

# The Integrated Process of Web Mining to Analyze Big Data to Optimize the Challenges of Web

Ms. Ragita Ranjit

Yuvakshetra, Mundur, Palakkad, Kerala, India

---

**Abstract:** This article gives the details about the analysis techniques used by Web Mining to get the details according to the user's requirement from the web. The article explains about how data is retrieved and the knowledge required by the user is retrieved from the web using Web Mining Techniques. The data is required to be retrieved from the big data from the web. The article explains about the techniques, the software and the analysis method used by Web Mining to optimize the biggest challenge of the web by retrieving data from big data.

**Keywords:** Big Data, Web Mining, Techniques, Web Mining Software, Content Mining, Usage Mining, Structure Mining.

---

## I. INTRODUCTION

Web Mining is the technique of extracting the useful data from the information gathered over the World Wide Web and the information gathered using the traditional data mining methodologies. The difficult techniques of web mining are content mining, structure mining and usage mining. Clustering and classification, association and examination of sequential pattern are the traditional data mining parameters used for evaluating the information gathered through Web Mining. The application that uses data mining to discover and analyze interesting pattern of user's usage data on the web is Web Usage Mining.

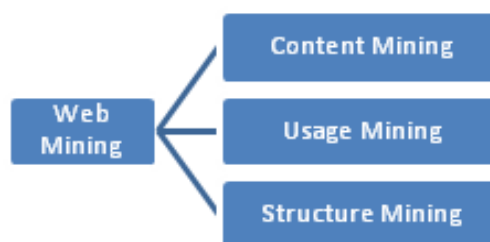


Fig: 1. Web Mining

## II. WEB MINING TECHNIQUES

### A. Content Mining:

An automatic process that also involves keyword extraction is called Web content mining. Techniques using lexicons for content interpretation are yet to come. There are two content mining strategies are:

- 1) That directly mine the content of the documents.
- 2) That improve the content search of the search tools like search engines.

**B. Structure Mining:**

Structure Mining deals with the how the content is organized on a site in the form of

- 1) Links between pages
- 2) Pattern of information on pages
- 3) Pattern of pages on the site
- 4) Link to the page outside of the site

**C. Usage Mining:**

This describes how the people use the site and the different ways how the web servers deal with the user's information. The web sites can understand the user behaviour and the web structure by analyzing the web access logs and thus improving the colossal collection of resources. General Access Pattern tracking and Customized Usage Tracking are the main techniques used in Web Usage Mining.

### III. BIG DATA

**A. Big Data and process techniques:**

Big Data indicates massive store and volume of structured and unstructured data which is so large that it is difficult to process it using traditional software and database techniques.

The most suitable technologies are:

- 1) A/B testing
- 2) Crowd sourcing,
- 3) Data fusion and integration
- 4) Genetic algorithms,
- 5) Machine learning
- 6) Natural language processing
- 7) Signal processing
- 8) Simulation
- 9) Time series analysis and Visualizations

**B. Content Mining for Big Data:**

The techniques used for content mining are:

- 1) Structured Data Extraction
- 2) Text Mining

**C. Structure Mining for Big Data:**

The technique used is Web graph which consist of Web pages as nodes and hyperlinks as edges connecting between two related pages

- 1) Web Page Rating
- 2) Web Clustering /Classification

**D. Usage Mining for Big Data:**

Usage Mining for Big Data consists of following techniques:

- 1) Navigational Pattern
- 2) Session And Visitor Analysis

#### IV. WEB MINING SOFTWARE

**TABLE I: Web Mining Software Categories**

No	Web Mining Techniques					
1	Content Mining	Screen Scraper	Automation Anywhere 6.1	Web Info Extractor	Mozenda	Web Content Extractor
2	Structure Mining	Web Data Extraction	Import.io	Cloudscrape	Webhose.io	Winautomation
3	Usage Mining	Springer	Speed Tracer	Web Log Explorer	Alter Wind Log Analyzer Professional	Amadea Web Mining

#### V. BIG DATA ANALYSIS BY WEB MINING SOFTWARE

**A. Analysis of Big Data by Content Mining:**



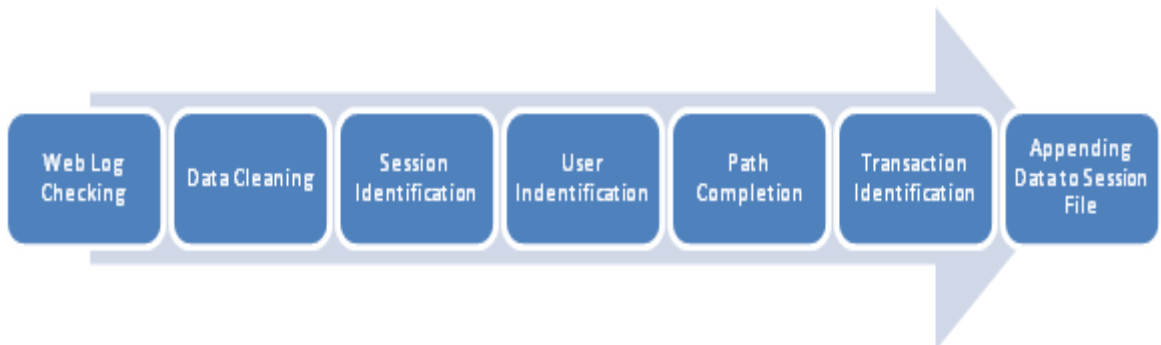
**Fig.2. Content Mining Analysis**

**B. Analysis of Big Data by Structure Mining:**



**Fig.3. Structure Mining Analysis**

**C. Analysis of Big Data by Usage Mining:**



**Fig.4. Usage Mining Analysis**

## VI. CONCLUSION

With the combination of various algorithms, new techniques Web Mining are optimising the challenges of Big Data. Web Data Mining is a process designed to explore big data by analysis. Web mining techniques searches for patterns and/or relationships between variables, and aims to validate data as per as the requirement. The ultimate goal of data mining is to help business applications using the prediction - and predictive data mining techniques. The process of data mining consists of three stages: (1) the initial analysis, (2) validation or verification, and (3) deployment

## REFERENCES

- [1] Liu, Bing (2007); Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, Springer, ISBN 3-540-37881-2.
- [2] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieved 17 December 2008.
- [3] Clifton, Christopher (2010). "Encyclopædia Britannica: Definition of Data Mining". Retrieved 2010-12-09.
- [4] Witten, Ian H.; Frank, Eibe; Hall, Mark A. (30 January 2011). Data Mining: Practical Machine Learning Tools and Techniques (3 ed.). Elsevier. ISBN 978-0-12-374856-0.
- [5] [http://en.wikipedia.org/wiki/Data\\_mining](http://en.wikipedia.org/wiki/Data_mining).