DIGITAL ASSIGNMENT-1

ITE2001

DATA MINING TECHNIQUES

19BIT0292-Bhaumik Tandan

Slot: - D1+TD1

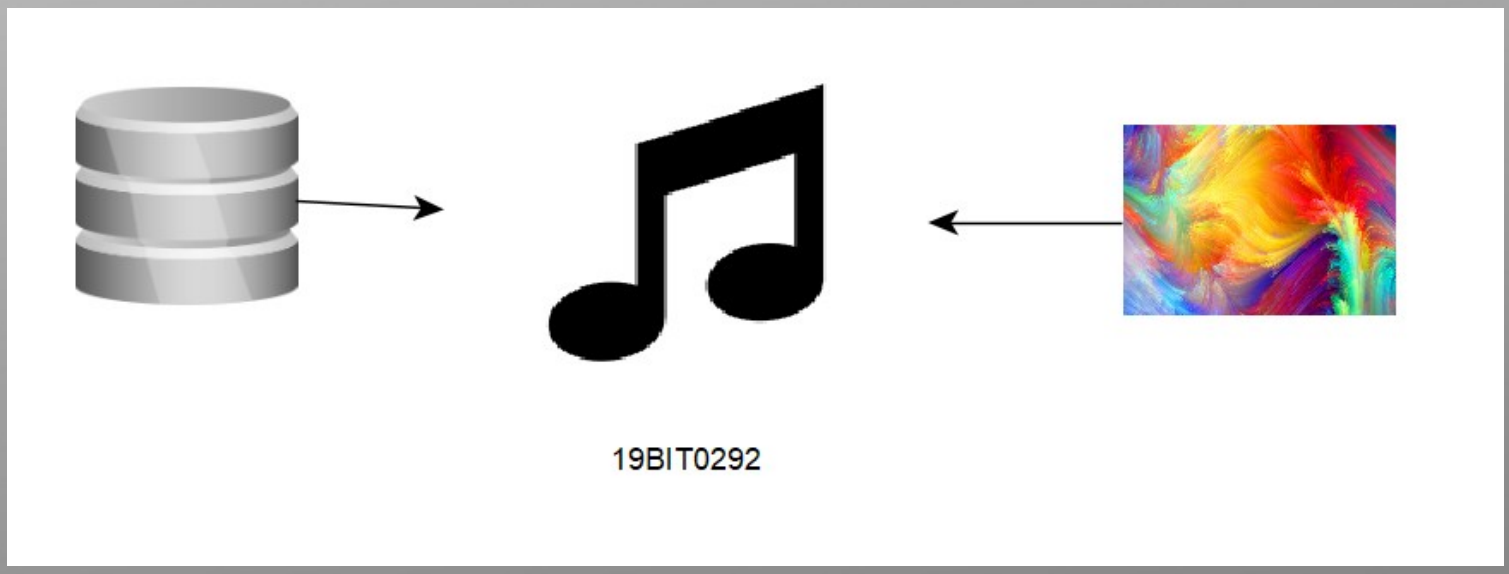
Q1. Propose a few implementation methods for audio data mining. Can we integrate audio and visual data mining to bring fun and power to data mining? Is it possible to develop some video data mining methods? State some scenarios and your solutions to make such integrated audio visual mining effective.

19BIT0292

Data mining has made significant progress and covered a broad spectrum of applications since the 1980s. Today, data mining is used in a vast array of areas. Numerous commercial data mining systems and services are available.

Audio data mining uses audio signals to indicate the patterns of data or the features of data mining results.
We can use the following approaches for audio data mining:-

1) Transform data to sound or music. When we receive the patterns, trends, structures and irregularity relationships from data mining, we need to transform these data features into different pitches, rhythms, tunes, and melody in order to identify anything interesting or unusual.



2) Establish mapping relations between color and music using a universal music and color understanding and mapping method .

Although visual data mining may disclose interesting patterns using graphical displays, it requires users to concentrate on watching patterns and identifying interesting or novel features within them. This can sometimes be quite tiresome.

It is also possible to develop video data mining or other methods that integrate sound and image. The major task involved is the integration of audio and visual data mining in order to bring power to data

19BIT0292

This requires combining visualization tools with the transformation of data patterns to sound, images, and videos, which also involves further development of methods for storing, accessing and demonstrating data efficiently in a multidimensional way. This is still a large and unexplored research domain.

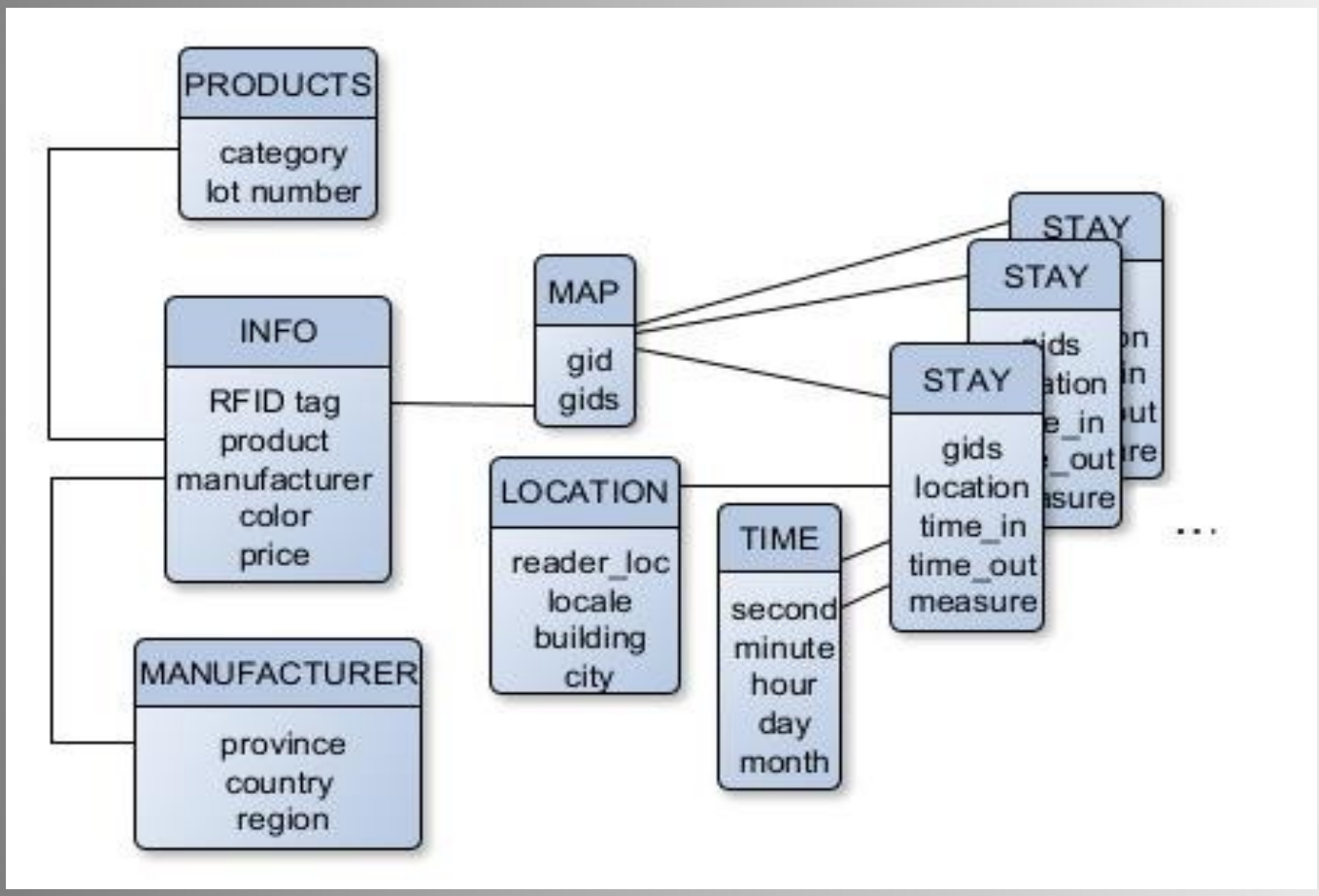
If patterns can be transformed into sound and music, then instead of watching pictures, we can listen to pitch, rhythm, tune, and melody to identify anything interesting or unusual. This may relieve some burden of visual concentration and be more relaxing than visual mining. Therefore, audio data mining is an interesting complement to visual

Q2. Radio-frequency identification is commonly used to trace object movement and perform inventory control. An RFID reader can successfully read an RFID tag from a limited distance at any scheduled time. Suppose a company wants to design a data warehouse to facilitate the analysis of objects with RFID tags in an online analytical processing manner. The company registers huge amounts of RFID data in the format of (RFID, at location, time), and also has some information about the objects carrying the RFID tag, for example, (RFID, product name, product category, producer, date produced, price).

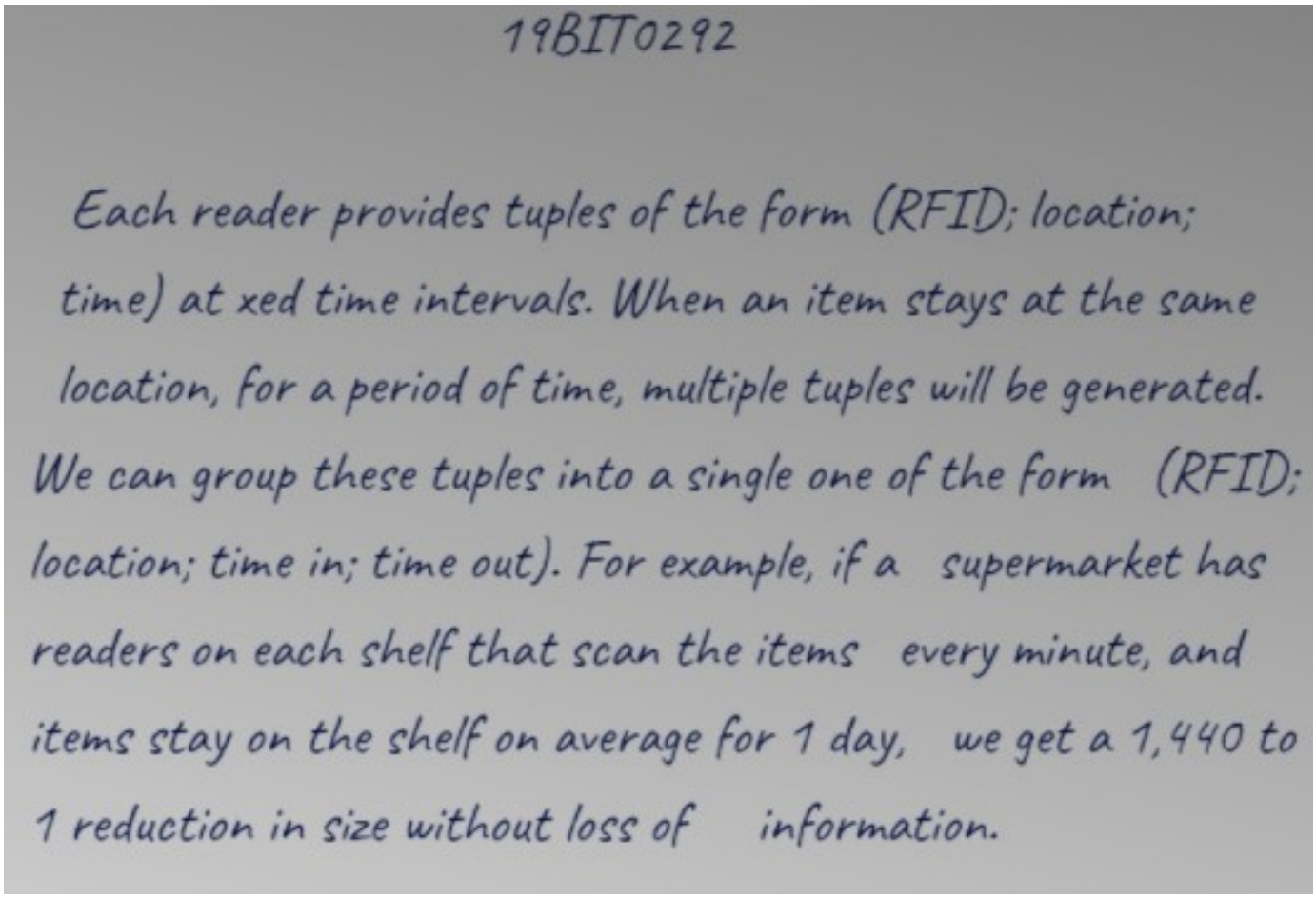
(a) Design a data warehouse to facilitate effective registration and online analytical processing of such data.

19BIT0292

A RFID warehouse need to contains a fact table, stay, composed of cleansed RFID records; an information table, info, that stores path-independent information for each item; and a map table that links together different records in the fact table that form a path. The main difference between the RFID warehouse and a traditional warehouse is the presence of the map table linking records from the fact table (stay) in order to preserve the original structure of the data.



(b) The RFID data may contain lots of redundant information. Discuss a method that maximally reduces redundancy during data registration in the RFID data warehouse.



(c) The RFID data may contain lots of noise such as missing registration and misread IDs. Discuss a method that effectively cleans up the noisy data in the RFID data warehouse.

19BIT0292

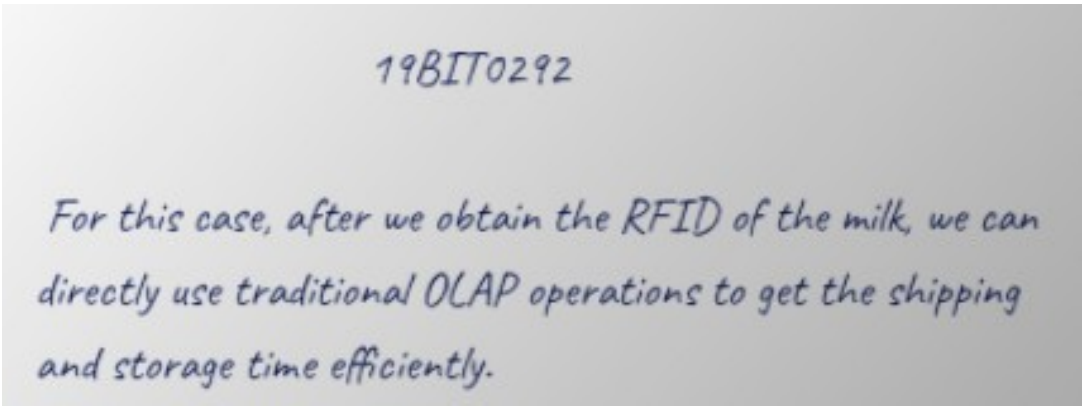
One can use the assumption that many RFID objects stay or move together, especially at the early stage of distribution, or use the historically most likely path for a given item, to infer or interpolate the miss and error reading.

(d) You may want to perform online analytical processing to determine how many TV sets were shipped from the LA seaport to BestBuy in Champaign, IL, by month, brand, and price range. Outline how this could be done efficiently if you were to store such RFID data in the warehouse.

19BIT0292

Compute an aggregate measure on the tags that travel through a set of locations and that match a selection criteria on path independent dimensions

(e) If a customer returns a jug of milk and complains that is has spoiled before its expiration date, discuss how you can investigate such a case in the warehouse to find out what the problem is, either in shipping or in storage.



CLICK HERE FOR IMAGES