

UNIT V PREDICTIVE MODELING OF BIG DATA AND TRENDS IN DATA MINING

9

Statistics and Data Analysis – EDA – Small and Big Data – Logistic Regression Model – Ordinary Regression Model – Mining complex data objects – Spatial databases – Temporal databases – Multimedia databases – Time series and sequence data – Text mining – Web mining – Applications in Data mining

Multimedia databases

- **Multimedia database** is the collection of interrelated multimedia data that includes text, graphics (sketches, drawings), images, animations, video, audio etc and have vast amounts of multisource multimedia data. The framework that manages different types of multimedia data which can be stored, delivered and utilized in different ways is known as multimedia database management system. There are three classes of the multimedia database which includes static media, dynamic media and dimensional media.

- Content of Multimedia Database management system :

1. **Media data** – The actual data representing an object.
2. **Media format data** – Information such as sampling rate, resolution, encoding scheme etc. about the format of the media data after it goes through the acquisition, processing and encoding phase.
3. **Media keyword data** – Keywords description relating to the generation of data. It is also known as content descriptive data. Example: date, time and place of recording.
4. **Media feature data** – Content dependent data such as the distribution of colors, kinds of texture and different shapes present in data.

- Types of multimedia applications based on data management characteristic are :

1. **Repository applications** – A Large amount of multimedia data as well as meta-data(Media format data, Media keyword data, Media feature data) that is stored for retrieval purpose, e.g., Repository of satellite images, engineering drawings, radiology scanned pictures.
2. **Presentation applications** – They involve delivery of multimedia data subject to temporal constraint. Optimal viewing or listening requires DBMS to deliver data at certain rate offering the quality of service above a certain threshold. Here data is processed as it is delivered. Example: Annotating of video and audio data, real-time editing analysis.
3. **Collaborative work using multimedia information** – It involves executing a complex task by merging drawings, changing notifications. Example: Intelligent healthcare network.

- There are still many challenges to multimedia databases, some of which are :
 1. **Modelling** – Working in this area can improve database versus information retrieval techniques thus, documents constitute a specialized area and deserve special consideration.
 2. **Design** – The conceptual, logical and physical design of multimedia databases has not yet been addressed fully as performance and tuning issues at each level are far more complex as they consist of a variety of formats like JPEG, GIF, PNG, MPEG which is not easy to convert from one form to another.
 3. **Storage** – Storage of multimedia database on any standard disk presents the problem of representation, compression, mapping to device hierarchies, archiving and buffering during input-output operation. In DBMS, a "BLOB"(Binary Large Object) facility allows untyped bitmaps to be stored and retrieved.
 4. **Performance** – For an application involving video playback or audio-video synchronization, physical limitations dominate. The use of parallel processing may alleviate some problems but such techniques are not yet fully developed. Apart from this multimedia database consume a lot of processing time as well as bandwidth.
 5. **Queries and retrieval** –For multimedia data like images, video, audio accessing data through query opens up many issues like efficient query formulation, query execution and optimization which need to be worked upon.

Areas where multimedia database is applied are :

- **Documents and record management** : Industries and businesses that keep detailed records and variety of documents. Example: Insurance claim record.
- **Knowledge dissemination** : Multimedia database is a very effective tool for knowledge dissemination in terms of providing several resources. Example: Electronic books.
- **Education and training** : Computer-aided learning materials can be designed using multimedia sources which are nowadays very popular sources of learning. Example: Digital libraries.
- Marketing, advertising, retailing, entertainment and travel. Example: a virtual tour of cities.
- **Real-time control and monitoring** : Coupled with active database technology, multimedia presentation of information can be very effective means for monitoring and controlling complex tasks Example: Manufacturing operation control.