Web Content Mining Techniques for Structured Data: A Review

Mustafa Ali Bamboat
High Court of Sindh, Karachi, Pakistan

E-mail: bamboat3@gmail.com

Ghulam Sarfaraz Khan
Working in Outriders (Pvt.) Ltd, Karachi, Pakistan

E-mail: ghulamsarfaraz@yahoo.com

Naadiya Mirbahar
Research Lab of Artificial Intelligence and Information Security, Faculty of Computing Science and Information Technology, Benazir Bhutto Shaheed University Lyari, Karachi, Pakistan

E-mail: naadiya.khudabux@yahoo.com

Sheeba Memon
Department of Information Technology, Government College University Hyderabad, Sindh, Pakistan

E-mail: sheeba.memon@gcuh.edu.pk

Received: 28th June, 2022; Accepted: 9th September, 2022; Published: 21st September, 2022

***Abstract*:** The Web accumulated vast volumes of data, making it difficult to extract data according to customers' needs; hence, web mining came to tackle these challenges. Web mining has involved databases, information retrieval systems, and artificial intelligence. Web Mining is an extensive, interdisciplinary, and dynamic area; it consists of three techniques: Web Content Mining, Structured Web Mining, and Web Usage Mining. This paper gives an overview of web mining techniques and explores the Web Content Mining Techniques, such as Wrapper Generation, Page Content Mining, and Web Crawler, including their classification and tool being used.

Index Terms: Web Mining, Content Mining, Structured Data Mining, Web Crawler.

1. INTRODUCTION

In the early days, there was no vast data over the Internet, so there was no need for web mining tools; as years passed, accumulated large amounts of data. It became a difficult task to retrieve data according to users' needs. Web mining came to this issue as a rescue. A big quantity of information is stored in the computer system, in the case of web mining the method of finding required information from available Web data.

Web mining uses data mining techniques to analyze web content patterns, such as texts, images, and videos [1]. It is classified into three techniques: Web Structure Mining, Web Content Mining, and Web-Usage Mining. Computers are a repository of knowledge that contains a huge amount of information; web mining i.e one of the operations of the data mining process that is utilized for finding information and knowledge present in the Web data. A big quantity of information is stored in the computer system, in the case of web mining, the method of finding required information from available Web data.

Web Structure Mining evaluates the relationship between web pages directly related or linked to each other by information. It addressed two big problems: irrelevant search results on the Web and the inability to index a large amount of information.

Web Content Mining is a method of exploring knowledge or resource from millions of Web-based dispersed data content. The data's content consists of text, images, audio, video, hyperlinks, and metadata, and as much of the web content is in text format, it is mainly concerned with text mining. There are three content mining techniques, such as (i) Classification: Supervised Learning techniques of ML are used in it, standard techniques are Nearest-Neighbor Classifier, Feature Selection, and Decision Tree. (ii) Clustering: Unsupervised Learning technique of ML and (iii) Association technique, which imply the association rules, for example, if the customer purchase tea then 50% of cases are that customer also buy sugar with it.

Web Usage Mining deals with the navigation patterns of the user. It predicts user behavior by mining the data from logs, user profiles, user sessions, cookies, user queries, bookmarks, mouse clicks, and scrolls [2]. Figure-1, demonstrate the Web Mining Taxonomy.

In this paper, we overview Web Content Mining and explore the Structured Data Mining Techniques, such as Wrapper Generation, Web Crawler, Page Content Mining, and their application are also discussed. This research paper is aligned as follows: for the relevant work, Section 2 provides a summary. Section 3-give an overview of Content Mining, Section 4-explore the Techniques, Section 5 discussed the tools used in structure data mining, and Section 6 presents the definitive conclusion.

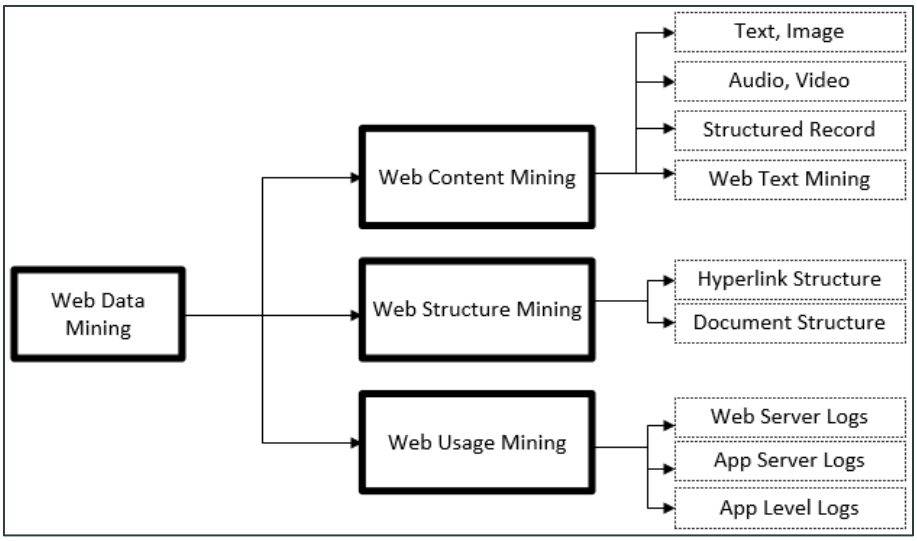

Fig 1. Web Mining Taxonomy

2. RELATED WORKS

In [2], the analysis was carried out on the various web content mining techniques, patterns, and the areas in which potentially web content mining can be used; this research also reflects web content mining in web usage mining aspects. The authors clarified the value of web mining in the overview section of [2], such as using web mining to acquire the required information, develop new data from the relevant information, customize the information according to user requirements, or discover knowledge about a client or individual users. Five content mining tools were also compared in the study: Automation Anywhere, Web Info Extractor, Web Content Extractor, Screen Scraper, and Mozenda. In the end, the authors highlight the possible areas in which content mining can be used, such as Online News sites, decision-making application of distance learning, or understanding between customer and seller by providing accurate knowledge of what customer needs.

A similar study is presented in [3]. The authors conducted a study on web mining tools and techniques, explaining all three web mining techniques, such as web content, web structure, and web usage mining, and briefly explained subtypes of each technique, as web content min consists of four types of data mining, such as Unstructured Data Mining, Structured Data Mining, Semi-Structured.

Data Mining, and Multimedia; moreover, this research also explains that the algorithms being used in these techniques, such as Google's PageRank Algorithm and HITS algorithm, are used in web structure mining techniques. Fourteen web mining tools are also considered in terms of their General Features, Advantages, and Limitation.

In [4], the author published a detailed report on the technique, architecture, and types of Web Crawler, used for structured data mining. He explained that Web Crawler plays an essential role in search engines, so it must be adequate, scalable, robust, extensible, and quality fetcher of content. The general architecture and types of web crawlers are also discussed in-depth in this report. This study also provides an overview of eight types of Web Crawlers, namely Customary Web Crawler, Deep Web Crawler, Incremental Web Crawler, RIA (Rich Internet Application)-, Web Crawler Unified Model, Focused Web Crawler, Parallel Crawler, and Distributed Web Crawler.

In [5], the different Wrapper approaches used to extract information from unstructured data were performed. Five wrapper extraction techniques, such as Automated Wrapper Generation, Semi-Automatic Generation, Wrapper Induction, Wrapper Maintenance, and Information Extraction approaches, were explored in this report. The comparative study was performed on the algorithms used in each wrapper approach for data extraction. Four algorithms consist of the automatic Wrapper Generation approach: Data Path Matching & Alignment, which reflected accuracy and recall as 96 percent and 93 percent respectively, while the Automatic Matching Trinity Algorithm had a significant extraction time than the Automatic Matching Roadrunner and Partial Tree Alignment algorithms. There are four sub-approaches to the Semi-Automatic Wrapper approach: Visual Extraction, Spatial Reasoning, Logic-Based and Regular-Based approaches. Findings have shown that visual extraction is the most time-consuming technique since it uses IE (Information Extraction) APIs to extract data. The Wrapper Induction method uses optimum computing resources and allows each web page to be labeled as it includes techniques of machine learning. The Wrapper Maintenance approach is based on the Schema-Guided approach that, after the web page has been shifted, retains the hyperlinks, extracted data objects, and lexical features. Information Extraction (IE), the fifth approach, uses NLP (Natural Language Processing) algorithms to automatically extract structured information from unstructured data.

Web structure mining focuses on a website's whole network of links. There may be links between sections of the same webpage or between the webpages of various websites. The two possible categories for web structure mining are document structure and hyperlink structure. Document structure refers to how HTML and XML use different tags to organize the same web page structure. Different sections of a single website or web pages from other websites are connected via link structures [18].

The main topics of discussion center on three types of ranking algorithms. Content-based page ranking is the first class of algorithms based on the content of online sites. Web structure-based page ranking algorithms are the second category of the algorithm, which uses the link structure of the global web, while the third category employed a combination of the first and second categories. Ranking systems heavily rely on web mining techniques, but due to incorrect data, a lack of mining tools, and other difficulties in classification and clustering approaches, web mining has some problems that need to be resolved [19].

3. WEB CONTENT MINING

In search engines, web content mining plays an important role and refers to the extraction from web pages of relevant material, such as texts, photos, or videos [2]. Web Content Mining helps classify web documents into various categories, recognize web documents' topics, and locate related web pages across various web servers and applications related to personal relevance.

3.1. Web Content Mining Steps

Web Content Mining includes four primary steps are as follows.

3.1.1. Pre-Processing of web contents

The contents on the web pages are unstructured or semi-structured; this process will convert these documents into adequately structured data. Pre-Processing includes these steps, Extraction of text from web documents, Clean up the data by filling the missing values, removing and redundancy, Tokenizing, Stemming, removing Stop Words,

calculating CFT (Collection Frequency Term), calculating DFT (Document Frequency Term), and Bag of words. After pre-processing, the machine learning (ML) algorithms can apply as per requirement, usually classification and clustering.

3.1.2. Classification

The class or category to which a new web document belongs from a set of predefined classes or categories is established. It uses algorithms to categorize the new data using the vector term.

3.1.3. Clustering

Grouping web documents with similar characteristics, using no predefined understanding of what the groups should be, and calculating the most common similarity between two web document vectors using a dot product.

3.1.4. Identifying the associations

It uses association rules to identify the correlation between web pages that occur mostly together. Content mining uses two types of approaches Agent-Based and Database. The Agent-Based approach has three types: (i) Intelligent Search Agents, (ii) Information Filtering/Categorization, and (iii) Personalized web. On the other hand, the Database approach involves structured data containing schemas and attributes; there are two main types of this approach: Multi-level databases and Web query systems.

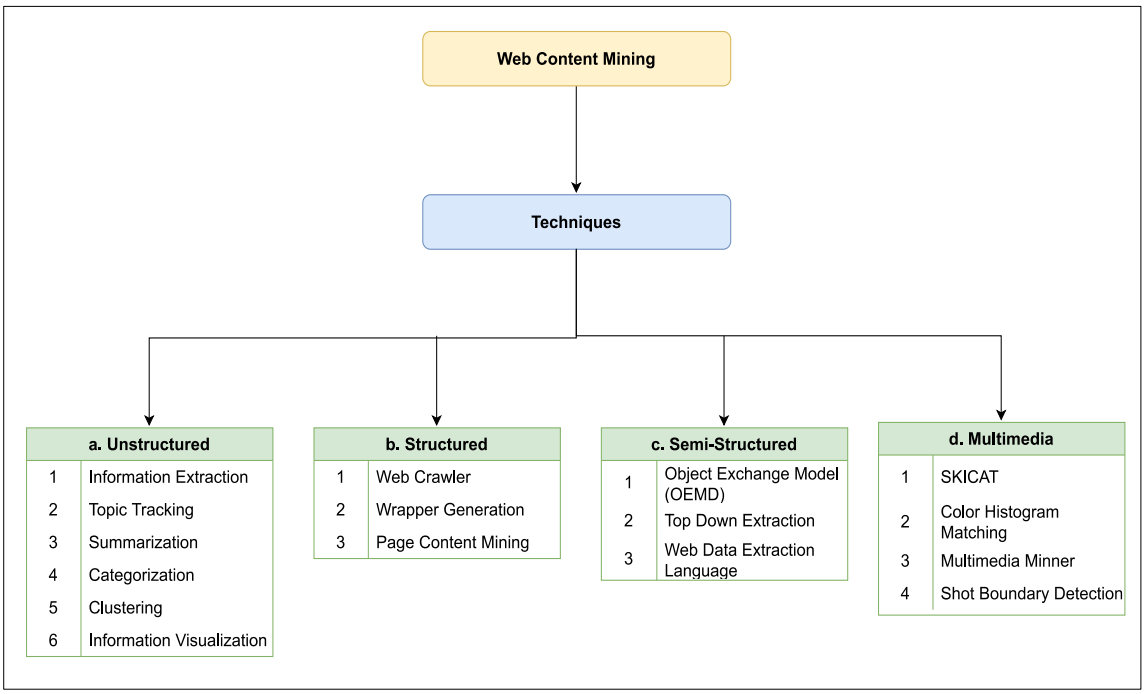

Fig 2. Web Content Mining Techniques

3.2. Web Content Mining Techniques

Web content mining techniques spread over four data mining styles, such as Unstructured, Structured, Semi-Structured, and Multimedia as shown in figure 2

3.2.1. Unstructured Data Mining

Content Mining` can be performed on unstructured data like email. The mining of unstructured data provides anonymous information [2]. There are various techniques to perform mining on unstructured data, such as Data/Information Extraction, Topic tracking, Clustering, Information Visualization, and Summarization [6].

3.2.2. Structured Data Mining

Structured data are highly ordered data. Within a file or record, it may refer to data. The level of organization is such that basic algorithms and searches are smooth and searchable for integration into databases [7]. Structured Data is often categorized as quantitative data, such as a relational database. It includes three techniques: Web Crawler, Wrapper Generation, and Page Content Mining [2, 3]. In this study, we explore these techniques in detail in Section 4.

3.2.3. Semi-Structured Data Mining

Semi-structured data is a type of structured data that does not follow the rules of relational databases or other types of data tables but still can be tagged with some other textual feature. It is also known as a form that defines itself [8]. It consists of three techniques, such as OEM (Object Extraction Model), Web Data Extraction Language, and Top-down Extraction [2, 3].

3.2.4. Multimedia Data Mining

It is a subfield of data mining used to find impressive implicit knowledge from multimedia databases. Two or more multimedia data types are needed to mine data, such as text and video or text video and audio. Mining multimedia is a kind of automated annotation [9]. Four techniques are used in multimedia data mining, such as Color Histogram Matching, Multimedia Miner, Shot Boundary Detection, and SKICAT.

4. TECHNIQUES FOR STRUCTURED DATA MINING

Structured Data Mining consists of three techniques, as in Section 3.2; in this Section, we explore these techniques and their tools.

4.1. Web Content Mining Steps

The web crawler will go through the details on the web page on its own. The web crawler's most crucial function is to crawl the web data, find successful information, and store the local database's necessary information data [10]. Web Crawler is also known as Spider, Spiderbot, or in-short Crawler. It is very resource-intensive to run and sometimes visits sites without permission [11]. Figure 3 depicts the Architecture of the Web Crawler.

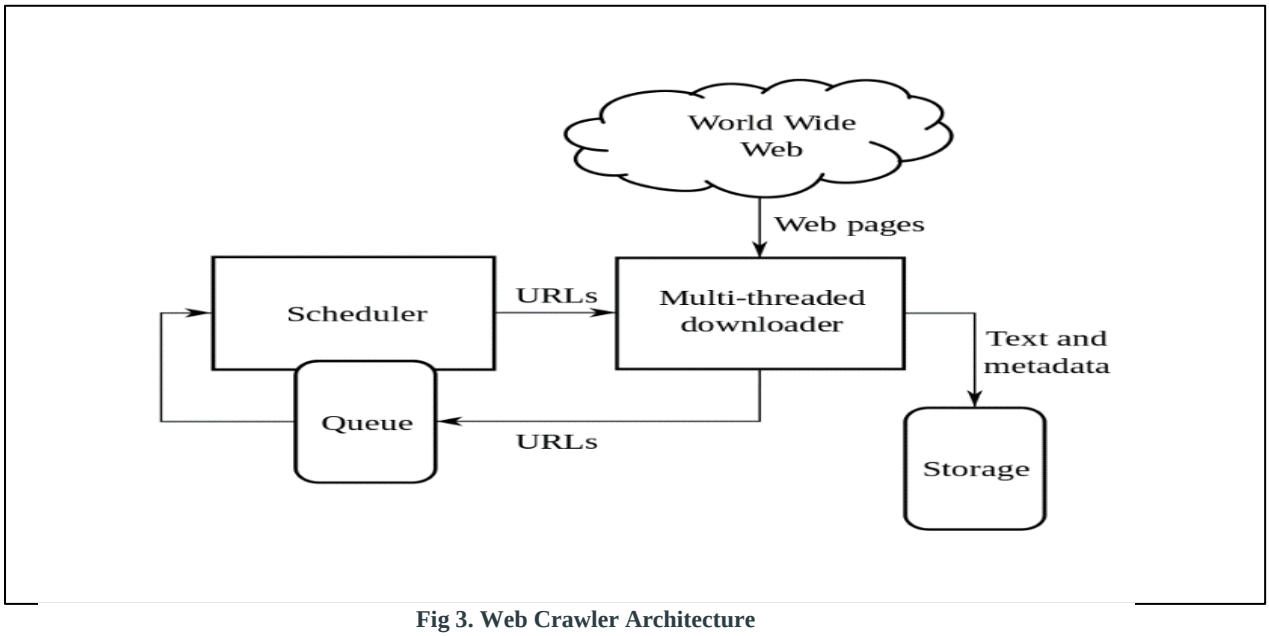

Fig 3. Web Crawler Architecture

Four steps, Scheduler, Queue, Downloader, and Storage, are part of WebCrawler. It starts with the Scheduler that provides the Seeds (list of URLs) to download. The downloader retrieves the page information from the Web and sends it to the data extractor, which detects all hyperlinks and adds them to the next URLs to visit. These next URLs are waiting for their turn in the Queue, where the Queue is busy filtering and sorting operations on the list and waiting for calls from the Scheduler [3, 10 – 11].

4.2. Crawler Classification

It is divided into six classes based on various characteristics.

4.2.1. Generic Web Crawlers

It fetches and stores all web documents and links to the topic on the hard disk, which consumes a great deal of space on the hard disk and network resources. One example of Generic Web Crawlers is Google's PageRank algorithm [10 – 12].

4.2.2. Focused Web Crawlers

Contrary to general web crawlers, it concentrates only on the specified websites and consumes fewer hardware and network resources than generic web crawlers [10 – 13]. The web page filtering module has two modules.

i. ***Web Judgement Module***: a high-quality arbitration module that compares the web page content with the specified topic.

ii. ***URL-Link Priority Ranking Module***: it compares the URL's quality with the topic and prioritizes them according to authentication and the number of citations of URLs.

4.2.3. Incremental Web Crawlers

It marks the existing collection based on the previous pages. It only updates the data of the expired web pages and replaces them with new pages. This approach makes it faster and utilizes a minimum amount of hardware resources than 'Focusing Web Crawler,' as well as significate improvement with the rate of data update. Incremental Web Crawler is responsible for the three types of cases, such as (i) websites with new pages, (ii) websites where page content will be updated, and (iii) deciding whether the content already exists when writing to the storage medium [10 – 14].

4.2.4. Distributed Crawlers

As the name suggests distributed, it executes on the different computers (nodes), and each node executes the focused crawler. The main problem is the coordination of each node's work to avoid work duplication [10]. Distributed Crawlers are divided into three modes.

i. ***Master-Slave Mode***: in this Mode, one computer would serve as a host, and other computers connected to the host will be nodes, much like a hierarchy, with one root, and multiple nodes. The host machine has to arrange the seeds (list of URLs), maintain communication with each node, disperse the task for crawling, and control each node's activities to counter an abnormality. There is no need for nodes to communicate with each other. Each node was expected to perform the tasks assigned and report the results to the host.

ii. ***Autonomous Mode***: there are two forms of it. In the first form, machines do one-way communication in a circular structure, and in the second form is the Unicom; each machine requires to communicate with each other to form a network, which is a complex structure. There is no host and salve situation as was in the previous Mode.

iii. ***Mixed Mode***: it is a combination of both Autonomous and Master-Salve Modes. Mixed Mode has a host machine and nodes, but each node can communicate with the others and have an assignment task function. If one node fails to perform a particular task, the host will be responsible for assigning it to the new node.

4.2.5. Parallel Crawlers

It is a time-consuming job for a single crawler to fetch all data from the Web; therefore, a parallel crawler reduces the time and speediest the process of crawling by running concurrently. It supports either embedded or distributed crawling; UbiCrawler is the best example of distributed crawler [10].

4.2.6. IoT Web Crawlers

The whole world is connected to the devices, such as computers, laptops, mobile phones, tablets, embedded devices, and sensors. Therefore web crawling is also introduced for IoT devices. The best example is the Shodan search engine, specifically designed for IoT devices, whereas Google performs Web Crawling for websites. Shodan fetches the data from various ports, such as RTSP, SIP, SNMP, HTTP, and FTP [10].

4.3. Wrapper Generation

Many web pages use specific HTML templates to shape organized data, such as telephone directories, and product catalogs. The wrapper generation extracts the information from web page sources and transforms them into a relational form that is efficiently processable by machines. The web pages can be retrieved by a query, such as fetching the web pages according to their PageRank values in descending order, where the pages already are ranked by traditional search engines. Wrapper Generation has been divided into two types of approaches: Wrapper Induction and Wrapper Automated [2 – 10].

i. ***Wrapper Induction***: trains itself by the pre-labeled data using supervised learning algorithms. This approach's downside is that it takes a hard time for the manual labeling process and difficulty for the wrapper to retain its quality. The ShopBot, WIEN, SoftMealy, and STALKER are tools that use Wrapper Induction [10 – 15].

ii. ***Wrapper Automated:*** overcomes the Induction approach's drawback, as it uses unsupervised learning algorithms because most web pages are organized using fixed templates [10 – 15].

4.4. Page Content Mining

Page Content Mining is used to extract structured data that operates on pages ranked by conventional search engines. The pages are categorized by comparing the rank of pages [16].

5. TOOLS FOR STRUCTURED DATA MINING

There are various tools for Web Content Mining as listed in [2]. In this study, we extract only the tools used for Structured Data Mining. **Table 1** lists out some tools available for Structured Data Mining in aspects of Web Content Mining. ***KNIME (Konstanz Information Miner)*** is open source and free with quality to use with a Modular data exploration platform for Pre-processing, modeling, and data mining, and is mostly used for Data mining, Business Intelligence. ***RapidMiner*** is an open source and royalty-free software tool for text and data mining with comprises learning algorithms from WEKA and its main applications are predictive analysis, and statistical computing. ***ORANGE v3.27.1*** is also free and open source Data mining and Machine Learning techniques, data mining is done through Visual programming or Python Scripting and most suited Data Visualization, text mining, bioinformatics, and data analytics of domains. ***Scrapy v2.4.1*** is free, Open Source, written in Python efficiently used in Structured Data Extraction. ***Web Information Extractor*** is paid commercial tool, with the ability to extract structured and unstructured data that is appropriate for web content extraction.

Web Data Extractor is a commercial tool having the features to extract data as well as web content extraction. ***Mozenda*** is also a Commercial/profitable tool with a privileged Web Console appropriate for Data mining and managing data. ***Web Content Extractor*** is a commercial Tool, identified for crawling and web spiders, collecting data from password-protected sites commonly used in the dominion of real estate data, online auctions, and job-seeking. ***Screen Scraper*** is a commercial tool as well, having the capability to explore content from databases with the help of meta-search engines.

Table 1. Web Content Mining Tools

Tool	*Feature*	*Suitable for*
KNIME v4.3.0	Modular data exploration platform Pre-processing, modeling, data mining	Data mining, Business Intelligence
Rapidminer v9.8	Includes learning algorithms from WEKA	Predictive analysis, statistical computing
ORANGE v3.27.1	Data mining and Machine Learning techniques, data mining is done through Visual programming or Python scripting	Data Visualization, text mining, bioinformatics, data analytics
Scrapy v2.4.1	Open Source, written in Python	Structured Data Extraction
Web Information Extractor	Commercial Tool, Extraction of structured and unstructured data	web content extraction
Web Data Extractor	Commercial Tool Extraction of URLs	web content extraction
Mozenda	Commercial Tool, Web Console provided	mine and manage data
Web Content Extractor	Commercial Tool, known for crawling and web spiders, collect data from password-protected sites	real estate data, online auctions, job-seeking
Screen Scraper	Commercial Tool can search for content from databases	meta-search engines

Table 2 lists the open-source Web Crawling Tools that *Apache Nutch* has written in Java language with a highly extensible and flexible system along with dynamically scalable with Hadoop. *Heritrix* has also been written in Java and has an easy setup, with distributed crawls. *StormCrawler* is also backed with Java and has large-scale recursive crawls with low latency web crawling. *Scrapy* is written in Python programming language and has properties of broad crawling with easy setup procedure as well as active Community. *Web-Harvest* is written in Java and supported by the real scripting language. *MechanicalSoup* simulates human behavior and is extremely fast for scraping relatively simple websites also written in Python programming language. *Apify SDK* (Service Development Kit) is programmed in JavaScript Scrape with mass level and high performance.

Table 2. Web Crawler - Open Source Tools

Tool	Programming Language	Features
Apache Nutch	Java	- Highly extensible and Flexible system - Dynamically scalable with Hadoop
Heritrix	Java	- easy setup - distributed crawls
StormCrawler	Java	- large-scale recursive crawls - Low latency web crawling
Scrapy	Python	- broad crawling - Easy setup - Active Community
Web-Harvest	Java	- Real scripting languages supported
MechanicalSoup	Python	- simulate human behavior - Blazing fast for scraping relatively simple websites
Apify SDK	JavaScript	- Scrape with largescale and high performance
Jaunt	Java	- Comfortable interfacing with REST APIs
Node-crawler	JavaScript	- priorities for URL requests
PySpider	Python	- Powerful WebUI with a script editor, task monitor, project manager, and result viewer - Distributed architecture

Jaunt is also backed with Java programming language having friendly interfacing with REST (REpresentational State Transfer) APIs. *Node-crawler* is programmed in JavaScript with priorities for URL requests. *PySpider* has a powerful Web user interface with a script editor, task monitor, project manager, and result viewer along with Distributed architecture written in Java programming language.

6. CONCLUSION

We have presented Web mining techniques such as Web Content Mining, Structured Web Mining, and Web Usage Mining in this paper. Web Content Mining, discussed in Section III, was the objective of the paper. Web Content Mining is categorized into four forms of data mining: Unstructured Data Mining, Structured Mining, Semi-Structured Mining, and Multimedia Mining. This research aims to investigate the Structured Data Mining techniques and tools used. Three essential techniques, such as Web Crawler, Page Content Mining, and Wrapper Generation, come under Structure Data Mining. We addressed Content Mining tools in Table 1, and Table 2 represents explicitly the tools used for Web crawling at the end of this article.

FUTURE SCOPE

The future work involves analyzing the Web Crawling Tools using performance metrics, such as throughput, Performance Cost, Scale-up, Latency, Durability, and Concurrency through Serial and Parallel Web Content Mining algorithms.

References

[1] Shukla, R. K., Sharma, P., Samaiya, N., & Kherajani, M. (2020). WEB USAGE MINING-A Study of Web data pattern detecting methodologies and its applications in Data Mining. In 2nd International Conference on Data, Engineering and Applications (IDEA). 2020 2nd International Conference on Data, Engineering and Applications (IDEA). IEEE. https://doi.org/10.1109/idea49133.2020.9170690

[2] WIKIPEDIA: "Web mining," URL: https://en.wikipedia.org/wiki/Web_mining

[3] Faustina Johnson and Santosh Kumar Gupta, "Web Content Mining Techniques: A Survey," 2012, International Journal of Computer Application (0975-888), DOI: 10.5120/7236-0266.

[4] Saleh Mowla, Ishita Bedi, and Nisha P.Shetty, "A Study on Web Mining Tools and Techniques," 2017, published in Journal of Engineering and Applied Sciences; DOI: 10.36478/jeasci.2017.6135.6142.

[5] Vandana Shrivastava, "A Methodical Study of Web Crawler," 2018, published in International Journal of Engineering Research and Applications; DOI: 10.9790/9622-0811 01 0108

[6] Mohd Amir Bin Mohd Azir, and Kamsuriah Binti Ahmad, "Wrapper Approaches For Web Data Extraction: A Review," 2017 published in IEEE Conference, DOI: 10.1109/ICEEI.2017.8312458.

[7] Mohd Shoaib1, and Ashish K. Maurya, "Comparative Study of Different Web Mining Algorithms to Discover Knowledge on the Web," 2014 Conference Paper @ ResearchGate;

[8] SEOBILITY-WIKI: Structured Data; URL: https://www.seobility.net/en/wiki/Structured_Data.

[9] WIKIPEDIA: Semi-structured Data; URL: https://en.wikipedia.org/wiki/Semi-structured_data#cite_note-1

[10] Mylavarapu Kalyan Ram, M.Venkateswara Rao, and Challapalli Sujana, "An Overview on Multimedia Data Mining and Its Relevance Today," 2017, published in IJCST (International Journal of Computer Science Trends and Technology)-Vol. 5; URL: http://ijcstjournal.org/Vol5Issue3No1.html

[11] Lin Xuan Yu, Yeli Li, Qingtao ZengQingdaobbbong tSun, Yuning Bian and Wei He, "Summary of web crawler technology research," 2020, Journal of Physics: Conference Series – IOP Publishing; DOI: 10.1088/1742-6596/1449/1/012036

[12] WIKIPEDIA: Web crawler; URL: https://en.wikipedia.org/wiki/Web_crawler

[13] Lu Zhang, Zhan Bu, Zhiang Wu, and Jie Cao, "Distributed and generic web crawler for online information extraction," 2017 – IEEE, DOI: 10.1109/BESC.2016.7804487

[14] Anish Gupta and Priya Anand, "FOCUSED WEB CRAWLERS AND ITS APPROACHES," 2015 – IEEE, DOI: 10.1109/ABLAZE.2015.7154936

[15] Kevin S. McCurley, "Incremental Crawling," Google Research; URL: https://research.google.com/pubs/archive/34403.pdf

[16] WIKIPEDIA: Wrapper, URL: https://en.wikipedia.org/wiki/Wrapper_(data_mining)

[17] Andemariam Mebrahtu, and Balu Srinivasulu, "Web Content Mining Techniques and Tools," 2017 – IJSCMC; URL: https://www.ijcsmc.com/docs/papers/April2017/V6I4201725.pdf

[18] Kumar, S., & Kumar, R. (2021).” A Study on Different Aspects of Web Mining and Research Issues. In IOP Conference Series: Materials Science and Engineering” (Vol. 1022, Issue 1, p. 012018). IOP Publishing. https://doi.org/10.1088/1757-899x/1022/1/012018

[19] Sharma, P. S., Yadav, D., & Thakur, R. N. (2022). “Web Page Ranking Using Web Mining Techniques: A Comprehensive Survey”. In M. P. Kumar Reddy (Ed.), Mobile Information Systems (Vol. 2022, pp. 1–19). Hindawi Limited. https://doi.org/10.1155/2022/7519573

Authors’ Profiles

Mustafa Ali Bamboat earned a higher degree in Software Development (B.Sc. Hons) from the University of Huddersfield, the United Kingdom, in 2002. The major field of study is software development, data mining, and IoT. The Author has published the following articles, i) Performance of RDF Library of Java, C# and Python on Large RDF Models published in International Journal on Emerging Technologies Journal [Scopus index Y Category Journal], (ii) Analysis of Traditional Hard Disk Scheduling Algorithms published in Quaid-e-Awam University and Engineering and Technology Journal.

Ghulam Sarfaraz Khan was born in Karachi, Sindh; he got secondary education in Karachi and did a bachelor’s degree from the University of Karachi in 1999. The major field of study is networking, systems administration, data mining, and quality of experience in multimedia. The author has published the article, Analysis of Traditional Hard Disk Scheduling Algorithms published in Quaid-e-Awam University and Engineering and Technology Journal.

Naadiya Mirbahar received her BSIT degree from Sindh Agriculture University, Tandojam, Hyderabad, Pakistan, in 2011, and the M.S. degree in computer science from Isra University, Hyderabad, in 2014. She is pursuing her PhD in Computer Science. Since February 2021, she has been working as Cooperative Lecturer at Faculty of Computing Science and Information Technology, Benazir Bhutto Shaheed University Lyari, Karachi. Her research interests include data mining, machine learning, and data privacy and utility protection.

Sheeba Memon obtained his Ph.D. degree (2021) In Computer Science from Central South University, Changsha, Hunan, China. She also received the B.S. degree in IT from the Mehran University of Engineering and Technology, Pakistan, in 2011, and the M.S. degree in computer science from Isra University, Hyderabad, Pakistan, in 2015. She is currently working as Assistant Professor, Government College University, Hyderabad. Her research interests include data center networks, load balancing in DCNs, Congestion Control.