

Introduction to Web Mining

Web mining is an application of data mining techniques to find information patterns from the web data.

Web mining helps to improve the power of web search engine by identifying the web pages and classifying the web documents.

Web mining is very useful to e-commerce websites and e-services.

There are three types of web mining:

1. Web Content Mining

- Web content mining can be used for mining of useful data, information and knowledge from web page content.
- Web structure mining helps to find useful knowledge or information pattern from the structure of hyperlinks.
- Due to heterogeneity and absence of structure in web data, automated discovery of new knowledge pattern can be challenging to some extent.
- Web content mining performs scanning and mining of the text, images and groups of web pages according to the content of the input (query), by displaying the list in search engines. **For example:** If an user wants to search for a particular book, then search engine provides the list of suggestions.

2. Web Usage Mining

- Web usage mining is used for mining the web log records (access information of web pages) and helps to discover the user access patterns of web pages.
- Web server registers a web log entry for every web page.
- Analysis of similarities in web log records can be useful to identify the potential customers for e-commerce companies.

Some of the techniques to discover and analyze the web usage pattern are:

i) Session and visitor analysis

- The analysis of preprocessed data can be performed in session analysis, which includes the record of visitors, days, sessions etc. This information can be used to analyze the behavior of visitors.
- Report is generated after this analysis, which contains the details of frequently visited web pages, common entry and exit.

ii) OLAP (Online Analytical Processing)

- OLAP performs Multidimensional analysis of complex data.
- OLAP can be performed on different parts of log related data in a certain interval of time.
- The OLAP tool can be used to derive the important business intelligence metrics.

3. Web Structure Mining

- The web structure mining can be used to discover the link structure of hyperlink.
- It is used to identify that the web pages are either linked by information or direct link connection.
- The purpose of structure mining is to produce the structural summary of website and similar web pages.

Example: Web structure mining can be very useful to companies to determine the connection between two commercial websites.