

Lesson 7

Web Usage Mining

Web Mining Components

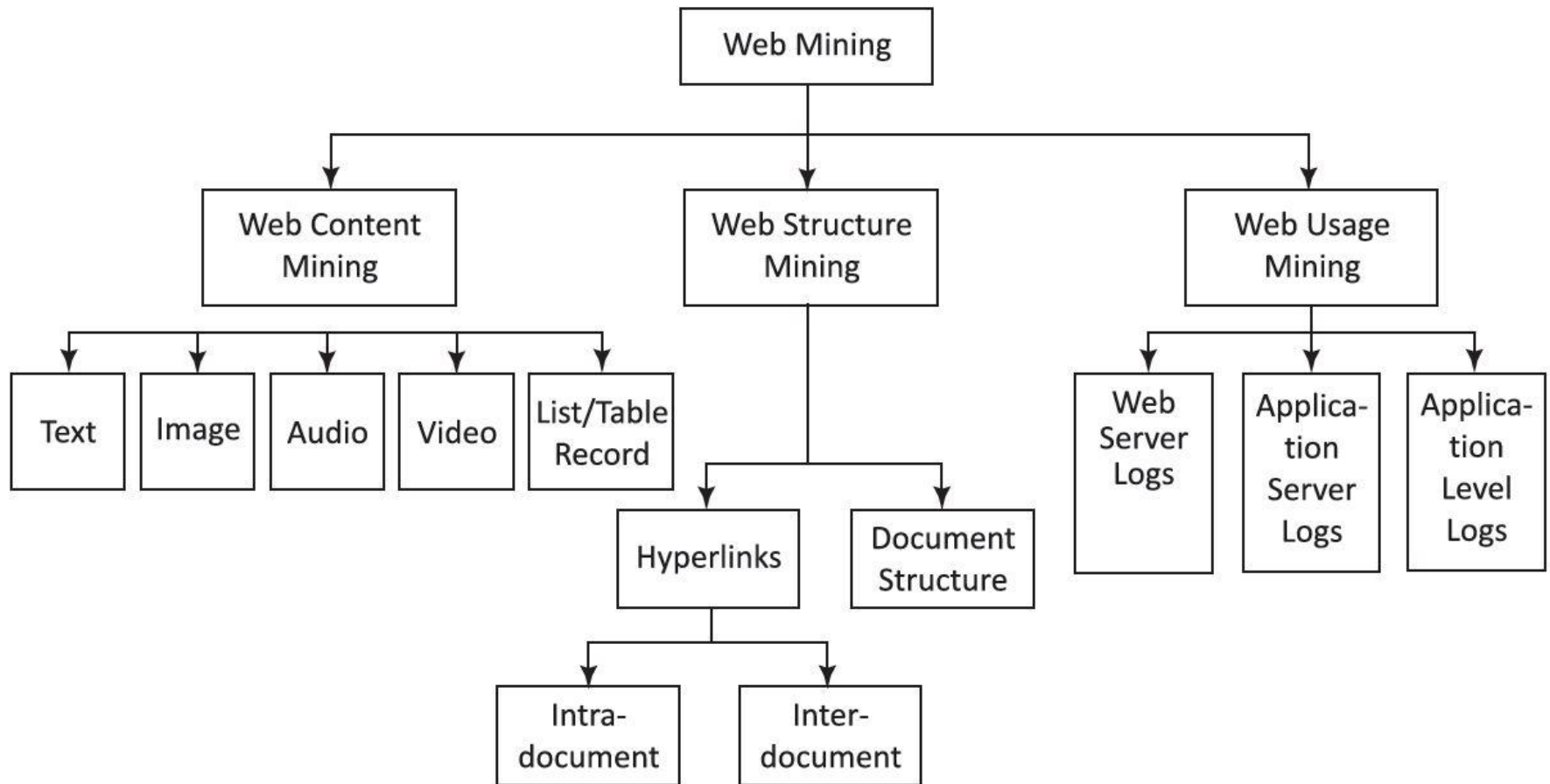
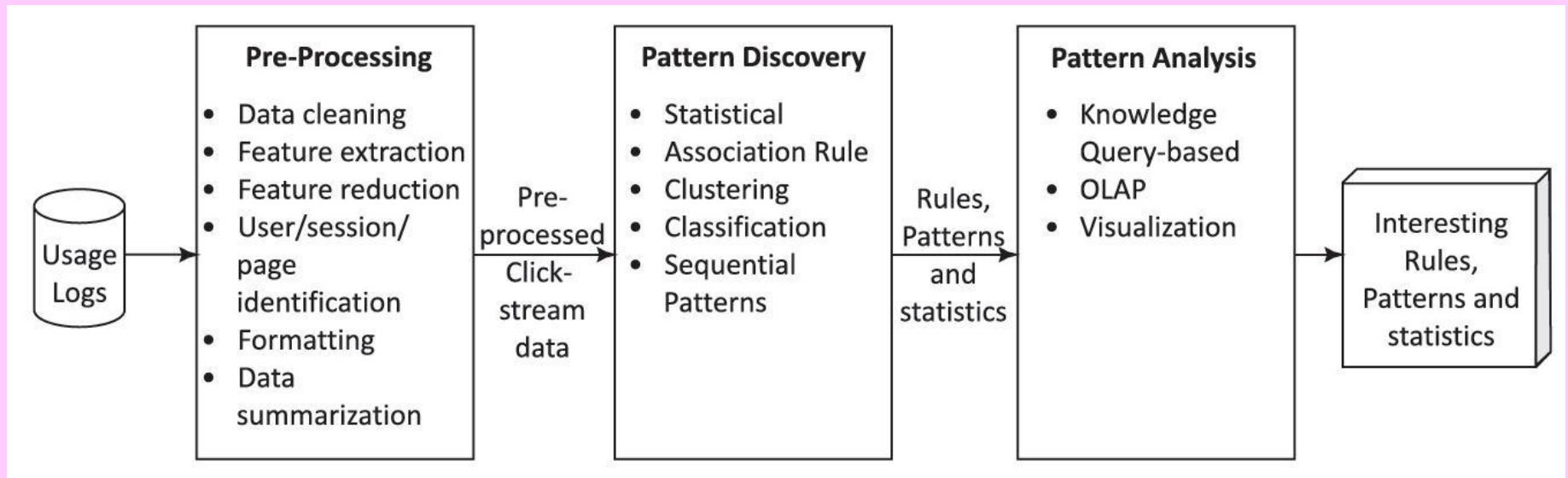


Figure 9.7 Process of web usage mining



Pre-processing

- using vector space model (Refer Example 9.2).
- Pre-processing is the data preparation task, which is required to identify
-

Pre-processing

- (iii) Content from server logs to obtain state variables for each active session
- (iv) Page references

Pre-processing

- (i) User through cookies, logins or URL information
- (ii) Session of a single user using all the web pages of an application
- (iii) Content from server logs to obtain state variables for each active session
- (iv) Page references.

Pattern Discovery

- Knowledge extraction algorithms based on statistics, ML and data mining algorithms using pre-processed data
Path analysis,
- Association rules,
- Sequential patterns
- clustering and classification of web usages.

Statistical Techniques

- (i) The most frequently accessed pages
- (ii) Average view time of a page or average length of a path through a site
- (iii) Providing support for marketing decisions

Association Rules Mining

- Enable relating the pages, which are most often referenced together in a single server session
- These pages may not be directly connected to one another using the hyperlinks

Association Rules

- Reveal a correlation between users who visited a page containing similar information
- Provide recommendations to purchase other products
- Provide help to web designers to restructure their websites

Association Rules

- Retrieve the documents in prior in order to reduce the access time when loading a page from a remote site.

Clustering

- (i) Developing a profile of users belonging to a particular class or category
- (ii) Discovery of interesting rules from server logs. For example, 3750 users watched a certain movie, out of which 2000 are between age 18 to 23 and 1500 out of these lives in metro cities

Clustering

- (iii) Provide help to web designers to restructure their websites
- (iv) Retrieve the documents in prior in order to reduce the access time when loading a page from a remote site

Sequential pattern discovery

- User navigation patterns in web usage data gather web page trails
- Markov Model can be used to model navigational activities in the website.
- Computation of a number of significant user or site metrics that can lead to useful rules, pattern, or statistics.

Pattern Analysis

- (i) A knowledge query mechanism such as SQL
- (ii) Another method is to load usage data into a data cube in order to perform Online Analytical Processing (OLAP) operations
- (iii) Visualization techniques, highlight overall patterns or trends in the data

Pattern Analysis

(iv) Content and structure information on filtering out patterns containing pages of a certain usage type, content type or pages that match a certain hyperlink structure

Summary

We learnt:

- Web Usage Mining
- Preprocessing of Usage data
- Pattern Discovery from Usage Data
- Association Rules mining
- Clustering
- Classifying

End of Lesson 7 on **Web Usage Mining**