Page No.
Date

Name : Dhiraj Bodake

Roll No : 18141216

TUTORIAL NO : 10

Q.1 Explain search algo and GEMINI algo in brief with respect in Genetic multimedia indexing approach.

→ Algorithm Search -

1) map the query object Ø into a point F(Q) in feacture space.

2) Using a spatial access method, retrieves all point within the desired tolerance.

3) Retrieve the corresponding object, compute their actual list. from Q and discard false alaram. Mathematically let O_1 and O_2 be two objects with distance function D() and $F(O_1)$ and $F(O_2)$ be feacture vectors then we have.

To gurantee no false dismissals for whole match queries the feature extraction function F() shared satisfy the formula

$$D_{feacture}(F(O_1), F(O_2) \leq D(O_1, O_2))$$

Date : / /
Page No.:

Explain one dimensional time series in detail.

→ Here the goal is to search a collection of time series, to find ones that are similar to desirable series. For eg. in a collection of yearly stock price movements find ones that are similar to IBM:

① Distance function:

According to GEMINI, the first step is to determine the distance measure between the time series.

② Feature extraction and lower-bounding –

Having decided on the Euclidean distance and dissimilarity measure, the next step is to find some features that can lower-bound it. The second requirement suggests that we use good features, that have much dissimilarity power. Applying first step & GEMINI algo we ask feature extracting question

2) Experiments:

Performance results with GEMINI approach on time series are reported. There the method is compared to a sequential scanning method.

Q.3 Explain in details : different challenges in searching the web

→ problems related to data in searching web:

① Distributed data : Due to intrinsic nature of the web, data spread over many computers and platforms. These computers are interconnected with no predefined topology.

② High percentage of volatile data : Due to internet dynamics, new computers and data can be added or removed easily.

③ Large volume : the exponential growth of web posed scaling issues that are difficult to cope with.

④ quality of data : The web can be considered as new publishing medium. Data can be false, invalid, written poorly or typically with many errors from diff sources.

⑤ Heterogeneous data : In addition to having dea with multiple data types and hence the with multiple formats we also have diff language

- problems faced by user during interaction with system

1) How to specify a query
2) How to interprete an answer provided by the system