential power of the World Wide Web turned out to be its universality though the use of **HTML**. The concept provided the ability to combine words, pictures, and sounds (i.e., to provide multimedia content) on Internet pages.

This excited many computer programmers who saw the potential for publishing information on the Internet with the ease of using a word processor, but with the richness of multimedia forms.

Berners-Lee and his collaborators laid the groundwork for the open standards of the Web. Their efforts included inventing and refining the **Hypertext Transfer Protocol (HTTP)** for linking Web documents, the **HTML** for formatting Web documents and the **Universal Resource Locator (URL)** system for addressing Web documents.

TIM BERNERS-LEE

Tim Berners-Lee was born in London, England, in 1955. His parents were computer scientists who met while working on the Ferranti Mark I, the world‘s first commercially sold computer. He soon developed his parents‘ interest in computers, and at Queen‘s College, at Oxford University, he built his first computer from an old television set and a leftover processor.

Berners-Lee studied physics at Oxford, graduated in 1976. Between 1976 and 1980, he worked at Plessey Telecommunications Ltd. followed by D. G. Nash Ltd. In 1980, he was a software engineer at CERN, the ***European Particle Physics Laboratory, in Geneva, Switzerland*** where he learned the laboratory‘s complicated information system. He wrote a computer program to store information and use random associations that he called, ***"Enquire-Within-Upon- Everything," or "Enquire."*** This system provided links between documents.

In 1989, Berners-Lee with a team of colleagues developed ***HTML***, an easy-to-learn document coding system that allows users to click onto a link in a document‘s text and connect to another document. He also created an addressing plan that allowed each Web page to have a specific location known as a ***URL***. Finally, he completed ***HTTP*** a system for linking these documents across the Internet. He also wrote the software for the first server and the first Web client browser that would allow any computer user to view and navigate Web pages, as well as create and post their own Web documents.

In the following years, Berners-Lee improved the specifications of URLs, HTTP, and HTML as the technology spread across the Internet.

HyperText Markup Language is the primary language for formatting Web pages. The author of a web page uses HTML to describe the attributes of the documents such as,

- what the web page should look like
- what types of fonts to use
- what color text should be
- where paragraphs begin

Hypertext Transfer Protocol

HyperText Transfer Protocol is the network protocol used to deliver files and data on the Web including: HTML files, image files, query results, or anything else. Usually, HTTP takes place through TCP/IP sockets. Socket is the term for the package of subroutines that provide an access path for data transfer through the network.

HTTP uses the client–server model: An HTTP client opens a connection and sends a request message to an HTTP server; the server then returns a response message, usually containing the resource that was requested. After delivering the response, the server closes the connection.

The result of an implementation of XML is referred to as SOAP. ***Simple Object Access Protocol (SOAP)*** is an implementation of XML that represents one common set of rules about how data and commands are represented and extended.

It consists of three parts:

1. **An envelope** (a framework for describing what is in a message and how to process it)
2. **Set of encoding rules** (for expressing instances of application-defined data types)
3. **A convention** (It is used for identifying remote procedure calls and responses.)

SEMANTIC ROADMAP:

Tim Berners - Lee, and his World Wide Web constortium (W3C) team are working collaboratively to develop, extend, and standardize the Web‘s markup languages and tools. In addition, what they are designing is the next generation Web architecture: the Semantic Web.

The goal of the Semantic Web architecture is to provide a knowledge representation of linked data in order to allow machine processing on a global scale. This involves moving the Web from a repository of data without logic to a level where it is possible to express logic through knowledge representation systems.

The vision of the Semantic Web is to expand or increased the existing Web with resources more easily interpreted by programs and intelligent agents.

The existing web involves *two methods* to gain information regarding documents.

The first is to use a directory, or portal site
The directory is constructed manually by searching the Web and then categorizing pages and links. The problem with this approach is that directories take a tremendous effort to maintain. Finding new links, updating old ones, and maintaining the database technology, all add to a portal‘s administrative burden and operating costs.

The second method uses automatic Web crawling and indexing systems.

The future semantic web approaches can produce effective results by using a system that combines the reasoning engine as well as search engine. It will be able to reach out to indexes that contain very complete lists of all occurrences of a given term, and then use logic to weed out all the terms of items that can be used to solve a given problem.

Hence, if the Semantic Web can produce such a structure and meaningful content to the Web, then an environment is created where software agents can perform sophisticated tasks for users.

Logic on the semantic Web

The goal of the Semantic Web is different from most systems of logic. Th eSemantic Web‘s goal is to create a unifying system where a subset is constrained to provide the tractability and efficiency necessary for real applications. However, the Semantic Web itself does not actually define a reasoning engine, but rather follows a proof of a theorem.

This mimics an important comparison between conventional hypertext systems and the original Web design. The original Web design dropped link consistency in favor of expressive flexibility and scalability. The result allowed individual Web sites to have a strict hierarchical order or matrix structure, but it did not require it of the Web as a whole.

As a result, a Semantic Web would actually be a proof validator rather than a theorem prover.

In other words, the Semantic Web cannot find answers, it cannot even check that an answer is correct, but it can follow a simple explanation that an answer is correct. The Semantic Web as a source of data would permit many kinds of automated reasoning systems to function, but it would not be a reasoning system itself.

The objective of the Semantic Web therefore, is to provide a framework that expresses both data and rules for reasoning for Web-based knowledge representation. Adding logic to the Web means using rules to make inferences, choose courses of action, and answering questions. A combination of mathematical and engineering issues complicates this task. The logic must be powerful enough to describe complex properties of objects, but not so powerful that agents can be tricked by being asked to consider a paradox.

The logic of the SemanticWeb is proceeding in a step-by-step approach building one layer on top of another. Three important technologies for developing the Semantic Web are,

1) Resource Description Framework 2) Ontology 3) Web Ontology Language

1. Resource Description Framework

Resource Description Framework is a model of statements made about resources and associated URI. Its statements have a uniform structure of three parts: subject, predicate, and object.

Using RDF, the statements can be formulated in a structured manner. This allows software agents to read as well as act on such statements. The set of statements can be expressed as a graph; a series of (subject, predicate, object) triples, or even in XML forms.

- The first form is the most convenient for communication between people
- The second for efficient processing
- The third one allows as flexible communication with agent software.

2. Ontology

Ontology is an agreement between software agents that exchange information. Thus, the required information is obtained by such an agreement in order to interpret the structure as well as understand the exchanged data and a vocabulary that is used in the exchanges.
Using ontology, agents can exchange new information can be inferred by applying and extending the logical rules present in the ontology.

An ontology that is complex enough to the useful for complex exchanges of information will suffer from the possibility of logical inconsistencies. This is considered as a basic consequence of the insights of Godel‘s incompleteness theorem.

3. Web Ontology Language [OWL]

This language is a vocabulary extension of RDF and is currently evolving into the semantic markup language for publishing and sharing ontologies on the World Wide Web. Web Ontology Language facilitates greater machine readability of Web content than that supported by XML, RDF, and RDFS by providing additional vocabulary along with formal semantics.OWL can be expressed in three sublanguages: OWL Lite, OWL DL, and OWL Full.