Abstract
Data mining (sometimes called data or knowledge discovery) is the method of analyzing data from different perspectives and summarizing it into useful information. For retrieving and analyzing the data from web, web data mining can be used which is an important field of data mining of extraction of knowledge from the web. Web data mining can be classified into three distinct categories, web content mining, web structure mining, and web usage mining. The aim of this paper is to evaluate the past present and future of each of these three categories of web data mining. Future trends of web data mining research have also been considered in this paper.

Keywords
Web, Data Mining, Web Data Mining, Categories of Web Mining

I. Introduction
Web mining is the application of data mining techniques to extract knowledge from Web data - including Web documents, hyperlinks between documents, usage logs of web sites, etc. There are roughly three knowledge discovery domains that pertain to web mining: Web Content Mining, Web Structure Mining, and Web Usage Mining. Web content mining is the process of extracting knowledge from the content of documents or their descriptions. Web document text mining, resource discovery based on concepts indexing or agent based technology may also fall in this category. Web structure mining is the process of inferring knowledge from the World-wide Web organization and links between references and referents in the Web. Finally, web usage mining, also known as Web Log Mining, is the process of extracting interesting patterns in web access logs. Taxonomy of web mining scenario is given as follows:

![Web Mining Taxonomy](image)

Fig. 1: Taxonomy of Web Mining

Web content mining [4] is an automatic process that goes beyond keyword extraction. Since the content of a text document presents no machine readable semantic, some approaches have suggested restructuring the document content in a representation that could be exploited by machines.

Web structure mining [2-3], one of three categories of web mining for data, is a tool used to identify the relationship between Web pages linked by information or direct link connection. This structure data is discoverable by the provision of web structure schema through database techniques for Web pages.

Web usage mining [4] is the third category in web mining. This type of web mining allows for the collection of Web access information for Web pages. This usage data provides the paths leading to accessed Web pages.

II. Web Mining: A Broad View
Web mining is the process of extracting structured information from unstructured or semi-structured web data sources. Web Extraction also referred as Web Data Mining or Web Scraping.

Web data mining is done by creating programme or script written in any programming language that processes the unstructured or semi-structured html web pages of a target web site to extract information or data for converting unstructured data into structured format. Web data mining scripts and applications will simulate a person viewing a web site with a browser. With help of web data mining we can connect to a website’s web pages and request a information or a pages, exactly as your browser would do. The web server will send back the html web page which you can then extract specific information from that web page.

Web mining can be decomposed to do the following tasks:
1. Resource /information finding: referred to as the task of finding information from web.
2. Information pre-processing: referred to as automatic selection and pre-processing of information from web.
3. Generalization: referred to as discovering the patterns of individual web sites.
4. Analysis: referred to as the interpretation of mined pattern.

Taxonomy of web mining reveals the different categories of it, which includes web content mining, web structure mining and web usage mining as described earlier.

Besides this web content mining can follow two approaches
5. Agent based approach
The agent-based approach to Web mining involves the development of sophisticated AI systems that can act autonomously or semi-autonomously on behalf of a particular user, to discover and organize Web-based information. Generally, the agent-based Web mining systems can be placed into the following three categories:
- Intelligent Search Agents
- Information Filtering/Categorization
- Personalized Web Agents

6. Database approach
The database approaches to Web mining have generally focused on techniques for integrating and organizing the heterogeneous and semi-structured data on the Web into more structured and high-level collections of resources, such as in relational databases, and using standard database querying mechanisms and data mining techniques to access and analyze this information. It may include

www.ijcst.com
Web mining can be subdivided into two categories based on the kind of structural data used:

- Hyperlinks: A Hyperlink is a structural unit that connects a Web page to a different location, either within the same Web page or to a different Web page.
- Document Structure: In addition, the content within a Web page can also be organized in a tree-structured format, based on the various HTML and XML tags within the page.

III. Advantages of Web Mining
Web mining can obviously be quite beneficial to both businesses and individuals [11]. Web mining is attractive for companies because of several advantages. In the most general sense it can contribute to the increase of profits, be it by actually selling more products or services, or by minimising the costs. In order to do this, marketing intelligence is required. This intelligence can focus on marketing strategies and competitive analyses or on the relationship with the customers. The different kinds of web data that are somehow related to customers will then be categorised and clustered to build detailed customer profiles.

IV. Historical Background of Web Mining
Web mining can be considered as an application of data mining techniques to extract knowledge from the web data which includes web content (hyperlink, images, records etc.), web structures (hyperlink, tags), web usage (https logs, app server log).

Web mining techniques are the result of long process of research and product development. This evolution begun when business data was first stored on computer and internet [12]. Basically data mining techniques are used in web mining. Web mining is an extended version of data mining. Data mining is work upon off-line where as web mining is a work upon on-line. The main component of the web mining technology has been under development for years. Web mining research includes research areas such as internet, artificial intelligence, business application [13] and machine learning.

Most of the research efforts nowadays, propose systems, algorithms that combine methods from these Web mining categories. Therefore, Web mining moves to a more abstract level, where data representation is achieved using semantics. These semantics are defined using tools that emerged along with the Semantic Web vision, such as XML, RDF, and most importantly, ontologies.

VI. Survey of Mining Applications
The expansion of the World Wide Web has resulted a large amount of data that is now in general freely available for user access. The different types of data have to be managed and organized in such a way that they cannot be accessed by all users efficiently. Therefore the application of data mining techniques on the web is now focus of an increasing number of researchers.

Table 1: Survey of Different Application Areas of Web Mining

<table>
<thead>
<tr>
<th>Paper name</th>
<th>Field of Work</th>
<th>Conference/ Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Web mining frame work for security in e commerce</td>
<td>Web security</td>
<td>IEEE International conference on recent trends in information technology, 2011</td>
</tr>
<tr>
<td>Knowledge discovery and retrieval on wide web using web structure mining</td>
<td>Information retrieval/ knowledge discovery</td>
<td>Fourth Asia international conference on mathematical/ analytical modeling and computer simulation, 2010</td>
</tr>
<tr>
<td>Web mining in search engines</td>
<td>Link analysis, web dynamics</td>
<td>27th Australian computer science conference (ACSC2004), 2004</td>
</tr>
<tr>
<td>Hyper-textual language model for web information retrieval</td>
<td>Knowledge discovery</td>
<td>IEEE, 2010</td>
</tr>
<tr>
<td>Web Mining Application in University Library Personalized Search Engine</td>
<td>Web usage mining, search engine</td>
<td>IEEE, International Conference of Soft Computing and Pattern Recognition, 2011</td>
</tr>
<tr>
<td>Research on E-commerce Application Based on Web Mining</td>
<td>Architecture of E-commerce based on web mining</td>
<td>International Conference on Intelligent Computing and Cognitive Informatics, 2010</td>
</tr>
<tr>
<td>An approach in web content mining for clustering web pages</td>
<td>Clustering web pages</td>
<td>IEEE, 2010</td>
</tr>
<tr>
<td>Terror Tracking Using Advanced Web Mining Perspective</td>
<td>Web Security</td>
<td>IEEE, IAMA2009</td>
</tr>
<tr>
<td>Web content mining using web design patterns</td>
<td>Content mining using web pattern</td>
<td>IEEE IR, Las Vegas, Nevada, 2008</td>
</tr>
<tr>
<td>web mining in search engines</td>
<td>Link analysis, web dynamics</td>
<td>27th Australian computer science conference (ACSC2004), 2004</td>
</tr>
<tr>
<td>web mining in search engines</td>
<td>Link analysis, web dynamics</td>
<td>27th Australian computer science, conference (ACSC2004), 2004</td>
</tr>
</tbody>
</table>
Research on web mining includes distinct application areas which includes information retrieval, optimizing the performance of search engine, e-commerce application, pattern analysis, link analysis-security etc. Most of papers surveyed by me concentrate on the field on knowledge discovery or information retrieval which is an emerging application area of web mining.

J.Srivastava, R.Colley concentrated on the pattern of web data using a well known toll known as WEBMINER. On the other hand, Milos Kudelka, Vaclav Snasel, Ondrej Lehecka concentrated on pattern discovery from the web using web content mining. They have used the concept of Preprocessing of the code of the html. Information retrieval or knowledge was the field of research of Sekhar babu boddu, V.P Krishna Anne, Rajasekhar Rao kurra, Durgesh Kumar Mishra using the concept of HITs and page rank. Ricardo baeza-yates enlightened the field of enhancing the performance of search engine by link analysis. Analyzing the behaviour of a visitor of web site was the field of research of Juan Velasquaz, Hiroshi Yasudha, Terumasa Aoki. Current trend of web mining research includes pattern discovery, terror tracking, and web security performance analysis of search engines.

V. Conclusion and Future Direction

This paper provides a more current evaluation and update of web mining research available. Literature has been reviewed based on the field of three categories of web mining namely, web content mining, web usage mining, web structure mining. Year wise development in each type of application areas in included in the table. Future trends depict that web data mining has some research scope in the field of web service optimization, fraud and threat analysis, identification of web matrices etc. The importance of web data continues increasing. Based on the requirement of the web application users, different fields of research can be introduced. Web data mining perhaps still in infancy and much search is being carried out in this field of research.

VI. Acknowledgment

I offer my sincere gratitude to Dr. Abhoy Chand Mondal, Professor, Department of Computer Science, University of Burdwan for constant encouragement and support provided throughout period of this research.

References
