Data Structure and Algorithm – Assignment -1

What I understood from the thesis:

In the introduction, we are performing every single task through online. The data are stored on the servers. For each transaction, there is a transaction id and set of item for the transaction dataset. ). Each transaction has its unique transaction id and set of item accessed by the user during his/her session. Data mining techniques are explained in the introduction part. Frequent pattern mining plays an important role in many data mining techniques. as association rules, correlations, sequences, classifiers, clusters, etc., in which, mining association rule is one of the most popular problems. The association rule mining is one of the researched techniques of data mining introduced in 1993 by R. Agrawal et al. Association rule generation is an easy task, frequent pattern generation is a challenging task. The frequent pattern mining task is computationally expensive when a large number of frequent patterns exist.

Frequent pattern mining is used to generate association rules. It finds the association betv\'een two or more items.

DATASET LAYOUT

The transaction dataset contains the Transaction IDs (TIDs) and the items (pages) accessed during the transaction. The dataset can be ordered in two ways, 1) horizontally 2) vertically. In horizontal data layout, the dataset contains two columns: TID and the list of items. Each row in the dataset contains TID of a transaction and a set of items (pages) accessed by the user during his/her session.

Some of the web items we are using for communication:

1. Internet
2. Web pages
3. Website
4. Web browser

Data mining:

Data Mining becomes the research area with increasing importance. It is used to search hidden information from databases. It is the process of extracting useful information from the larger dataset and convert it into an understandable format for future use. It is a process with applied intelligent methods to the database and extract the useful information, also known as pattern from the database. It also refers to knowledge discovery in database because it extracts previously unknown and potentially useful information like knowledge rules, constraints, and regularities from the data in databases. Knowledge discovery consists of the following steps:

1) Data cleaning

2) Data integration

3) Data selection

4) Data transformation

5) Data mining

6) Pattern evaluation

7) Knowledge presentation

FREQUENT PATTERN

Frequent patterns are the patterns that occur frequently. The kind of frequent patterns are frequent itemsets,- frequent subsequences (sequential patterns), and frequent substructures.

This frequent patterns are useful in association clustering. Frequent patterns are also useful in correlation mining.

WEB MINING

Web mining as the name suggest mines the meaningful information from tlie web data lilce web documents, usage log of - web sites and hyperlink between web pages, etc.

TYPES OF WEB MINING

1. WEB CONTENT MINING
2. WEB STRUCTURE MINING
3. WEB USAGE MINING

DATA PREPROCESSING TASKS:

1. DATA CLEANING (REMOVING UNWANTED INFORMATION)
2. USER IDENTIFICATION (IDENTIFYING THE USER BY VARIOUS TECHNIQUES)
3. SESSION IDENTIFICATION (FIND USER'S ACTIVITY USING SESSION)
4. PATH COMPLETION
5. TRANSACTION IDENTIFICATION
6. FORMATTING
7. PATTERN DISCOVERY

This section represents the different types of mining activities that can be applied to the web domain.

1) Statistical Analysis

2) Rules Discovering

3) Clustering

4) Classification

5) Sequential Patterns Finding

6) Dependency Modeling.

FREQUENT PATTERN MINING TERMINLOGIES

1. ITEM
2. PATTERN
3. ITEMSET
4. TRANSACTION
5. TRANSACTION DATASET