UNIT 1 EMERGING DATABASE MODELS, TECHNOLOGIES AND APPLICATIONS-I

Structure **Page Nos.**

1.0 INTRODUCTION

Database technology has advanced from the relational model to the distributed DBMS and Object Oriented databases. The technology has also advanced to support data formats using XML. In addition, data warehousing and data mining technology has become very popular in the industry from the viewpoint of decision making and planning.

Database technology is also being used in advanced applications and technologies. Some of this new application includes multimedia based database applications, geographic applications, Gnome databases, knowledge and spatial databases and many more such applications. These applications require some additional features from the DBMS as they are special in nature and thus are categorised as emerging database technologies.

This unit provides a brief introduction of database requirements of these newer applications.

1.1 OBJECTIVES

After going through this unit, you should be able to:

- define the requirements of a multimedia database systems;
- identify the basic features of geographic databases;
- list the features of Gnome databases;
- differentiate various knowledge databases and their advantages, and
- define the terms information visualisation and spatial databases.

1.2 MULTIMEDIA DATABASE

Multimedia and its applications have experienced tremendous growth. Multimedia data is typically defined as containing digital images, audio, video, animation and graphics along with the textual data. In the past decade, with the advances in network technologies, the acquisition, creation, storage and processing of multimedia data and its transmission over networks have grown tremendously.

A Multimedia Database Management System (MMDBMS) provides support for multimedia data types. It also provides the facilities for traditional DBMS functions like database creation, data modeling, data retrieval, data access and organisation, and data independence. With the rapid development of network technology, multimedia database system, multimedia information exchange is becoming very important. Any such application would require the support for a string multimedia database technology. Let us look into some of the factors that influence the growth of multimedia data.

1.2.1 Factors Influencing the Growth of Multimedia Data

(i) Technological Advancements

Some of the technological advances that attributed to the growth of multimedia data are:

- computers, their computational power and availability,
- availability of high-resolution devices for the capture and display of multimedia data (digital cameras, scanners, monitors, and printers),
- development of high-density storage devices, and
- integration of all such technologies through digital means.

(ii) High Speed Data Communication Networks and Software

Secondly high-speed data communication networks are common these days. These networks not only support high bandwidth but also are more reliable and support digital data transfer. Even the World Wide Web has rapidly grown and software for manipulating multimedia data is now available.

(iii) Applications

With the rapid growth of computing and communication technologies, many applications have come to the forefront. Thus, any such applications in future will support life with multimedia data. This trend is expected to go on increasing in the days to come.

1.2.2 Applications of Multimedia Database

Multimedia data contains some exciting features. They are found to be more effective in dissemination of information in science, engineering, medicine, modern biology, and social sciences. They also facilitate the development of new paradigms in distance learning, and interactive personal and group entertainment.

Some of the typical applications of multimedia databases are:

- media commerce
- medical media databases
- bioinformatics
- ease of use of home media
- news and entertainment
- surveillance

- wearable computing
- management of meeting/presentation recordings
- biometrics (people identification using image, video and/or audio data).

The huge amount of data in different multimedia-related applications needs databases as the basic support mechanism. This is primarily due to the fact that the databases provide consistency, concurrency, integrity, security and availability of data. On the other hand from a user perspective, databases provide ease of use of data manipulation, query and retrieval of meaningful and relevant information from a huge collection of stored data.

Multimedia Databases (MMDBs) must cope with the large volume of multimedia data, being used in various software applications. Some such applications may include digital multimedia libraries, art and entertainment, journalism and so on. Some of these qualities of multimedia data like size, formats etc. have direct and indirect influence on the design and development of a multimedia database.

Thus, a MMDBs needs to provide features of a traditional database as well as some new and enhanced functionalities and features. They must provide a homogenous framework for storing, processing, retrieving, transmitting and presenting a wide variety of multiple media data types available in a large variety of formats.

1.2.3 Contents of MMDB

A MMDB needs to manage the following different types of information with respect to the multimedia data:

Media Data: It includes the media data in the form of images, audio and video. These are captured, digitised, processes, compressed and stored. Such data is the actual information that is to be stored.

Media Format Data: This data defines the format of the media data after the acquisition, processing, and encoding phases. For example, such data may consist of information about sampling rate, resolution, frame rate, encoding scheme etc. of various media data.

Media Keyword Data: This contains the keyword related to the description of media data. For example, for a video, this might include the date, time, and place of recording, the person who recorded, the scene description, etc. This is also known as content description data.

Media Feature Data: This contains the features derived from the media data. A feature characterises the contents of the media. For example, this could contain information on the distribution of colours, the kinds of textures and the different shapes present in an image. This is also referred to as content dependent data.

The last three types are known as meta data as, they describe several different aspects of the media data. The media keyword data and media feature data are used as indices for searching purpose. The media format data is used to present the retrieved information.

1.2.4 Designing MMDBs

The following characteristics of multimedia data have direct and indirect impacts on the design of MMDBs:

- the huge size of MMDBs,
- temporal nature of the data,
- richness of content through media, and

- complexity of representation and subjective interpretation specially from the viewpoint of the meta data.

Challenges in Designing of Multimedia Databases

The major challenges in designing multimedia databases are due to the requirements they need to satisfy. Some of these requirements are:

1) The database should be able to manage different types of input, output, and storage devices. For example, the data may be input from a number of devices that could include scanners, digital camera for images, microphone, MIDI devices for audio, video cameras. Typical output devices are high-resolution monitors for images and video, and speakers for audio.

2) The database needs to handle a variety of data compression and storage formats. Please note that data encoding has a variety of formats even within a single application. For example, in a medical application, the MRI image of the brain should be loss less, thus, putting very stringent quality on the coding technique, while the X-ray images of bones can be coded with lossy techniques as the requirements are less stringent. Also, the radiological image data, the ECG data, other patient data, etc. have widely varying formats.

3) The database should be able to support different computing platforms and operating systems. This is due to the fact that multimedia databases are huge and support a large variety of users who may operate computers and devices suited to their needs and tastes. However, all such users need the same kind of user-level view of the database.

4) Such a database must integrate different data models. For example, the textual and numeric data relating to a multimedia database may be best handled using a relational database model, while linking such data with media data as well as handling media data such as video documents are better done using an object-oriented database model. So these two models need to co-exist in MMDBs.

5) These systems need to offer a variety of query systems for different kinds of media. The query system should be easy-to-use, fast and deliver accurate retrieval of information. The query for the same item sometimes is requested in different forms. For example, a portion of interest in a video can be queried by using either:

   (a) a few sample video frames as an example
   (b) a clip of the corresponding audio track or
   (c) a textual description using keywords.

6) One of the main requirements for such a Database would be to handle different kinds of indices. The multimedia data is in exact and subjective in nature, thus, the keyword-based indices and exact range searches used in traditional databases are ineffective in such databases. For example, the retrieval of records of students based on enrolment number is precisely defined, but the retrieval of records of student having certain facial features from a database of facial images, requires, content-based queries and similarity-based retrievals. Thus, the multimedia database may require indices that are content dependent key-word indices.

7) The Multimedia database requires developing measures of data similarity that are closer to perceptual similarity. Such measures of similarity for different media types need to be quantified and should correspond to perceptual similarity. This will also help the search process.

8) Multimedia data is created all over world, so it could have distributed database features that cover the entire world as the geographic area. Thus, the media data may reside in many different distributed storage locations.

9) Multimedia data may have to be delivered over available networks in real-time. Please note, in this context, the audio and video data is temporal in nature. For example, the video frames need to be presented at the rate of about 30 frames/sec for smooth motion.

10) One important consideration with regard to Multimedia is that it needs to synchronise multiple media types relating to one single multimedia object. Such media may be stored in different formats, or different devices, and have different frame transfer rates.

Multimedia data is now being used in many database applications. Thus, multimedia databases are required for efficient management and effective use of enormous amounts of data.

1.2.5 State of the Art of MMDBMS

The first multimedia database system **ORION** was developed in 1987. The mid 90s saw several commercial MMDBMS being implemented from scratch. Some of them were **MediaDB**, now **MediaWay**, **JASMINE**, and **ITASCA** (the commercial successor of **ORION**). They were able to handle different kinds of data and support mechanisms for querying, retrieving, inserting, and updating data. However, most of these products are not on offer commercially and only some of them have adapted themselves successfully to hardware, software and application changes.

These software are used to provide support for a wide variety of different media types, specifically different media file formats such as image formats, video etc. These files need to be managed, segmented, linked and searched.

The later commercial systems handle multimedia content by providing complex object types for various kinds of media. In such databases the object orientation provides the facilities to define new data types and operations appropriate for the media, such as video, image and audio. Therefore, broadly MMDBMSs are extensible **Object-Relational DBMS (ORDBMSs)**. The most advanced solutions presently include **Oracle 10g**, **IBM DB2** and **IBM Informix**. These solutions purpose almost similar approaches for extending the search facility for video on similarity-based techniques.

Some of the newer projects address the needs of applications for richer semantic content. Most of them are based on the new MPEG-standards MPEG-7 and MPEG-21.

MPEG-7

MPEG-7 is the ISO/IEC 15938 standard for multimedia descriptions that was issued in 2002. It is XML based multimedia meta-data standard, and describes various elements for multimedia processing cycle from the capture, analysis/filtering, to the delivery and interaction.

MPEG-21 is the ISO/IEC 21000 standard and is expected to define an open multimedia framework. The intent is that the framework will cover the entire multimedia content delivery chain including content creation, production, delivery, presentation etc.

Challenges for the Multimedia Database Technologies: Multimedia technologies need to evolve further. Some of the challenges posed by multimedia database applications are:

- the applications utilising multimedia data are very diverse in nature. There is a need for the standardisation of such database technologies,

- technology is ever changing, thus, creating further hurdles in the way of multimedia databases,
- there is still a need to refine the algorithms to represent multimedia information semantically. This also creates problems with respect to information interpretation and comparison.

☞ **Check Your Progress 1**

1) What are the reasons for the growth of multimedia data?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) List four application areas of multimedia databases.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) What are the contents of multimedia database?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

4) List the challenges in designing multimedia databases.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

1.3 SPATIAL DATABASE AND GEOGRAPHIC INFORMATION SYSTEMS

A spatial database keeps track of an object in a multi-dimensional space. A spatial database may be used to represent the map of a country along with the information about railways, roads, irrigation facilities, and so on. Such applications are known as Geographic Information Systems (GIS). Let us discuss GIS in this section.

The idea of a geographic database is to provide geographic information; therefore, they are referred to as the Geographic Information System (GIS). A GIS is basically a collection of the Geographic information of the world. This information is stored and analysed on the basis of data stored in the GIS. The data in GIS normally defines the physical properties of the geographic world, which includes:

- spatial data such as political boundaries, maps, roads, railways, airways, rivers, land elevation, climate etc.
- non-spatial data such as population, economic data etc.

But what are the Applications of the Geographic Databases?

The applications of the geographic databases can be categorised into three broad categories. These are:

- **Cartographic Applications:** These applications revolve around the capture and analysis of cartographic information in a number of layers. Some of the basic applications in this category would be to analyse crop yields, irrigation facility planning, evaluation of land use, facility and landscape management, traffic monitoring system etc. These applications need to store data as per the required applications. For example, irrigation facility management would require study of the various irrigation sources, the land use patterns, the fertility of the land, soil characteristics, rain pattern etc. some of the kinds of data stored in various layers containing different attributes. This data will also require that any changes in the pattern should also be recorded. Such data may be useful for decision makers to ascertain and plan for the sources and types of and means of irrigation.

- **3-D Digital Modelling Applications:** Such applications store information about the digital representation of the land, and elevations of parts of earth surfaces at sample points. Then, a surface model is fitted in using the interpolation and visualisation techniques. Such models are very useful in earth science oriented studies, air and water pollution studies at various elevations, water resource management etc. This application requires data to be represented as attribute based just as in the case of previous applications.

- The third kind of application of such information systems is using the geographic objects applications. Such applications are required to store additional information about various regions or objects. For example, you can store the information about the changes in buildings, roads, over a period of time in a geographic area. Some such applications may include the economic analysis of various products and services etc.

Requirements of a GIS

The data in GIS needs to be represented in graphical form. Such data would require any of the following formats:

- **Vector Data:** In such representations, the data is represented using some geometric objects such as line, square, circle, etc. For example, you can represent a road using a sequence of line segments.
- **Raster Data:** Here, data is represented using an attribute value for each pixel or voxel (a three dimensional point). Raster data can be used to represent three-dimensional elevation using a format termed digital elevation format. For object related applications a GIS may include a temporal structure that records information about some movement related detail such as traffic movement.

A GIS must also support the analysis of data. Some of the sample data analysis operations that may be needed for typical applications are:

- analysing soil erosion
- measurement of gradients
- computing shortest paths
- use of DSS with GIS etc.

One of the key requirements of GIS may be to represent information in an integrated fashion, using both the vector and raster data. In addition it also takes care of data at various temporal structures thus, making it amenable to analysis.

Another question here is the capturing of information in two-dimensional and three-dimensional space in digital form. The source data may be captured by a remote sensing satellite, which can then be, further appended by ground surveys if the need arises. Pattern recognition in this case, is very important for the capture and automating of information input.

Once the data is captured in GIS it may be processed through some special operations. Some such operations are:

- Interpolation for locating elevations at some intermediate points with reference to sample points.
- Some operations may be required for data enhancement, smoothing the data, interpreting the terrain etc.
- Creating a proximity analysis to determine the distances among the areas of interest.
- Performing image enhancement using image processing algorithms for the raster data.
- Performing analysis related to networks of specific type like road network.

The GIS also requires the process of visualisation in order to display the data in a proper visual.

Thus, GIS is not a database that can be implemented using either the relational or object oriented database alone. Much more needs to be done to support them. A detailed discussion on these topics is beyond the scope of this unit.

1.4 GNOME DATABASES

One of the major areas of application of information technology is in the field of Genetics. Here, the computer can be used to create models based on the information obtained about genes. This information models can be used to study:

- the transmission of characteristics from one generation to next,
- the chemical structure of genes and the related functions of each portion of structure, and
- the variations of gene information of all organisms.

Biological data by nature is enormous. Bioinformation is one such key area that has emerged in recent years and which, addresses the issues of information management of genetic data related to DNA sequence. A detailed discussion on this topic is beyond the scope of this unit. However, let us identify some of the basic characteristics of the biological data.

Biological Data – Some Characteristics:

- Biological data consists of complex structures and relationships.
- The size of the data is very large and the data also has a lot of variations across the same type.
- The schema of the database keeps on evolving once or twice a year moreover, even the version of schema created by different people for the same data may be different.
- Most of the accesses to the database would be read only accesses.
- The context of data defines the data and must be preserved along with the data.
- Old value needs to be kept for future references.
- Complex queries need to be represented here.

The Human Genome Initiative is an international research initiative for the creation of detailed genetic and physical maps for each of the twenty-four different human chromosomes and the finding of the complete deoxyribonucleic acid (DNA) sequence of the human genome. The term Genome is used to define the complete genetic information about a living entity. A genetic map shows the linear arrangement of genes or genetic marker sites on a chromosome. There are two types of genetic maps– genetic linkage maps and physical maps. Genetic linkage maps are created on the basis of the frequency with which genetic markers are co-inherited. Physical maps are used to determine actual distances between genes on a chromosome.

The Human Genome Initiative has six strong scientific objectives:

- to construct a high-resolution genetic map of the human genome,
- to produce a variety of physical maps of the human genome,
- to determine the complete sequence of human DNA,
- for the parallel analysis of the genomes of a selected number of well-characterised non-human model organisms,
- to create instrumentation technology to automate genetic mapping, physical mapping and DNA sequencing for the large-scale analysis of complete genomes,
- to develop algorithms, software and databases for the collection, interpretation and dissemination of the vast quantities of complex mapping and sequencing data that are being generated by human genome research.

Genome projects generate enormous quantities of data. Such data is stored in a molecular database, which is composed of an annotated collection of all publicly available DNA sequences. One such database is the Genbank of the National Institutes of Health (NIH), USA. But what would be the size of such a database? In February 2000 the Genbank molecular database contained 5,691,000 DNA sequences, which were further composed of approximately 5,805,000,000 deoxyribonucleotides.

One of the major uses of such databases is in computational Genomics, which refers to the applications of computational molecular biology in genome research. On the basis of the principles of the molecular biology, computational genomics has been classified into three successive levels for the management and analysis of genetic data in scientific databases. These are:

- Genomics.
- Gene expression.
- Proteomics.

1.4.1 Genomics

Genomics is a scientific discipline that focuses on the systematic investigation of the complete set of chromosomes and genes of an organism. Genomics consists of two component areas:

- **Structural Genomics** which refers to the large-scale determination of DNA sequences and gene mapping, and
- **Functional Genomics**, which refers to the attachment of information concerning functional activity to existing structural knowledge about DNA sequences.

Genome Databases

Genome databases are used for the storage and analysis of genetic and physical maps. Chromosome genetic linkage maps represent distances between markers based on meiotic re-combination frequencies. Chromosome physical maps represent distances between markers based on numbers of nucleotides.

Genome databases should define four data types:

- Sequence
- Physical
- Genetic
- Bibliographic

Sequence data should include annotated molecular sequences.

Physical data should include eight data fields:

- Sequence-tagged sites
- Coding regions
- Non-coding regions
- Control regions
- Telomeres
- Centromeres
- Repeats
- Metaphase chromosome bands.

Genetic data should include seven data fields:

- Locus name
- Location
- Recombination distance
- Polymorphisms
- Breakpoints
- Rearrangements
- Disease association
- Bibliographic references should cite primary scientific and medical literature.

Genome Database Mining

Genome database mining is an emerging technology. The process of genome database mining is referred to as computational genome annotation. Computational genome annotation is defined as the process by which an uncharacterised DNA sequence is documented by the location along the DNA sequence of all the genes that are involved in genome functionality.

1.4.2 Gene Expression

Gene expression is the use of the quantitative messenger RNA (mRNA)-level measurements of gene expression in order to characterise biological processes and explain the mechanisms of gene transcription. The objective of gene expression is the quantitative measurement of mRNA expression particularly under the influence of drugs or disease perturbations.

Gene Expression Databases

Gene expression databases provide integrated data management and analysis systems for the transcriptional expression of data generated by large-scale gene expression experiments. Gene expression databases need to include fourteen data fields:

- Gene expression assays
- Database scope
- Gene expression data
- Gene name
- Method or assay
- Temporal information

- Spatial information
- Quantification
- Gene products
- User annotation of existing data
- Linked entries
- Links to other databases
    - o Internet access
    - o Internet submission.

Gene expression databases have not established defined standards for the collection, storage, retrieval and querying of gene expression data derived from libraries of gene expression experiments.

Gene Expression Database Mining

Gene expression database mining is used to identify intrinsic patterns and relationships in gene expression data.

Gene expression data analysis uses two approaches:

- Hypothesis testing and
- Knowledge discovery.

Hypothesis testing makes a hypothesis and uses the results of perturbation of a biological process to match predicted results. The objective of knowledge discovery is to detect the internal structure of the biological data. Knowledge discovery in gene expression data analysis employs two methodologies:

- Statistics functions such as cluster analysis, and
- Visualisation.

Data visualisation is used to display the partial results of cluster analysis generated from large gene expression database cluster.

1.4.3 Proteomics

Proteomics is the use of quantitative protein-level measurements of gene expression in order to characterise biological processes and describe the mechanisms of gene translation. The objective of proteomics is the quantitative measurement of protein expression particularly under the influence of drugs or disease perturbations. Gene expression monitors gene transcription whereas proteomics monitors gene translation. Proteomics provides a more direct response to functional genomics than the indirect approach provided by gene expression.

Proteome Databases

Proteome databases also provide integrated data management and analysis systems for the translational expression data generated by large-scale proteomics experiments. Proteome databases integrate expression levels and properties of thousands of proteins with the thousands of genes identified on genetic maps and offer a global approach to the study of gene expression.

Proteome databases address five research problems that cannot be resolved by DNA analysis:

- Relative abundance of protein products,
- Post-translational modifications,
- Subcellular localisations,
- Molecular turnover and
- Protein interactions.

The creation of comprehensive databases of genes and gene products will lay the foundation for further construction of comprehensive databases of higher-level mechanisms, e.g., regulation of gene expression, metabolic pathways and signalling cascades.

Proteome Database Mining

Proteome database mining is used to identify intrinsic patterns and relationships in proteomics data. Proteome database mining has been performed in areas such as Human Lymphoid Proteins and the evaluation of Toxicity in drug users.

Some Databases Relating to Genome

The following table defines some important databases that have been developed for the Genome.

Database Name	Characteristics	Database Problem Areas
GenBank	Keeps information on the DNA/RNA sequences and information on proteins	Schema is always evolving. This database requires linking to many other databases
GDB	Stores information on genetic map linkages as well as non-human sequence data	It faces the same problem of schema evolution and linking of database. This database also has very complex data objects
ACEDB	Stores information on genetic map linkages as well as non-human sequence data. It uses object oriented database technology	It also has the problem of schema evolution and linking of database. This database also has very complex data objects

A detailed discussion on these databases is beyond the scope of this Unit. You may wish to refer to the further readings for more information.

☞ Check Your Progress 2

1) What is GIS? What are its applications?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

2) List the requirements of a GIS.

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

3) What are the database requirements for Genome?

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

……………………………………………………………………………………

1.5 KNOWLEDGE DATABASES

Knowledge databases are the database for knowledge management. But what is knowledge management? Knowledge management is the way to gather, manage and use the knowledge of an organisation. The basic objectives of knowledge management are to achieve improved performance, competitive advantage and higher levels of innovation in various tasks of an organisation.

Knowledge is the key to such systems. Knowledge has several aspects:

- Knowledge can be implicit (called tacit knowledge) which are internalised or can be explicit knowledge.
- Knowledge can be captured before, during, or even after knowledge activity is conducted.
- Knowledge can be represented in logical form, semantic network form or database form.
- Knowledge once properly represented can be used to generate more knowledge using automated deductive reasoning.
- Knowledge may sometimes be incomplete. In fact, one of the most important aspects of the knowledge base is that it should contain upto date and excellent quality of information.

Simple knowledge databases may consist of the explicit knowledge of an organsiation including articles, user manuals, white papers, troubleshooting information etc. Such a knowledge base would provide basic solutions to some of the problems of the less experienced employees.

A good knowledge base should have:

- good quality articles having up to date information,
- a good classification structure,
- a good content format, and
- an excellent search engine for information retrieval.

One of the knowledge base technologies is based on deductive database technology. Let us discuss more about it in the next sub-section.

1.5.1 Deductive Databases

A deductive database is a database system that can be used to make deductions from the available rules and facts that are stored in such databases. The following are the key characteristics of the deductive databases:

- the information in such systems is specified using a declarative language in the form of rules and facts,
- an inference engine that is contained within the system is used to deduce new facts from the database of rules and facts,
- these databases use concepts from the relational database domain (relational calculus) and logic programming domain (Prolog Language),
- the variant of Prolog known as Datalog is used in deductive databases. The Datalog has a different way of executing programs than the Prolog and
- the data in such databases is specified with the help of facts and rules. For example, The fact that Rakesh is the manager of Mohan will be represented as:

    Manager(Rakesh, Mohan) having the schema:
    Manager(Mgrname, beingmangaed)

Similarly the following represents a rule:

Manager(Rakesh, Mohan) :- Managedby(Mohan, Rakesh)

- Please note that during the representation of the fact the data is represented using the attribute value only and not the attribute name. The attribute name determination is on the basis of the position of the data. For instance, in the example above Rakesh is the Mgrname.
- The rules in the Datalog do not contain the data. These are evaluated on the basis of the stored data in order to deduce more information.

Deductive databases normally operate in very narrow problem domains. These databases are quite close to expert systems except that deductive databases use, the database to store facts and rules, whereas expert systems store facts and rules in the main memory. Expert systems also find their knowledge through experts whereas deductive database have their knowledge in the data. Deductive databases are applied to knowledge discovery and hypothesis testing.

1.5.2 Semantic Databases

Information in most of the database management systems is represented using a simple table with records and fields. However, simple database models fall short of applications that require complex relationships and rich constructs to be represented using the database. So how do we address such a problem? Do we employ object oriented models or a more natural data model that represents the information using semantic models? Semantic modeling provides a far too rich set of data structuring capabilities for database applications. A semantic model contains far too many constructs that may be able to represent structurally complex inter-relations among data in a somewhat more natural way. Please note that such complex inter-relationships typically occur in commercial applications.

Semantic modeling is one of the tools for representing knowledge especially in Artificial Intelligence and object-oriented applications. Thus, it may be a good idea to model some of the knowledge databases using semantic database system.

Some of the features of semantic modeling and semantic databases are:

- these models represent information using high-level modeling abstractions,
- these models reduce the semantic overloading of data type constructors,
- semantic models represent objects explicitly along with their attributes,
- semantic models are very strong in representing relationships among objects, and
- they can also be modeled to represent IS A relationships, derived schema and also complex objects.

Some of the applications that may be supported by such database systems in addition to knowledge databases may be applications such as bio-informatics, that require support for complex relationships, rich constraints, and large-scale data handling.

1.6 INFORMATION VISUALISATION

Relational database offers one of the simplest forms of information visualisation in the form of the tables. However, with the complex database technologies and complex database inter-relationship structures, it is important that the information is presented to the user in a simple and understandable form. Information visualisation is the branch of Computer Graphics that deals with the presentation of digital images and interactions to the users in the form that s/he can handle with ease. Information visualisation may result in presentation of information using trees or graph or similar data structures.

Another similar term used in the context of visualisation is knowledge visualisation the main objective of which is to improve transfer of knowledge using visual formats that include images, mind maps, animations etc.

Please note the distinction here. Information visualisation mainly focuses on the tools that are supported by the computer in order to explore and present large amount of data in formats that may be easily understood.

You can refer to more details on this topic in the fifth semester course.

☞ **Check Your Progress 3**

1) What is a good knowledge base?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

2) What are the features of deductive databases?

……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………
……………………………………………………………………………………

3) State whether the following are True or False:

(a) Semantic model is the same as an object. ☐

(b) IS A relationship cannot be represented in a semantic model. ☐

(c) Information visualisation is used in GIS. ☐

(d) Knowledge visualisation is the same as information visualisation. ☐

1.7 SUMMARY

This unit provides an introduction to some of the later developments in the area of database management systems. Multimedia databases are used to store and deal with multimedia information in a cohesive fashion. Multimedia databases are very large in size and also require support of algorithms for searches based on various media components. Spatial database primarily deals with multi-dimensional data. GIS is a spatial database that can be used for many cartographic applications such as irrigation system planning, vehicle monitoring system etc. This database system may represent information in a multi-dimensional way.

Genome database is another very large database system that is used for the purpose of genomics, gene expression and proteomics. Knowledge database store information either as a set of facts and rules or as semantic models. These databases can be utilised in order to deduce more information from the stored rules using an inference engine. Information visualisation is an important area that may be linked to databases from the point of visual presentation of information for better user interactions.

1.8 SOLUTIONS/ANSWERS

Check Your Progress 1

1) a) Advanced technology in terms of devices that were digital in nature and support capture and display equipment.
   b) High speed data communication network and software support for multimedia data transfer.
   c) Newer application requiring multimedia support.

2) Medical media databases
   Bio-informatics
   Home media
   News etc.

3) Content can be of two basic types:
   a) Media Content
   b) Meta data, which includes media, format data, media keyword data and media feature data.

4) Some of the challenges are:
   a) Support for different types of input/output
   b) Handling many compressions algorithms and formats
   c) Differences in OS and hardware
   d) Integrating to different database models
   e) Support for queries for a variety of media types
   f) Handling different kinds of indices
   g) Data distribution over the world etc.

Check Your Progress 2

1) GIS is a spatial database application where the spatial and non-spatial data is represented along with the map. Some of the applications of GIS are:
   - Cartographic applications
   - 3-D Digital modeling applications like land elevation records
   - Geographic object applications like traffic control system.

2) A GIS has the following requirements:
   - Data representation through vector and raster
   - Support for analysis of data
   - Representation of information in an integrated fashion
   - Capture of information
   - Visualisation of information
   - Operations on information

3) The data may need to be organised for the following three levels:
   - **Geonomics:** Where four different types of data are represented. The physical data may be represented using eight different fields.
   - **Gene expression:** Where data is represented in fourteen different fields
   - **Proteomics:** Where data is used for five research problems.

Check Your Progress 3

1) A good knowledge database will have good information, good classification and structure and an excellent search engine.

2) They represent information using facts and rules
   New facts and rules can be deduced
   Used in expert system type of applications.

3) a) False   b) False   c) True   d) False.