Journal of Science and Engineering Research
Vol. 1, No. 1, 2022, pp. 15-28
eISSN: 2957-7187

Web-based Data Mining Technology and its Application

Andrzej Olczak[1] and Zdzisław Pólkowski[2]

[1,2]Pedagogical University of Cracow, Kraków, Poland
andrzej.olczak@up.krakow.pl

DOI: http://dx.doi.org/10.56828/jser.2022.1.1.2
Article history: *Received (September 9, 2021); Review Result (October 25, 2021); Accepted (December 2, 2021)*

Abstract: Data mining is a brand-new information technology emerging with the development of database technology and artificial intelligence technology in recent years. It is also an important subject that needs to be solved urgently by computer science and technology, especially the development and universal application of computer networks. Data mining is to discover interesting, hidden, and previously unknown knowledge from a large number of databases. Data mining technology mainly studies structured data mining, while Web data mining is applied technology research, which is to extract interesting and potential patterns from semi-structured or unstructured Web pages. With the increasing resources on the World Wide Web, how to use effective methods to obtain the required information has become an important problem for researchers. Mining refers to data from the Web, such as Web logs and Web page content. Page content discovers users' browsing patterns or searches for related Web pages. Web mining can not only provide reasonable suggestions for site managers, investors, and advertisers but also provide powerful intelligent search engines and personalized services for users. Because it is an unstructured, dynamic, and distributed information system, it is difficult to mine it directly, and the acquired knowledge is unreliable. However, the log of the server has a complete structure. Therefore, we can mine the usage of data by mining web log files. The performance of the Web server system can be enhanced, network service quality for users can be improved, and potential customers in e-commerce can be identified by evaluating and investigating the rules in log data. In this paper, the dynamics of Web mining are analyzed, issues are presented, data mining and Web mining technologies are introduced, and the connection between XML and Web mining is discussed.

Keywords: Data mining, Web mining, Web data, Computer network

1. Introduction

In recent years, with the development of the Internet, computers, networks, and communication are integrated into one, and the emergence of new concepts such as network economy has become the most striking research topic of information science with its great social benefits and rich connotation of challenges and opportunities. However, while the network brings a lot of information quickly and conveniently, it also brings a lot of problems, such as too much information, difficulty in digesting information, difficulty in identifying true and false information, difficulty in ensuring security, inconsistency of information forms and difficulty in unified processing. How to comprehend historical data and utilize it to anticipate future behavior, how to rapidly and reliably get valuable network information, how to extract

knowledge from these enormous amounts of data, etc. These concerns are what give rise to knowledge discovery and data mining. Knowledge Discovery Abbreviation and Data Mining Abbreviation are cross-cutting research fields of statistics, artificial knowledge, pattern recognition, parallel computing, machine learning, database, and other technologies. Data mining is the process of retrieving hidden, unknown, and possibly beneficial knowledge from massive databases or data warehouses that pertains to people. It is currently one of the most cutting-edge study areas in the world for databases and decision-making. It was created and developed quickly in response to the pressing need to use information and resources more thoroughly and fully, garnering significant interest from both academics and industry. With the rapid development, the amount of information in the seventh field is increasing at an alarming rate. Facing the abundant information resources in the seventh field, which are abundant, distributed, dynamic, heterogeneous, complex, and unstructured, how to find and extract the data and useful information that users want has become a big problem, and it is this problem that accelerates the birth of begging mining technology. Mining is the process of identifying and extracting implicit, unexplored, and possibly beneficial patterns and information from a sizable number of Web pages and Web activity. By thoroughly integrating the technologies of computer network, database and data warehouse, artificial intelligence, information retrieval, information extraction, machine learning, statistics, probability theory, visualization, computer linguistics, natural language understanding, and other fields, it combines the traditional data mining technology with the Web. It is based on data mining, text mining, and multimedia mining [1][2][3][4][5][6][7][8].

Data mining gives way to web mining, which has numerous distinctive properties as compared to conventional data mining. Traditional data mining just looks at structured data in the database, and it uses storage structures like relational tables to find knowledge. However, the object of mining is a large number of heterogeneous, distributed, and semi-structured data and its particularity determine that mining cannot directly apply traditional mining methods and models in the database field. How to solve the problem of data standardization and preprocessing, make the mining system and database closely combined, and provide an integrated information processing environment has become the premise of Web mining.

Web mining and Web information retrieval are also different, and the differences are mainly shown in the following aspects, as shown in Table 1.

Table 1: Differences between web mining and web information retrieval

	Web Mining	Web information retrieval
Purpose	Associating a large number of seemingly irrelevant data, applying data mining methods to discover the rules and knowledge for decision support.	Find a subset of documents that can meet the user's query request from a large number of documents, and usually can't find the links hidden behind the data
Focus	Learn more about document content, structure, and usage.	Words and links explicitly stored in documents
Methodology	The features of the target information are extracted from the text, and the result is independent of the user's information. Demand is often unpredictable for users.	Goal-driven, users need to put forward query requirements, simple targets expressed in the form of keywords can't handle complex fuzzy targets in the form of samples given by users.
Evaluation methodology	Gain, certainty, simplicity	Precision, recall

Web mining is different from dead information retrieval; it is a higher technology than information retrieval. However, they complement each other. On the one hand, they have their strengths and have their applicable occasions. On the other hand, we can use the research results of Web mining to improve the accuracy and efficiency of information retrieval, improve the organization of retrieval results and make the information retrieval system develop to a new level.

2. Related Work

Web Utilization Miner (WUM) of Humboldt University mines sequential patterns in weblogs and provides an environment integrating log preparation, pattern query, and visualization.

WebWatcher proposed a personalized service based on Web log mining. Every page request of users has to go through the proxy server of WebWatcher. The proxy server guides users to browse along the appropriate path based on their interests of users, and the location and relevance of collected web pages.

WEBMINER system provides a framework for mining logs and can analyze the relationship between files accessed by users.

Jian Pei et al. proposed a concise and highly compressed WAP-Tree structure, and the mining algorithm based on this structure simplified the generation process of user browsing path patterns.

Ming-Syan Chen and others put forward browsing path mining and introduced the concept of the maximum forward path to analyze Web site logs.

Information retrieval research has made significant strides since the 1960s, including the development of the index model, document content representation, matching technique, etc. Search engines like Yahoo, Google, and others have been created as a result of these accomplishments being successfully applied to the Web. At present, most search engines focus on searching for text information, trying to traverse the whole Web in a centralized way, generating indexes for all documents on it, finding the corresponding information from the index database according to a certain algorithm according to the user's query request, and returning it to the user to complete the right search, such as locating the page containing a certain keyword.

Search engine partially solves the problem of resource discovery, but their coverage is limited, their accuracy is not high, their hardware facilities are consumed greatly, their maintenance is difficult, and their effect is far from satisfactory. Although the improved meta-search engine method can make multiple dead search engines query simultaneously and expand the coverage of queries, the meta-search engine can't fundamentally solve the above problems because of its dependence on search engines.

In addition, the purpose of search engines is to discover the resources on the flight, but as far as knowledge discovery is concerned, even if the retrieval accuracy is higher, search engines are not competent. Therefore, it is necessary to develop new technologies higher than information retrieval. Effective, novel, useful, and understandable patterns are found from a large number of data sets, and data mining technology is adopted in the database field. However, most of the work of data mining involves structured databases, and there is little work to deal with heterogeneous and semi-structured information. One way to solve these problems is to combine traditional data mining technology with Web mining.

3. Data Mining and Web Mining

3.1. Data mining

By examining the data that is already in the database, data mining helps to address issues. The practice of identifying data trends is known as data mining. This procedure must be fully or partially automated.

Another field of computer technology, artificial intelligence [9][10][10][11], has made great progress since its birth Knowledge discovery in databases has benefited from the integration of artificial intelligence and database technologies. Expert systems, machine learning, pattern recognition, statistics, intelligent databases, knowledge acquisition, data visualization, and other subjects are all involved in the interdisciplinary topic of knowledge discovery in databases. The knowledge discovered from the database can be used in many aspects such as information management, process control, scientific research, decision support, and so on.

Because of the importance of data mining, data mining has concentrated the main energy of researchers, and related research work has also made great progress. Usually, according to the difference between discovery task and discovery goal, we divide data mining into the following categories [12][13][14].

(1) Classification

Classification refers to mapping data to predefined groups or classes. Because the category has been determined before the analysis and test, classification is usually called guided learning. Data classification is to dig out the description or model of each kind of data in the database, and the classes in these databases are established by training data in advance.

In data mining, there are many research results of classification algorithms, such as CART, C45, ID3, SLIQ, Knn, GA-Knn, and so on.

(2) Regression

Regression refers to mapping a data item to a real-valued predictive variable. Regression involves learning a function that can complete the mapping. Firstly, regression assumes that some known types of functions can fit the target data, and then uses some error analysis to determine a function that fits the target data best.

(3) Time series analysis

The attribute values of the data are dynamically changing over time in time series analysis. Three fundamental tasks are performed by time series analysis: first, comparing the similarity of various time series using distance measurements; second, analyzing the structure of the lines in the time series diagram to ascertain the behaviour of occasionally distinct time series; and third, predicting the future value of data using the historical series diagram.

(4) Forecast

Many practical data mining applications need to predict the future data state based on past and current data. Prediction can be regarded as a classification. Prediction is mainly to predict the state of future data rather than the current state. Applications include flood forecasting, speech recognition, machine learning, and pattern recognition. Besides time series analysis and regression analysis, other techniques can be used to predict future value.

(5) Clustering

While clustering falls under the category of unsupervised learning, classification is a type of directed learning. Data clustering separates tangible or intangible items into a number of groups. Only a high degree of resemblance exists between objects within each group, however there is a low degree of similarity between groupings. A group is typically a class, however unlike data classification, the outcomes of clustering are primarily reliant on the data that is now being processed; we are unaware of the category structure and the category to which each item belongs beforehand. Additionally, data clustering requires a significant amount of processing and has a far higher time complexity than data classification.

At present, there are many different methods and technologies for data clustering. Common algorithms are PAM, K-means, Clara, BIRCH, etc.

(6) Summary

The summary is to map data to a subset with a simple description. Aggregation is sometimes called characterization or generalization. Summarize to extract or get representative information from the database, which can be done by searching some data, or getting some summary information such as the average value of some numerical attributes from the data. Summarize and concisely characterize the contents of the database.

(7) Association rules,

The aim of identifying rules is to identify those powerful rules whose confidence and support are greater than a given value from the database. The so-called association rules refer to the dependency between data objects. Discovering association rules from databases has been the most studied in recent years. At present, it has developed from the discovery of association rules at a single conceptual level to the discovery of association rules at multiple conceptual levels. With the deepening of the concept level, the information provided by the discovered association rules becomes more and more specific. This is a process of gradually deepening the discovered knowledge.

At present, there are many technologies available for data mining. The technologies and methods in machine learning, pattern recognition, artificial intelligence, and other fields have been improved accordingly, and most of them can be applied to data mining. Commonly used methods include decision tree, rough set, neural network, genetic algorithm, and concept tree.

3.2 Web data mining

Since the emergence of the Web, it has developed at an alarming rate. In the early days of the Web, information sharing was the mainstay. In recent years, e-commerce, e-library, distance education, and so on have become the main applications of the Web, which has promoted the development of PANI at a faster speed. At the same time, higher requirements have been put forward for the design and functions of touch sites. It is required to be intelligent, to be able to quickly and accurately find the information required by users, to provide different services for different users, to allow users to customize pages according to their own needs, and to provide users with a product marketing strategy information and so on. It is difficult to fully realize all functions, and it needs a breakthrough in artificial intelligence and natural language understanding.

Data mining is the process of applying data mining technology to massive amounts of data and discovering meaningful, original, potentially practical, and lastly, intelligible pattern rules. Web data mining is the process of removing information from server data files that

people are interested in. Web mining is analyzing the content of documents, the use of available resources, and the relationship between resources.

3.3. Object and classification of web data mining

The objects of data mining mainly include log data, e-commerce data, page data, and other data.

(1) Log when individuals browse the site, the server will generate several types of files, such as, to record the basic information of users' access. It records all kinds of access information of users and saves the data of user request failure, such as lost connection, authorization failure or timeout, etc.

(2) E-commerce data Structural data stored in a traditional relational database is used to store e-commerce information in daily business activities.

(3) A page whose data meets the side standard.

(4) Other information mainly includes a series of information such as user registration and survey information. It is some additional information to get better mining results.

Web mining realizes the search for several access patterns, Web structures and rules, and dynamic content. Generally, mining can be divided into three categories: content mining, structure mining, and using mining fields. The structure can also be considered as a part of content mining, so mining can be simply divided into two categories, namely, content mining and usage mining. Content mining, also called page mining, is to mine useful knowledge from page content. Using mining, also called log mining, is to discover the patterns of users visiting pages by mining log records. Figure 1 shows the classification diagram of Web mining.

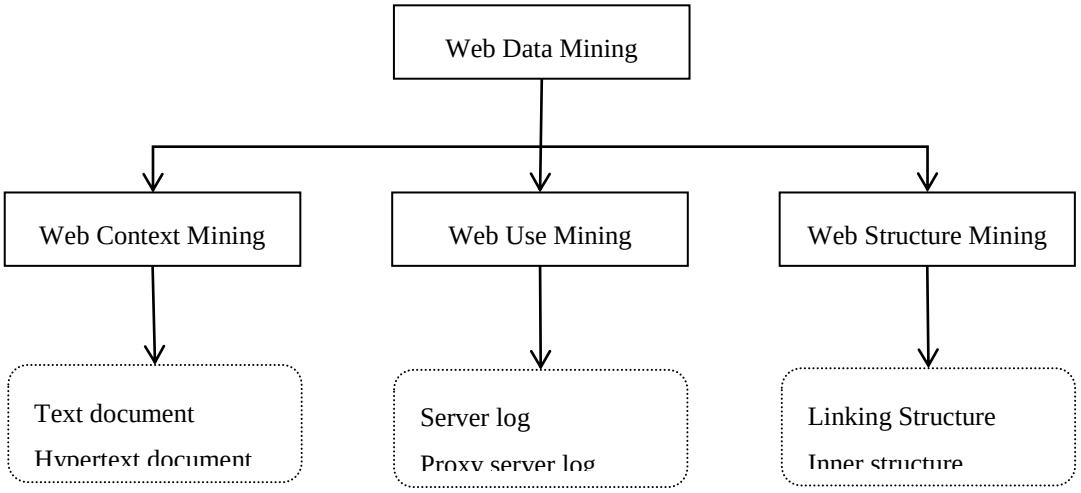

Figure 1: Web mining classification diagram

(1) Web content mining

Web content mining mainly refers to Web text mining and multimedia mining. This paper mainly introduces Web text mining. Web text mining is mainly to summarize, classify, cluster, analyze association rules and predict the trend of a large number of documents on the Web.

Text feature representation is the foundation, and text classification and clustering are the most basic and core functions.

Text summarization is text summarization, which refers to extracting key information from documents and summarizing and explaining the contents of documents in a concise and clear form so that users can understand the overall contents of documents or document collections without browsing the full text, and its purpose is to condense the text information and give its compact description. Commonly used summarization algorithms include part-of-speech tagging, word segmentation analysis, and statistical extraction of high-frequency words to determine the summarization. The general process of text classification is shown in[Fig. 2].

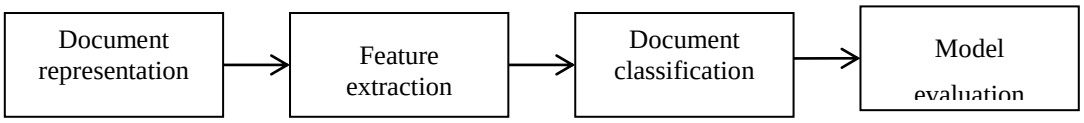

Figure 2: General process of text classification

Text clustering refers to grouping a group of objects into several categories according to similarity. Hearst et al.'s research has proved that documents related to users' queries are usually clustered close to each other, and far away from documents irrelevant to users. The current text clustering algorithms are mainly divided into two categories: the hierarchical aggregation method represented by G-HAC and other algorithms; the Plane division method represented by algorithms such as K-Means.

Commonly used research methods of Web content mining include word frequency statistics, classification algorithm, clustering algorithm, machine learning, pattern recognition, and metadata. Web content mining is mainly used in topic extraction and text classification, semi-structured query language and pattern extraction, Web heterogeneous data integration learning patterns or rules, and information discovery based on specific knowledge fields. The main research achievements in this field are as follows: Bayes model-based on keyword frequency: the improvement of the Maximum entropy algorithm -Bayesian algorithm; Understanding of natural language; Co-training utilizes the combination of hyperlink information and text content.

(2) Web structure mining

Web structure mining is a process of mining potential link structure patterns and deriving information and knowledge from Web organization structure and link relationships. By classifying the hyperlinks of pages, we can judge and identify the attribute relationship between page information. Because there is more or less structural information in the Web pages, by studying the internal information structure of pages, we can find out other page information patterns related to the page set information selected by users, and use the structural mining method to detect the completeness of the information displayed by the site in the database.

Jon Kleinberg of Cornell University [15] proposed a widely adopted technology to explain the information in hyperlinks. He is based on the assumption that when establishing a link from one site to another, the manager of the website will think that the linked website is important, and the two websites usually have similar or related content. A station linked to many authoritative sites is called a hub, and the sites connected by many hubs become authorities. Based on the above theory, Kleinberg created the algorithm hits (hyperlink-included topic search)' to identify authoritative sources.

In addition, many academic research institutions have studied and analyzed hyperlinks on the Web and put forward many algorithms for Web structure mining. For example, Weiss analyzed the link structure of the clustering method. By mapping the link structure to the information in the standard relational database, the query of the Web was realized with SQL statements. Kleinberg searches for the Authorities page and Hubs page by calculating the eigenvector of the Web corresponding correlation matrix; Brin and Page calculate PageRank values of Web pages by using the inlink and outlink of pages. Based on this, they find authoritative pages and improve their algorithms by using Markov chain theory and propose a random algorithm SALSA to analyze hyperlink structure.

At present, the main method to analyze the structure of Web hyperlinks is to map several pairs into the form of a directed graph or undirected graph and then analyze them by graph theory according to certain heuristic rules. Compared with typical powerful algorithms, this algorithm is an important tool to evaluate the authority of web pages. That is, the algorithm is used to rank the relevance of a large number of retrieved results by combining with factors such as tags and word frequency statistics, and the authoritative web pages are put in front. Behind the high efficiency, it also has its shortcomings. It completely ignores the content of the web page. A twist algorithm is proposed to evaluate the importance of webpage content, and content mining is added.

The application fields of Web structure mining mainly include the ranking of search engine query results, searching related documents, calculating pages, and determining the crawling priority of the main content and features of a site.

(3) Use Web mining

Web usage mining is a technology of mining logs of Web servers to obtain information and predict users' browsing behavior, that is, mining users' access patterns from user access logs. The process of mining Web access patterns is generally divided into four stages, namely, data source determination, data preprocessing, pattern discovery, and pattern analysis. As shown in Figure 3.

The log records on the Web server are stored in the form of text, in which a large amount of data irrelevant to mining work is stored, so it is difficult to use them directly for mining work. Because of the different servers or setting parameters, the information in the log files will be different, but they all contain the basic information of the accessing users. Log records usually have two formats: general log format and extensible log format. Its main structure is shown in Table 2.

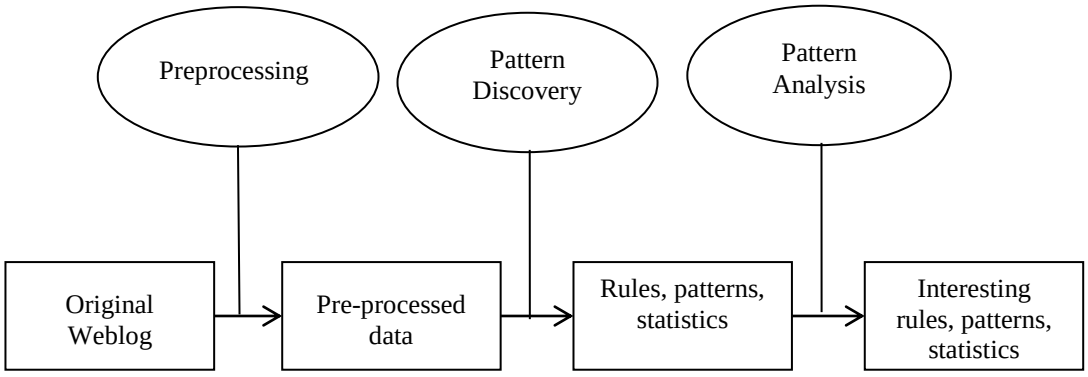

Figure 3: Process of web user access mode

Table 2: Weblogging format

Field	Description
date	Date the operation was performed.
time	The time when the activity took place
C-IP	Client address of access server
S-IP	The address of the server that generated the log record.
method	Operations performed by the client
S-port	Port number of the client connection server
URI	Resources visited by
status	HTTP or FTP
agent	Browser used by the client
Number of bytes sent	Number of bytes sent by the server
Number of bytes received	Number of bytes received by the server
Time spent	Time is taken by the operation
version	The version of HTTP or FTP protocol used by the client
host	Displays the contents of the host title.
Cookie	Received cookie content.
Reference	The previous site visited by the user
Sitename	Number of Internet services and instances visited by the client
S-IP	The IP address of the server that generated the log record.

In Table 2 the items in italics are the records of the extended format log, some of which are hard to get in daily application. For example, because it involves personal privacy, its use requires the cooperation of users, and now most browsers can prohibit its use. A general log format record is shown in Table 3:

Table 3: Records in general log format

2006-11-06 00:00:52211.28.60.135-192.168.2.12 80 GET /images/stylescss-200 Mozilla/4.0+(compatible; +MSIE+6.0; +Windows+NT+5.1)
2006-11-06 00:00:52 211.28.60.135-192.168.2.12 80 GET /images/jd bggif-200 Mozilla/4.0+(compatible; +MSIE+6.0; +Windows+NT+5.1)
2006-11-06 00:00:52 211.28.60.135-192.168.2.12 80 GET /images/jd osgif-200 Mozilla/4.0+(compatible; +MSIE+6.0; +Windows+NT+5.1)
2006-11-06 00:00:52211.28.60.135-192.168.2.12 80 GET /images/jd spgif- 200 Mozilla/4.0+(compatible; +MSIE+6.0; +Windows+NT+5.1)
2006-11-06 00:00:53 211.28.60.135-192.168.2.12 80 GET /images/jd argif- 200 Mozilla/4.0+(compatible; +MSIE+6.0; +Windows+NT+5.1)

4. Design of Web Data Mining System

4.1. System design

Compared with data mining activities, the data objects handled by traditional database methods are structured data stored in relational databases, and the amount of data is relatively small. It is considered that the mining object of Web data mining is massive data, and the general storage format of its data source-access log is text. The development scheme of a data

mining system is to combine the features of the relational database, such as convenient operation, structured data storage, convenient data processing, etc., with the algorithm of Web data mining by taking advantage of both database technology and data mining.

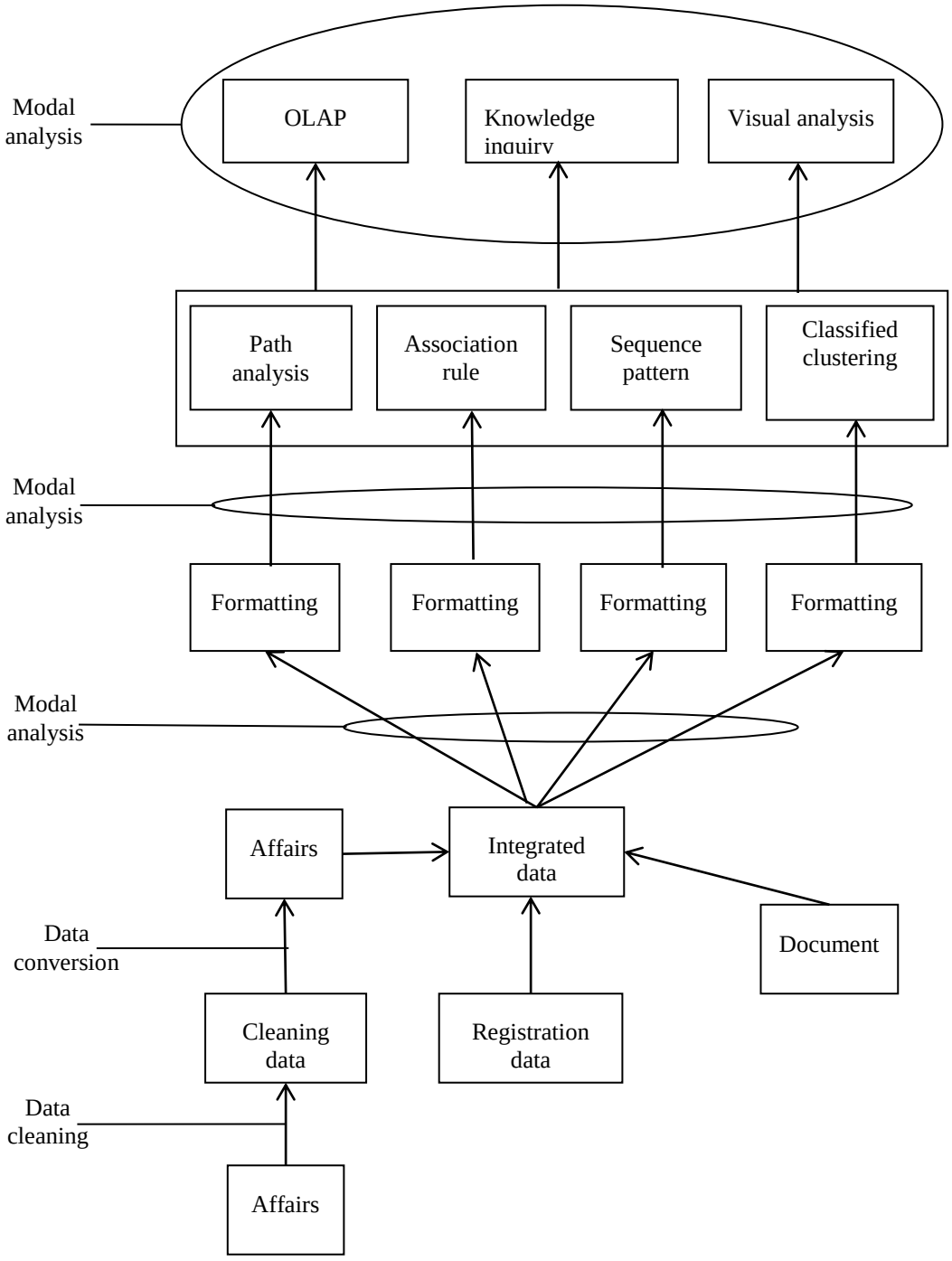

Figure 4: General process of web data mining

Traditional data mining algorithms [16] are all applied to specific data transaction models, but when traditional algorithms are introduced into data mining for data use, the data's characteristics, such as dynamic, timeliness, and heterogeneity, should be considered. The general process of data mining is shown in the figure. This process can be divided into three stages. The first stage is the data preparation stage, and the main task is to transform data into appropriate transaction forms, including data cleaning, transaction identification, data integration, data conversion, etc. The second stage applies various data mining algorithms to mine knowledge, and the third stage is pattern analysis, which is used to explain and evaluate the mined knowledge.

As can be seen from Figure 4, the access log in the Web server is the main data source of Web data usage mining, but it is not necessarily the only data source of Web mining activities. Generally speaking, when users visit a specific page, they will access many other data items irrelevant or redundant to data mining tasks at the same time. We need data cleaning to remove redundant or irrelevant data from the original data. Next is transaction identification. The cleaned data needs to be divided into many logical units to represent a user transaction. The process of user transaction identification is very important in our application because our mining activities are aimed at specific users. After the transaction identification task is finished, the next step is to convert the transaction data obtained to meet the data formats required by different data mining tasks, for example, the data formats required by association rule discovery are different from those found by clustering. Convert the data into a certain Web data mining format, so that the corresponding data mining algorithm can be applied to the Web transaction data.

In the second stage, the main work adopts the corresponding mining algorithm to mine the processed data. Various mining algorithms include association rules, clustering classification, sequence patterns, and so on. Because there is a huge amount of data stored in the log, timeliness should be considered when the mining algorithm is adopted, and the characteristics of begging log data itself should be considered.

The last step of mining activity is pattern analysis, including online analysis (OLAP), knowledge query, visualization technology, and so on. OLAP, which is the same as OLAP in a traditional database, mainly analyzes the mining results. It can better understand the mined knowledge and provide users with convenient statistical analysis functions. As technology for managing large-scale data, data mining naturally needs an effective query mechanism. The query mechanism allows users to select only the patterns they need or are interested in, thus greatly improving efficiency. Visualization technology of data mining refers to the use of a graphical user interface to help users mine and understand a large number of complex data. Visualization technology provides great convenience for users to manage and understand a large number of patterns.

4.2. System structure

Considering the expansion of system functions, the system we designed is divided into the following modules according to functions: master control module, data acquisition module, data cleaning module, a transaction identification module, various mining modules, and visualization modules. The system structure is shown in Figure 5. The processing flow of the system conforms to the general process of using and mining dead data, and the main control module controls the running process of other modules. First, the data acquisition module reads the access log from the proxy server A into the relational database, and the data cleaning module removes useless or irrelevant data from the original data. Then the

transaction identification module converts the cleaned data into a specific transaction format according to different mining tasks. According to users' requirements, the system calls specific mining modules by passing parameters and carries out specific data mining activities based on the transaction database. The mining module processes the data, and then returns the obtained mining results to the relational database, or directly serves as the input of other application systems. We can intuitively describe the mining results stored in the database through the visualization module.

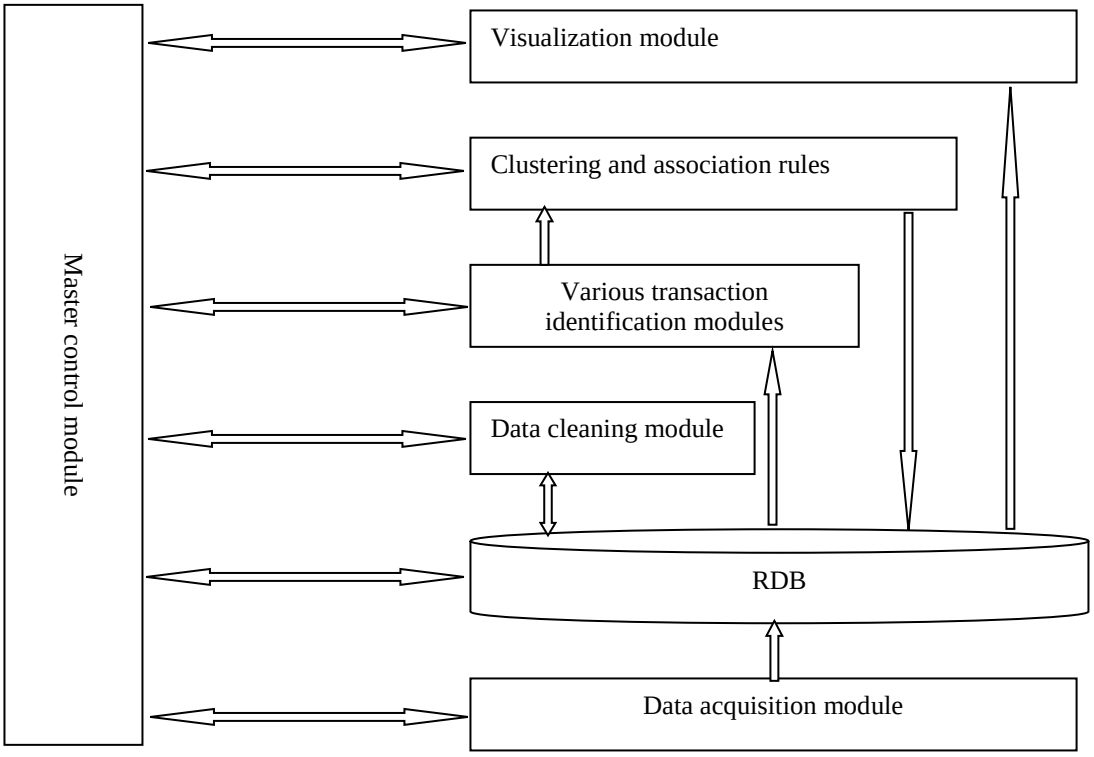

Figure 5: System structure diagram

5. Conclusions

The web has become an important way for people to obtain information. With the increase in Web information, it is difficult for people to search and browse the information they need. Although the search engine has become the most popular auxiliary information retrieval tool, it still can't meet people's requirements. The web is a huge and widely distributed global information service center, which contains abundant data and provides abundant resources for data mining. However, due to the characteristics of the Web, the dead mining is significantly different from the traditional data mining. Obtaining information from the Web will be the first choice in the next few years. Therefore, in the research process of this paper, the author deeply feels that self-mining, as a new development field of data mining, has little personal strength in the practice of the project, and the work we have done is only a small part of it. This paper transitions from data mining technology to Web mining technology. According to the process of Web data usage mining, firstly, the technology of the data preprocessing stage of Web mining is studied. The role of this stage is to remove a large amount of useless and

irrelevant data from the logged user access information and to structure the unstructured or semi-structured information into a database in the form of transactions or user sessions.

References

[1] E. Iosif, K. Christodoulou, & A. Vlachos. (2021). Web mining for estimating regulatory blockchain readiness.

[2] S. Yadao, A. V. Babu, & M. Janarthanan. (2021). A proposal on semantically enhanced web mining-based deep neural network framework. *Materials Today: Proceedings*.

[3] G. Kumar & V. Rishiwal. (2020). Malicious user nodes detected by web mining-based artificial intelligence technique. *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems*, 28(5).

[4] N. Thongmaun & W. Thamviset. (2020). Supplement products data extraction and classification using web mining.

[5] B. Vka & C. Esa. (2020). Scenario development using web mining for outlining technology futures. *Technological Forecasting and Social Change*, 156.

[6] F. A. Baltoumas, Z. Sofia, & K. Evangelos. (2021). OnTheFly2.0: A text-mining web application for automated biomedical entity recognition, document annotation, network, and functional enrichment analysis. *NAR Genomics and Bioinformatics*, 4, 4.

[7] B. Zhu, L. Zhang, & Y. Zhang. (2021). The early warning method of drug adverse reaction monitoring based on data mining algorithm, 1852(3), 032052.

[8] S. Luo & S. Y. He. (2021). Using data mining to explore the spatial and temporal dynamics of perceptions of metro services in China: The case of Shenzhen, 1023, 26-29.

[9] M. Mamdani & A. S. Slutsky. (2020). artificial intelligence in intensive care medicine. *Intensive Care Medicine*, 5.

[10] S. Korver, E. Schouten, & O. A. Moultos. (2020). Artificial intelligence and thermodynamics help solve arson cases. *Scientific Reports*, 10(1).

[11] C. Nicholas. (2021). Artificial intelligence as a general-purpose technology: An historical perspective. *Oxford Review of Economic Policy*, (3), 3.

[12] A. Dardzinska & M. Zdrodowska. (2020). Classification algorithms in material science and engineering data mining techniques. *IOP Conference Series Materials Science and Engineering*, 770, 012096.

[13] A. Al-Rasheed. (2021). Identification of important features and data mining classification techniques in predicting employee absenteeism at work. *International Journal of Electrical and Computer Engineering*, 11(5), 4587-4596.

[14] A. Ishak, K. Siregar, & Aspriyati. (2020). Orange software usage in data mining classification method on the dataset lenses. *IOP Conference Series: Materials Science and Engineering*, 1003(1), 012113.

[15] A. Rambachan, J. Kleinberg, & S. Mullainathan. (2020). An economic approach to regulating algorithms. *NBER Working Papers*.

[16] M. M. Kavitha & B. Ananthi. (2020). Fuzzy SVM clustering using plane sweep algorithm in big data mining.

This page is empty by intention.