

# Lesson 5

## Web, Web-Contents, Web- Structure and Web-Usages Mining

# Data Mining

- Is a process of discovering patterns in large datasets to gain knowledge
- The process can be shown as [Raw Data → Patterns → Knowledge].

# Web Mining

- Refers to the use of techniques and algorithms that extract knowledge from the web data available in the form of web documents and services.

# Web Mining Applications

- (i) Extracting the fragment from a web document that represents the full web document

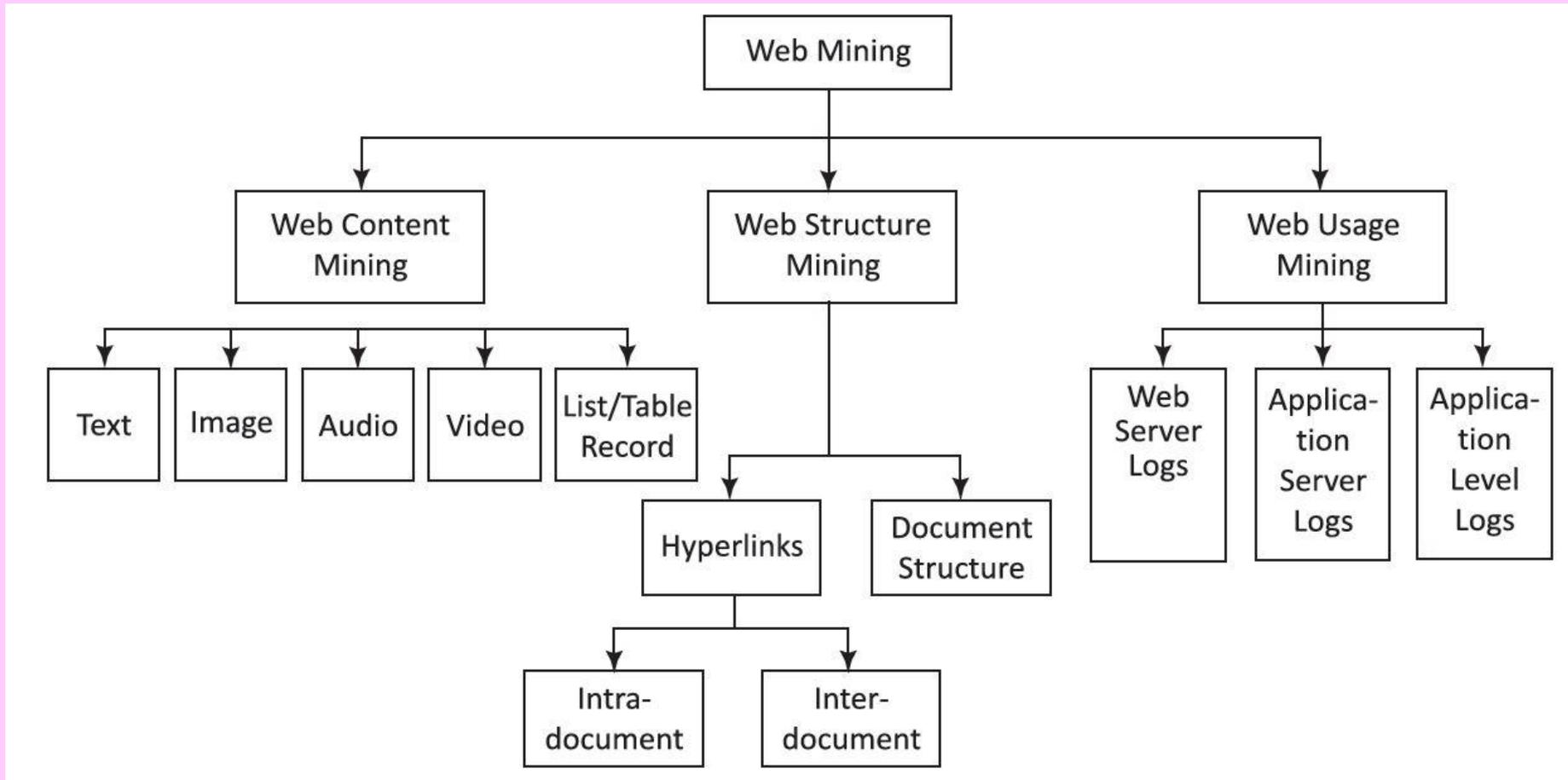
# Web Mining Applications

(ii) Identifying interesting graph patterns or pre-processing the whole web graph to come up with metrics, such as PageRank

# Web Mining Applications

(iii) User identification, session creation, malicious activity detection and filtering, and extracting usage path patterns

# Figure 9.6 Web Mining Taxonomy



# Web Content Mining

- Is the process of extracting useful information from the contents of web documents
- Content may consist of text, images, audio, video or structured records, such as lists and tables.

# Web Structure Mining

- Is the process of discovering structure information from the web. Based on the kind of structure-information present in the web resources, web structure mining can be divided into:
  1. Hyperlinks:
  2. Document Structure: .

# Web Usage Mining

- Is the application of data mining techniques which discover interesting usage patterns from web usage data
- The data contains the identity or origin of web users along with their browsing behavior at a web site

# Web Usage Mining Applications

- (i) Web Server logs: Collected by the web server and typically include IP address, page reference and access time

# Web Usage Mining Applications

(ii) Application Server Logs: Application servers typically maintain their own logging and these logs can be helpful in troubleshooting problems with services

# Web Usage Mining Applications

(iii) Application Level Logs: Recording events usually by application software in a certain scope in order to provide an audit trail that can be used to understand the activity of the system and to diagnose problems.

# Summary

We learnt:

- Web Mining Taxonomy
- Web Content Mining
- Web Structure Mining
- Web Usage Mining
- Applications of Web-Contents, Web-Structure and Web-Usage Mining

End of Lesson 5 on

**Web, Web-Contents, Web-  
Structure and Web-Usages Mining**