Fuzzy Systems and Soft Computing
ISSN : 1819-4362

WEB MINING TECHNIQUES AND APPLICATION

Prof Smruti Ranjan Dash. Department of CSE, Synergy Institute of Engineering and Technology, Dhenkanal, Odisha, India.

Abstract.
The purpose of this paper is to provide a more current evaluation and update of web mining research and techniques available. In the today's view of web mining is to extract useful information from various websites and organizations websites. This paper will elaborate various tools and techniques of web mining that can be used now a day for web mining. The Data mining techniques can be used by system to extract the required knowledge. Now a days web mining is useful in Social media applications, Electronic commerce websites, various web transaction, etc. and it's not only extracting data but also discover the knowledge about any topic required by web user. The proposed paper is focusing on various tools, techniques and applications of web mining.

Keywords.
Web similarity index, Content data, structured data.

1. Introduction

Web mining is a world famous technique of data mining that can extract knowledge and patterns as per the needs of user [1-3]. In this system the proposed work describe a web mining is an application of data mining and it uses various techniques to discover data. A web mining can used website or documents as a resource to extraction of data.

Since 1995, many publishers are focused on web mining applications. In the previous tool and techniques, it's authorized that website mining is a tactic of various other fields like information search & recover, storage, (AI)Artificial Intelligence that are recognized and adapt mining tools and techniques [3-5].

The large amount of data can be widely maintained and preserve in a Data-Warehouse, it considering and including all the motive and purpose of creation is to gathered selected data to identify the structure of material, directions and navigations [6]. we can differentiate the data used in four important types:

- Context Data: This type of data can carry(textual data, image dataand graphics data).
- Structure Data: Page Structure, Inner Structure of pages.
- Use of Data: It can describe to use of data by using sources of internet.
- User profile related data.

In the proposed work we do the previous survey on web mining technologies, then its tactics: recent trend, its tools, and an analysis of many areas and application.

Then, we will make proposal of an approach for the analysis of web data and discover the latest trends which can help user to discover appropriate knowledge from web.

2. The Types of Web Mining

The data can be mine using the four types of data mining such as follows:

Web mining- Content Mining

To extraction of knowledge and related information from the defined material of documents and website pages one mechanism is used called ad web content mining such as format capturing and preserving, picture data, textual data, audio data, videos data, etc [7-8].

Web mining- Structure Mining

Web Structure Mining is an overview of overall design of web i.e. the flow, design and the hyperlinks that available between the different web pages and websites [9]. The identification of the way of transportation allows, for example, determining how many web pages to reviewed and indexed the internet subscribers on the mean and it adapt the site tree structure for the first pages of the site. Many researchers can insert the links between the pages for growing and increasing economy. We

used two famousweb structures mining algorithms, Page Rank and HITS.

Web mining- usage mining
This is a most important technique of web mining that can we elaborate the actual use of web data [10-14]. The effectively and efficiently user can perform web mining using three stages. These steps added the pre-processing of data, discover the pattern and identification of pattern.

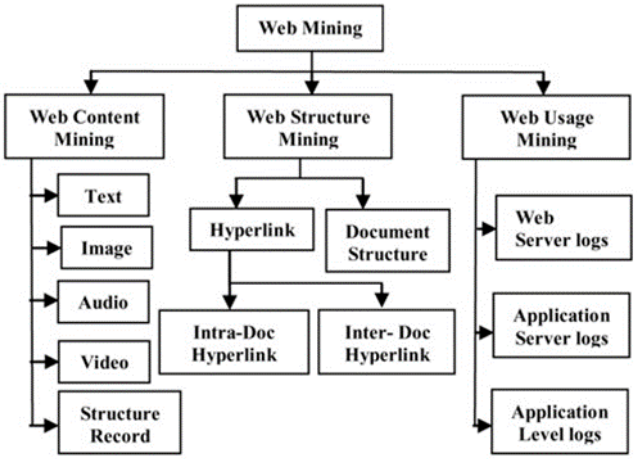

Figure 1. Website Mining Architecture

3. Webmining Mining Tactics: Preceding Trends
Now a days many of the techniques and algorithms are in used for finding,distinguishing, observing website data [9-15].
Types of classification rules and models
- Tree classification
- Naïve Bayesian Classification
- Semantic Networks
- Support Vector Machines (SVM)
- Classification and Associations

Types of clustering methods
- Cluster-Divisivetechniques
- Cluster-Random techniques
- Cluster-Data Grid view techniques
- Cluster-Data Modelbased techniques

Types of association methods
- Multilevel union rule.
- Multidimensional union rule.
- Quantityunion rule.

Data Preprocessing
- Data Cleaning
- User & Session Identification-Pattern Discovery
- Statistical Analysis
- Association Rules
- Clustering
- Classification
- Sequential Patterns - Pattern Analysis

- Knowledge Query Mechanism
- OLAP (Online Analytical processing)
- Intelligent Agents

Many of the web mining techniques gives different results and that can be discoverrequired knowledge and gives directions to the web users.

4. Website Mining Tools

The tools and techniques of website mining permits user to fetch and download datafrom web storage and whatever information is required that can gathered from appropriate place [16]. In the era of web mining various tools are used and no need to used encryption tools to deals with robot, which would require a strong knowledge of web development languages and database development languages; some of them have limited period and free for life time [17-22].

As per the rules of web mining number of techniques as invented and number of operating systems are supported to mining applications.

Latest Web Mining Tools

- Data Miner
- Google Analytics
- SimilarWeb
- Majestic
- Scrapy
- Bixo
- Oracle data Mining
- Tableau

Table 1. Toll and Techniques of Web Mining Tools

Mining Tools	Web Mining Area	Development Language	Supported O.S	Functionality
Screen-Scraper	Web Content Mining	Java Python. NET,VB, ASP, PHP	Linux, Windows	The SQL Server database can be used to discover the knowledge from web
WebContent Extractor	Web Content Mining	Python	Windows	This is related to professional businesses and used standard techniques to extract data.
Web Mozenda	Web Content Mining	JavaScript	Windows	This tools can use for storage and to transfer information to various places.
WebInfo Extracto r	Contextual Mining	--	Windows	It used various properties of mining line extraction, analysis, etc.
scrappy	Web Content Mining	Python	Unix, Windows, Apple	Web Scrapyis an open source framework for extracting data very quickly.
R-Reserve	Usage of Web	language like Python, Ruby, Perl	UNIX platform Windows AppleOS	The language R b a s i c a l l y u s e d f o r statistical analysis

Web Information Filter	Usage of Web	Basic Java, Procedure Databases	Linux, Windows	The web information system used filtration techniques to filter data.

5. Web Mining Applications

In the Last decade [23-26], the use of web search is increased which describes important and techniques of web mining such as follows

Electronic-Learning

Electronic learning facilitates the user for leaning online using web applications. This will provide quality and quantity of data to the user and now day's users are taking more advantage of this facility.

Customization of web Design and data

Today's web sites are more powerful and dynamic, this websites are providing facility to personalize and customized the settings. Usage of customization help user to retrieve only relevant data for extraction and discovery of patterns.

Duplication and Fraudulent detection

Duplicated users can easily detected by web logs data and it can be stored in logs for long period. Illegal access is easily detected using latest tools and techniques of web mining.

Website Robotics Identification

Websites robotics are application software which can act like human for searching websites. This types of software's are dangerous for system and can easily brake down the passwords and credentials.

Electronic-commerce

It is an electronic facility provided by web to buy and sale the products by using web. This will provides very good user interface with many more facilities to the web users.

Customized Web Portal

The Customized portal is extraordinary facility provided by web for look-and-feel of websites. This will reframe the website as per user demand.

Searching on Web—Google, Yahoo

Google is world famous and powerful search engine. It provides connectivity of more than 3 billion web pages that has stored on different servers. The google gives very fast response and quality of data search to the user. Old search engines are not responsible and not responding properly as compared to newest one. The link mining framework is first introduced by Google. The page ranking framework which gives ranking and assigns priorities to the web sites and it gives best quality output.

6. Experimental Result & Methodology

Year	Researcher	Algorithm/ Method	Input	Results
2019	Web Mining: Information and Pattern Discovery on the World Wide Web	Pre- Processing	Web Sites	Extraction and Retrieval of data
2018	Web Usage Mining: A Review of Recent Works	Mining association rules from WUM	Social Media Web Sites	These Technique is used for time analysis future events based on past events'

2017	Web Mining tactics and usage: review and a proposed system to improve system performance of recruitment of freshers.	Priority Algorithm	Apriori K-Apriori K-means	The result of this system is to express the efficient way of discovery.

7. Future Development Of Web Mining

Some of the future scopes of web mining are

- Forensics Identification.
- Crime investigation.
- Automated data cleaning.
- Robot detection and filtering.
- Transaction identification.
- Cloud mining.
- Temporal Evolution of the Web.
- Web Metrics and Measurements.

8. Conclusion & Future Work

In this paper, we introduced and describe some of the new web mining techniques and applications. We also focused on last techniques are used in the web mining and briefly distinguished old and new techniques. Our main target is to present the latest trends of web mining its applications, methodology, tools and techniques used for the mining. Web mining is a widely used approach now days to extract, discover, analyzed data andshow results very quickly in front of user. This proposed paper is survey of recent paper and on the basis of that we proposed new methodology and approaches to make updation of web mining framework. This technique contains various new algorithms and that will help user to quickly discover the desired output within a limited time period.

References

[1] S. Chakrabarti, Mining the Web: Discovering Knowledge from Hypertext Data (ElsevierScience & Technology Books, 2017),ISBN-13: 9781558607545.

[2] R. Chau, C. Yeh and K. Smith, Personalized multilingual web content mining, KES(2015), p. 155–163.

[3] L. Chen, W. Lian and W. Chue, Using web structure and summarization techniquesfor web content mining, Inform. Process. Management: Int. J. 41(5) (2014) 1225–1242.

[4] T. S. Clendaniel, Profitability and mining web data: Avoiding the path to red ink,The Data Administration Newsletter (R.S. Seiner, Publisher, 2015), http://www.tdan.com/i019fe02.htm.

[5] S. Cyrus and F. Banaei-Kashani, Efficient and anonymous web usage mining for webpersonalization, INFORMS J. Comput. Special Issue on Data Mining 15(2) (2014)123–147.

[6] J. Darmont, O. Boussaid and F. Bentayeb, Warehousing Web Data (2014),http://www. arxiv.org/ftp/arxiv/papers/0705/0705.1456.pdf.

[7] O. Etzioni, The World Wide Web: Quagmire or gold mine, Commun. ACM 39(11)(1996) 65–68.

[8] X. Fang and O. Sheng, LinkSelector: A web mining approach to hyperlink selectionfor web portals, ACM Trans. Internet Tech. 4(2) (2015) 209–237.

[9] K. Fenstermacher and M. Ginsburg, Client-side monitoring for web mining, J. Am.Soc. Inform. Sci. Tech. 54(7) (2010) 625–637.

[10] J. Furnkranz, Web structure mining Exploiting the graph structure of the worldwideweb .OGAI-J. 21(2) (2009) 17–26.

[11] S. Guan and P. McMullen, Organizing information on the next generation webdesign and implementation of a new bookmark structure, Int. J. Inform. Technol.Decision Making 4(1) (2010) 97– 115.

[12] J. Han and C. Chang, Data mining for web intelligence, Computer (November 2013),pp. 54–60, http://www.faculty.cs.uiuc.edu/~hanj/pdf/computer02.pdf.

[13] B. Hay, G. Wets and K. Vanhoof, Mining navigation patterns using a sequence alignmentmethod, Knowledge Inform. Syst. 6(2) (2010) 150–163.

[14] A.Joshi, Web mining (2012), www.cs.umbc.edu/~ajoshi/web mine.

[15] A. Joshi, Web/data mining and personalization, University of Maryland BaltimoreCounty (UMBC) eBiquity Research Area (2013), http://ebiquity.umbc.edu/project/html/id/17/Web-Data-Mining-and- Personalization.
[16] D. Kanellopoulos and S. Kotsiantis, Semantic web: A state of the art survey, Int. RevComput. Software 3(1) (2001) 428–442.
[17] B. Liu, Web content mining (2012), http://www.cs.uic.edu/~liub/WebContentMining.html.
[18] MegaputerIntellegenceInc.,WebAnalystarchitecture(2014),
http://www.megputer.com/products/wa/architechture.php3.
[19] Z. Pabarskaite and A. Raudys, A process of knowledge discovery from web log data:Systematization and critical review, J. Intell. Inform. Syst. 28(1) (2015) 79–104.
[20] S. Palmer, The semantic web: An introduction (2014), http://infomesh.net/2001/swintro/.
[21] D. Pierrakos, G. Paliouras, C. Papatheodorou and C. Spyropoulos, Web usage miningas a tool for personalization: A survey, User Model. User-Adapt. Interact. 13(4) (2015)311–372.
[22] F. Faccaand and P. Lanzi, Mining interesting knowledge from weblogs: A survey, DataKnowledge Eng. 53 (2012) 225–241.
[23] P. Giudici and R. Castelo, Association models for web mining, Data Mining KnowledgeDiscov. 5 (2010) 183–196.
[24] W. Grossmann, M. Hudec and R. Kurzawa, Web usage mining in e-commerce, Int.J. Electron. Business 2 (2011) 480–492.
[25] H. Han and R. Elmasri, Learning rules for conceptual structure on the web: Specialissue on web content mining, J. Intell. Inform. Syst. 22 (2011) 237–256.
[26] B. Haruechaiyasak and M. Shyu, A web-page recommender system via a data miningframework and the semantic web concept, Int. J. Comput. Appl. Tech. 27 (2010) 298-311.