International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (Jan-2022)

https://dx.doi.org/10.12785/ijcds/110149

Web Information Extraction methods using Web Content Mining (WCM) for Webapplications

Raghavendra R[1] and Dr. Niranjanamurthy M[2]

[1]*Research Scholar, Department of MCA, Ramaiah Institute of Technology (Affiliated to Visvesvaraya Technological University, Jnana Sangama, Belgavi) Bangalore, 560054, India, ORCID:0000-0003-3538-2339*
[2]*Assistant Professor, Department of MCA, Ramaiah Institute of Technology (Affiliated to Visvesvaraya Technological University, Jnana Sangama, Belgavi) Bangalore, 560054, India*

Received 12 Jun. 2021, Revised 16 Nov. 2021, Accepted 4 Jan. 2022, Published 20 Jan. 2022

Abstract: In the digital world era, data was generated by humans and machines are huge in volume and have been accessed through websites on the internet platform. Most of the transactions happened on web product items, web news, and web advertisements. Web Information Extraction (WIE) is the technique where the information on websites is extracted accurately within a time using Web Content Mining (WCM) concept. Every second, new data has been generated in different locations and the contents of the websites have changed rapidly at various intervals during processing time. The live time and location of the data have changed each time when internet users processing web applications. So extracting the information from the web page or website is a challenging one with accuracy and latency on websites. Classic algorithms and data mining techniques are used to preprocess the generated data with a certain time but the validity of those has not been maintained on the web server. Perhaps, their special features have taken for doing extraction using web mining techniques. The recently advanced concepts such as Deep Learning with Recurrent Neural Networks (RNN) are used to perform Web Information Extraction on various websites over the large network by keeping hold of the data status at each second in memory while doing the processing. The technique Long Short-Term Memory (LSTM) is used to hold the status in intermediate memory then all generated data in web applications send this status to RNN for further classifications. Classification methods are used in Artificial Neural Networks (ANN), it would train the input data from the large network and segregate them based on the algorithms used by the user. Finally, the deep learning concept is combined with all recent trends with input models as an embedded layer. Social media information is up-to-date with its originality and validity also keeps track fully in larger networks by using this technique. This paper suggested the best methods to implement the web information extraction concepts in web content mining from different websites on larger clusters/networks using deep learning LSTM techniques.

Keywords: Web Content Mining (WCM), WIE (Web Information Extraction), RNN (Recurrent Neural Network), LSTM (Long – Short Term Memory), ANN (Artificial neural Network), web server

1. Introduction

A collection of information and related content has identified with a common domain name using the internet is called a web page. More number of web pages are linked or connected with a centralized server and provide contents to the users through the internet are called websites. The name of the server is called web server and the services provided by all the websites are called web services. Generally deriving the data from the websites using the internet is called mining and it will happen on all websites connected under different networks known as web mining[1]. To extract information from this web server is known as Web Information Extraction (WIE). This WIE has done in this web mining for extracting accurate data with low latency over the internet. WIE has done with several extraction algorithms and methods using data mining techniques for getting exact data over the network [2]. Web mining is used to retrieve the contents over the internet easily and used for optimization purposes on larger network applications. Web mining classifies 3 types of mining on websites on a large network. The content, structure, and usage of websites on different networks created a huge log file for authentication purposes[3]. To optimize the content of those web pages in real-time the factors such as search, results are used for mining concepts. The other mining concepts discussed the structure of the webpage and usage analysis of them simultaneously. Web content mining is the technique used to derive the information from the web pages in three ways namely web page content, based on search the web page content, and page contents as a result[4]. Web structure mining consists of link structures and URL classification on the websites using HTML tags.

E-mail address: raghuramesh88@gmail.com , niruhsd@gmail.com

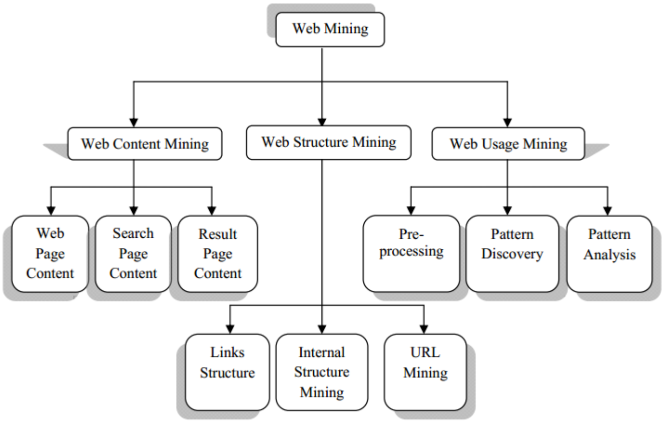

Figure 1. Types of web mining

Last web usage mining is the concept that is used to utilize the contents of the websites by preprocessing, pattern discovery, and pattern analysis. The optimization concept is to reduce the time taken and exact data-driven from the huge network plays a major role in web mining techniques[5]. The figure 1 denotes the types of mining on websites from huge networks.

Normally, Web pages have contained different forms of content like text, audio, video, animated, and gif format files in networks[6]. The contents which are created by users or machines would be processed as news, advertisements, and other forms of data that could be stored as a web page. The server gets the data from real-world entities continuously then it will be uploaded on the web page at regular intervals of time. As the information is integrated on web pages as contents in a network, the changes are dynamic and will be noted frequently[7]. Web mining concepts are used on a huge volume of data such as big data and other platforms for extracting the exact information from the different sources accurately with low latency. But the contents of the web pages are changed dynamically in nature in real-time and the changes that have been made being controlled are a big challenge for all[8]. Because the data all are used is varied at some locations. The figure 2 explains the nature of file formats used in web pages.

The figure 2 explains the file formats used in web mining techniques. In web content mining all the text, image, audio, video and, structured records are extracted from the data ware house. In web structure mining the hyperlinks and documents are extracted with the help of HTML and HTTP protocols as a collected documents on the same website. In web usage mining the log files of the web server, application server and application levels have extracted with web mining concepts[9]. Now a day, extracting information on web-based applications is a well-known technique among internet users. But huge data processing like big data as a web news/web advertisement of the product-item information is a challenging task in this modern world. The up-to-date information on the internet can access using media's through web applications[10]. While extracting the data from this huge volume will take more time and the accuracy of the data is very low due to classic approaches of mining principles. Their special features can be taken for further classification and it will process the mining techniques on web applications. The information collected for processing from different sources on the internet and their behaviors is stored as attributes[11]. The major problem is to find the location of the data where it was generated in mining is quite complicated to extract correct information. Web mining approached used in data mining techniques are processed by the characteristics and features of whole data which are stored in the data warehouse. While retrieving the data from the warehouse it could be generally difficult to get the exact one on a huge network. Challenging for all users is to extract the data from different sources in less time[12]. The following figure 3 denotes the approaches used in web mining

The figure 3 discussed the approaches the web mining techniques on the websites. Web content mining there are two main approaches are used namely agent and database. The agent-based approach is used to extract the information which is available in the local system and network system based on the agent verification techniques. Database approach is used to extract the information as a collection of database namely data warehouse or data repository with queries for structured formats and data mining algorithms used for other formats [13]. Web usage mining used pattern tracking and customized usage tracking for data extraction. When web structure mining is accessed their links have taken as the main input for file classifications either statically or dynamically. There are several algorithms and classification techniques are available for optimization but the time consumed and accuracy of extracted data is a major role for web mining researchers [14]. The people are using social media contents on the larger networks, got their exactly matched data after some time with some noisy output. Because proper extraction methods are not

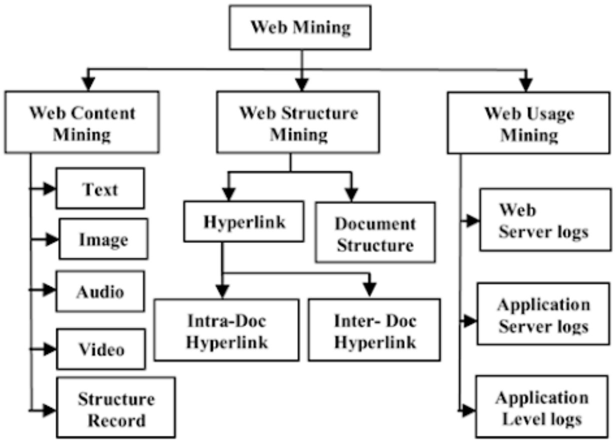

Figure 2. Web mining file formats

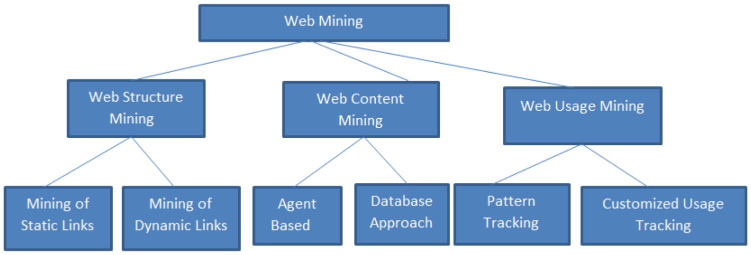

Figure 3. Approaches used in web content mining

available web mining concepts even though there are some classical methods are used. The huge volume of data is called unstructured which is normally no specific methods or formats in nature but it is very difficult to retrieve them from a big data warehouse [15]. But different techniques like information extraction, topic tracking, clusters, summarization, and virtualization are used to optimize their size as well as reducing extraction time on networks while extracting. The next type of data has its structure for classification namely; structured data and it would be derived by the concepts of web crawlers, web content generations, and data mining applications [16]. Semi-structured data represented by HTML, XML tags are extracted by Object Exchange Model (OEM) and XML tags, email contents are derived under top-down extraction method during real-time applications running time [17]. A specific language is used to retrieve the facts from the data repository is Web Data Extraction languages [18].

In this paper author has discussed the literature survey on section 2 and algorithms used for web content mining on Section 3, and then web information extraction approaches in web content mining on Section 4. Afterwards, web content mining algorithms used in previous approaches on section 5, then conclusion of WCM methods on and future work will be in Section 6.

2. LITERATURE SURVEY

Maghdid, H. S [2] proposed Machine learning algorithms or rule-based algorithms are performed to enhance the accuracy of data mining but time and space complexity or accuracy level is the challenged one in optimized web mining applications. Classical data mining approaches like KNN and ANN are used for web loggers but not suitable for keep track of data status in real-time applications. Various data types like text, audio, video, and other formats have approached the data mining process for extraction but the time taken is very high due to its size. LSTM is a technique that is the alternative of RNN used for storing data status and it will be monitored using deep learning concepts based on the geo-location of the data generation. But it is a big challenge to do with deep learning concepts.

Song J et al [3] proposed social media data has to be generated a lot and it will be changed due to the validity of time. To maintain the validity of data time it is going to be controlled by a centralized system with the help of data mining concepts. Also, CNN is a large network used to train the data sets with different computations for further classification. Classical approaches of classification techniques are used but it will extract the data with high delay. The latency of the optimization process takes high due to the size of the data on a larger network.

Moradi et al [4] proposed Support Vector Machine (SVM) and Page Rank algorithms are used to find the current status of the data generated in the large network but it will be monitored while getting changed. Also, the geo-location has not been identified by the system which will create issues in data mining techniques while performing preprocessing. All the monitoring concepts in the network have to do their job only when the data is in an active position.

Mulahuwaish et al [5] proposed Random Forest algorithm is used in data mining techniques to find the best values from the random one. But the data status has changed rapidly; the selection of random samples from the input is very difficult to understand. In this case, validity is a major point to address its functions. Random forest algorithm has selected the data based on the values not from the location or time generation period. So the status of the alive data monitoring is not possible over the larger network.

Umer et al [6] proposed Deep learning models are used to train the data set for geo-located data formats with the help of combined LSTM and RNN networks but their validity does not maintain all time due to its larger network. Various steps have been taken to overcome this issue but the accuracy cannot be maintained constantly. The level of accuracy varies depends on the size of the data taken as input. If larger data has to be processed time taken is very high due to traffic on a network. LSTM keeps track of the status of data in the network using gates for monitoring purposes.

Wang et al [7] proposed Convolution Neural networks (CNN) and Natural Language Processing (NLP) is the classical approaches for data mining and preprocessing works but their level of monitoring and control has not been done at a certain level. So the combination of these techniques will improve the level at a certain period in a small network. In larger networks, their accuracy is still not working properly due to the file size of the input. In a neural network, the trained data sets are given as input to the next levels from the hidden layer for better results. Hidden layer data have not been processed until the output of it will be taken as input for the next processed data.

Hu et al [8] proposed LSTM and CNN are the approaches that are working under data mining techniques but the status level of all data generated from different sources on a network is a challenging one for researchers. If collected, to maintain their status level all time is a challenging one using the predicted values shall not be accurate. The status of the data over the entire network is monitored and controlled at each stage during transmission. But the intermediate memory which is not located between the nodes of the cluster leads to create data loss or corruption problems frequently. To overcome this LSTM is a technique proposed to provide memory for short-term and long-term data as a result of it.

Hong et al [9] proposed LSTM on the stock market for predicting the level of data generated at a different level was the biggest challenge using CNN. The changes made in CNN and LSTM along with RNN network concepts have a big problem while creating trained data sets. CNN is used to provide the training sets of incoming data with less accurate results due to the changes made dynamically on websites. LSTM will overcome this problem with gates are used in this system for holding the data status continuously over the network.

Hong et al [10] proposed in health care services Privacy-free Data Fusion and Mining (PDFM) is used to find the status of the data can be calculated at the initial level and it will be stored in a separate memory. Duplicate data can be removed using Privacy-free Data Fusion and Mining (PDFM) methods based on the generation time. The duplication of data can be rectified by the above-said method and the extraction done at an accurate level in optimizing websites. Using this technique health care sector peoples are getting benefitted to secure their database at a huge level and extraction also done on time.

Shahrivari et al [11] proposed in big data using Map reduce the concept, the data input will be divided then it will do the preprocessing methods by the k-clique method. It will create small information about the data status on a whole large network and it will remove the duplication with the help of k-clique mining on web applications. The input data has split into small sectors to identify the data formats and it has taken for preprocessing methods. The K-clique method is used to generate the Metadata of the input information as a small snippet. There must be traffic or congestion that will occur due to the heavy input process at the preprocessing but the k-clique method will take care of merging all those files and accessed it for optimization.

Kumar et al [12] proposed to extract the features from the text classification can be done by the feature extraction algorithms. Some of the methods like Principal Component Analysis (PCA), Backward/ forward feature extraction, and random forest algorithms are used for extracting the text from the huge databases using classification principles. But the time taken to complete the process has not been done on time due to classification principles. To rectify this issue, the combined techniques like feature extraction and feature selection principles will help to reduce the dimensions of the vector space features.

Huang et al [13] proposed the feature extraction methods have developed with the help of Meta architectures used in CNN, RNN, and RCN neural network techniques. SSD (Single Shot Detector) is used to find the exact information from the huge database using the feature extraction technique effectively. But the task of extraction has not been completed on time due to the methods of SSD. To overcome this COCO detection task is used to find the exact match using their Metadata available on the network.

Anandhan et al [14] proposed Recommender Systems (RS) for detecting the feature extraction principles on a larger network. It uses data mining concepts like Association rules, KNN, K means and other principles to extract from the data warehouse. Collaborative Filtering (CF), Hybrid Collaborative Filtering (HB-CF), and Knowledge-Based CF are the new concepts used to get accurate information from the data warehouse on time. But the filtering concept used to get the information on the network within the stipulated time is a challenging one due to the dynamic changes on the websites.

Tang et al [15] proposed CPSS (Cyber-Physical Social System) method to optimize the web mining approach on websites and the algorithms used to perform this are LCS, N-Gram, Natkasu, and Lk analyzer. All are defined the feature extraction level of the CPSS system was implemented with page rank algorithm with the help of page ranking factors available on the websites. The page ranking factors have its guidelines to set their values over the network to improve the ranking of the websites in real-time.

Tan et al [16] proposed the method for extraction and accuracy level optimization based on DOM tree or Text density algorithm. But the accuracy level has not reached the value which drained the latency time on a huge network. MCSTD (Maximum Continuous Sum of Text Density) method is used to find the leverages of features extraction over the large network. This method is not giving the extraction data on time due to the text density. To overcome this problem TWCEM (Title Based Web Content Extracting Model) is used for retrieving the data on web pages. Title is declared in input has to be taken care of for further processing of feature extraction. So the level can be maintained and the extraction happened at right time.

Zhang et al [17] proposed to find the shortest path for getting more information from the huge network using sentimental analysis. Because the reviews of the customer about the product are collected from various sources and the size is very high. To get the exact information on that place it used the entity sentiment word pair extraction method. There are other methods such as direct trust computing and propagation trust computing is used to perform the sentimental analysis from the huge network on websites. To perform optimizing web mining in this concept sentimental analysis is the best method but the time to completion varied based on the inputs given by the user.

Si et al [18] proposed a method to find the large number of spammers who release the wrong data from the network and phishing websites. Spammers are used to removing pornographic contents and unwanted speech from the whole database warehouse. For running this method Spammer Identification Technology algorithm is used. To implement this LPA (Label Propagation Algorithm) is used to extract the information from a larger network on time. The accuracy level of this algorithm was not effectively executed by this approach due to the duplicate data integrated into the website frequently.

Ye et al [19] proposed the methods to find the web services discovery using the deep mining concepts. In this method, web services classification can be done by TF-IDF, LDA, and WE-LDA algorithms from the data warehouse. It will help to find the information which is hidden on the network can be easily rectified by deep mining concepts. The accuracy level of web discovered services is not reducing the time of mining on websites. So the implementation can be done by web service classification method and Bi-LSTM method. To hold and tracking of the data travel on the network can be monitored by this method and it will be helpful to optimize the web mining on the websites over large networks.

Da'u et al [20] proposed the model to find the optimization techniques using Review Based Analysis. Collaborative Filtering (CF), Matrix Factorization (MF), and CNN are used to perform the optimization on the network. This method is used to filter the contents from the huge networks using matrix factorization concepts like text contents are arranged in a row and column-wise to extract those on time. CNN is used to create trained data sets based on the output given by the hidden layer to the same layer for improving the better results. To implement this approach for getting good results it was used SDRA (Sentiment Aware Deep Recommender System with NN) and LSTM for holding data status throughout the network.

Xu et al [21] proposed a method to find extraction of reviews using sentimental analysis. TF-IDF algorithm is used to find out the vectors of the weighted word on bi-directional LSTM (Bi-LSTM) techniques to capture the context of the information available in the network. The extraction accuracy level was not increased due to the wrong calculations on the network and it would be done with the help of RNN, CNN, and NB method.

Li et al [22] proposed the methods to find the behavioral patterns and the operations of the mining on the websites. One to one and one to many relationships have failed in this method to deal with social media websites for mining. But Community Detection Information Diffusion and Topic detection Monitoring can be done by sentimental analysis to perform the pattern matching using web mining on larger networks. Opinion mining is also used to perform the sentimental analysis while larger data warehouse on the network. To overcome all the problems in this model Object Centric Behavioral Constraint Model is used to perform the optimization on the websites on huge data networks with the help of sentimental analysis.

Li et al [23] proposed a method to find the noisy question-answer pairs and their filters for irrelevant and incorrect data. The knowledge triples from the wrong question-pairs have been extracted based on high-quality knowledge and the time taken to complete the answers for the questions by the doctors is varied according to the nature of question pairs. To optimize the web content mining on networks in this medical images domain their repository extraction must be easy to get all the driven data at regular intervals to the doctors for avoiding the delay in answering the questions of the patient. For this approach or to overcome this problem Medical Knowledge Extraction (MKE) is used. It should be done by multiple linked truths and noisy input removals.

Bai et al [24] proposed the method to solve the web services recommendation problem by a long-tail approach. To do this process it used deep learning concepts as a joint encoder principle and feature representation techniques. A framework has been created for content-usage learning concepts using deep learning approaches. But the feature extraction principle was not performing well due to the recommendation problem of the reviews given by the user. To improve this method SDAE (Stacked denoising auto encoders) have developed for feature extraction using stack principles. But the incoming and delivered web page contents are changed dynamically in nature it reflects recommendation system problem over the networks. So the new concept DLSTR (Deep Learning for long-tail web recommendation system) has been developed to improve feature extraction on websites effectively. It has also not given accurate results among the websites on larger networks due to file sizes.

Zhang et al [25] proposed the methodologies for event mining based on the videos on the web. The number of videos has stored in data repositories and it was extracted by the users in real-time have not been processed effectively due to the size of the file. Video files repository takes enormous time for extraction during network transmission and can control by MCA (Multiple Correspondence Analysis) terminologies. Here it calculates the frame numbers of the video and stores it in a common index table then selected for consideration. The correlation concept has been used in this approach for betting exact videos using NDK (Near Duplicate Key Frames). The concept of this NDK is to find cross-correlation using high-level semantic analysis with the help of MCA to perform an extraction.

Imtiaz et al [26] proposed the concept of Google News Vector for duplicate question pair detection issues on a large network. It uses vector concepts to find the updated news contents from the web pages. The fast text crawl embedding concept is used to detect the dynamic contents on web mining reflected from various nodes. There are multiple algorithms are used to get these web mining concepts on the network but if the question pairs are more than a million the speed of the retrieval is delayed. To overcome this approach LSTM approach is used as the Manhattan distance has been calculated among the neural network model between the websites stored in a repository.

Liu et al [27] proposed a method to detect malicious websites on the networks using machine learning-based algorithms. It uses KNN, RNN, CNN, and SVM classifiers to perform web classification for creating the trained data sets from the input data. The trained data sets as a input to the ANN models for extracting the exact and accurate information from the websites. Naïve Bayes theorem and their concepts are used to get the feature extraction from the trained sets. In this approach, CNN is used to detect the malicious websites from the millions of the website repository based on the web classification on the trained data sets.

Uzun et al [28] proposed new strategies for the classification of irrelevant images from the image repository on networks. An error-prone process is used to perform the above problem and it will be done using the number of design variations taken from the web pages. It will give the exact images to the users through websites but the time delay is very high on large networks. The straightforward approach has been used to extract the features of the websites and the f-measure score can be calculated for finding the minimum, maximum score to update the features of the web pages dynamically. The classification concept giving wrong input because of the features consider from the web contents. To add a minimum of 15 features to that website for increasing the page ranking effectively. Ada Boost classifier is used to perform this by adding extra features to the websites randomly.

Cerqueira et al [29] proposed the techniques for finding the news on the websites and reading it without the users among all. To perform this fuzzy systems are used to get the changes made on websites in real-time. Rule-based textual databases are used to find the news on the websites and rectify if any issues on the websites without the user's presence. For selecting the news and advertisement over the network instance selection from the multi-objectives is used based on Genetic Algorithms. Multiple objectives have been taken as a feature of the website and it will be considered as a page ranking factor during search engine optimization.

Es-Sabery et al [30] proposed the methodology to driven the emotions and opinions from the posted data on social media networks by the users using sentimental analysis. This will be used to find the exact output from the reviews given by the reviewers on social media platforms for product surveys or product selection. But there are a lot of reviews given by the users it would be difficult to find an exact match from the sets so that opinion mining and sentimental analysis are used. The entire process is done by linguistic processing and some Natural Language Processing (NLP) principles with their common features. The problems in this approach are rectified by FDLC (Fuzzy Deep Learning Classifiers) using CNN to find the features of the web pages in web mining. MFS (Mamdhani Fuzzy Systems) is used to measure the small changes in the website to classify the website dynamically from remote places.

Mirtsch et al [31] proposed the techniques for the ISO/IEC standards were not given the expected growth rate of page ranking factors. The standards like ISO/IEC 27000 for improving the growth of web content mining on large networks have not given the web page growth rate concerning the presence of data on web pages. TOE (Technological Organizational Environmental) Framework is helped to create databases with multiple objectives which are created by the users. The results given by these standards are not fixed because the status of the data is not updated regularly in the repository. It is integrated with web page contents only at the end of the repository configuration until that it will not store all the changes.

Ma et al [32] proposed the concepts of the cyber security entities can be extracted from the unstructured text and it has a lot of challenges due to the structure of the data is unstable. NER (Named Entity Recognition) is used to perform these operations in cyber security along with Bi-LSTM concepts. These techniques are used to keep hold of the status of the data in the entire transmission stage of the network so that intruders are not disturbed the network system often. CRF (Conditional Random Field) is atopic used to select the features of the web pages based on their contents which are created by the users. It is especially used in unstructured text because the formats of the files which are loaded on the website are very high. To develop this with extended Bi-LSTM techniques gives word embedding layer for feature extraction and CRF inputs have to be accessed twice for getting better results over the web sites.

Liu et al [33] proposed the technology for extracting the information from the news or headlines on the web pages. Based on this technique it may help to make the decision and assist to extract the fields on web pages. Bi-LSTM is the technique used in this approach to extract the content features and feature expressions through the self-attention mechanism. The weight of the network must be adjusted using the model archive accurate classification method to avoid fault detection or wrong decision taken on the web classifications. LSTM keeps tracking all the data status into the live-action and if any changes have occurred it will be monitored and sent to the controller on a remote network.

Long et al [34] proposed the methodology to find the text detection and recognition using deep learning algorithms. The insights of the benchmark repository or website directory extraction have not developed and optimized properly due to the insufficient trends are used in current scenarios. The identification of imperfect imaging features and uncontrolled conditions are taken as challenges using these techniques and it will be solved using detection and recognition methods. MSER (Maximally Stable External Regions) is helped to extract the textures from the image and SWT Stroke Width Transform) is a technique used to recover the consistency of the changes on the network continuously. RNN based deep learning network model is used to predict the features of the images on large clusters.

Pandey et al [35] proposed the techniques of text extraction from the scene images using adaptive galactic swarm optimization with the hybrid deep neural network. The text identification and recognition on those images are much challenged and will not work for huge datasets. So WNBC (Weighted Naïve Bayes Classifier) is used to perform these operations using deep learning concepts. The metrics for improving the performance of the entire system like accuracy, F-measure score, mean square error, and mean absolute error have to be calculated with a deep learning algorithm and classifiers then it will use the SWT(stroke width transform techniques) evaluated to find the addictiveness of the whole system.

MBWS (Marker-based Watershed Segmentation) is used to enhance the texture contrast on websites for an easy data-driven process. A fuzzy system also includes in this approach to get better results with accurate data.

3. PROBLEM STATEMENT AND ALGORITHMS USED FOR WCM

Web content mining has done optimization using several techniques for the different types of data. All the text, audio, video, and graphics files are working under one umbrella named multimedia contents and that will be extracted using their textures such as histogram matching concepts, shot boundary detection, and SKICAT tools effectively. The size of the data used here is huge so modern tools are used for extracting accurate data[19]. The following figure 4 describes the techniques used in web content mining concepts.

Web mining has been done with different concepts but accurate results will be given by certain techniques in the research fields. First, it has to separate the input data into a group of data then it will be accessed for extraction or optimization. Finally, clustering, classification, and association features of web mining techniques are used to ensure the extraction on websites can be done [20]. Classical approaches like the k-means algorithm, KNN, and other data mining approaches have given the clustering techniques of the nodes and the path of the data but not the validity time of data. But the later techniques such as artificial neural networks and deep learning are used to find the status of the data at each state[22]. The location of data with its generation time can be stored in short term memory for next-level processing. The various levels of networks trained the data with different locations and times of data then it will be stored in a memory for the next step.

To manage web news, an advertisement as a web page contents online, data extraction within a stipulated time is not possible in real-time. Because when the internet

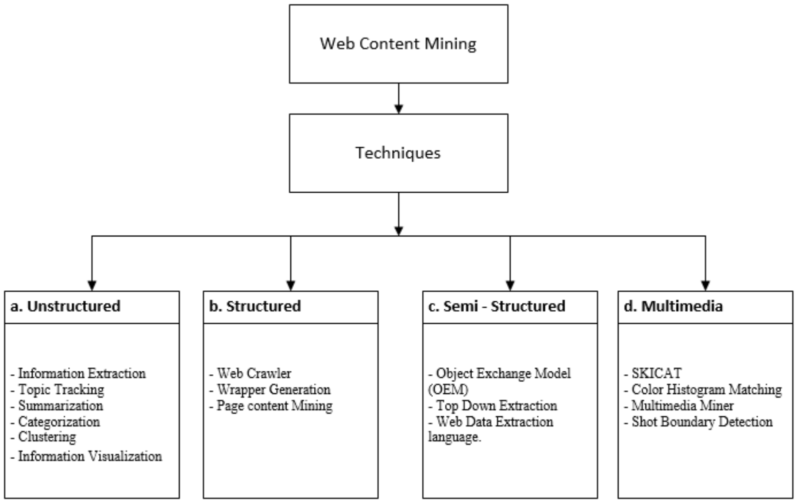

Figure 4. Techniques in web content mining

data will be updated frequently, identify those data is a challenging one. It also drains the network and power of the smart devices during extraction[22]. Generally, the raw information has been entered into the selection process by the user. For that, the data will be divided into several subsets and partitions to initiate the extraction process. The next step to select the appropriate data is to preprocessing the incoming data on a larger network for optimization. The preprocessed data is to change with different transformations as per requests given by the users. Finally, data mining concepts are used to extract accurate information based on their patterns on the larger network to optimize from the huge data warehouse. Then the data must be evaluated for interpretation and it will be converted into knowledge. The entire process has been done on the larger networks dynamically with the help of classical approaches like data mining and other algorithms [23]. The figure 5 describes the steps involved in data extraction.

Web content mining has used multiple methods to extract information from the huge database. These algorithms used classification techniques for fetching the information as knowledge and used to extract the web contents. The algorithms[[24] - [27]] used for classification on websites for web content mining are listed below

- Decision Tree (DT)
- Naïve Bayes (NB)
- Support Vector Machine (SVM)
- Neural Network (NN)
- Convolution Neural Network(CNN)
- Recurrent Neural Network(RNN)
- Deep Learning (DL)

A. Decision Tree

It is a classification approach used for web content mining which has a root node and leaf node as branches in a tree structure. Based on the root node data all the data will be divided into sub-leaf nodes and hierarchy can be maintained by the nodes while extracting data from huge databases.

B. Naïve Bayes

Based on Bayes Theorem a classifier has created for predefined data set values on the network called Naïve Bayes classifier. It is the most commanding algorithm used for counting probabilities of occurrences as a combination between the values in websites.

C. Support Vector Machine

A simple algorithm with machine learning concepts and classification techniques but especially used for Linear and nonlinear data sets on the networks. Separate classification features have been taken as decision boundaries for selecting the values on a hyper plane.

D. Neural Network

This algorithm used in the web content mining approach using multiple layers such as input , hidden and output. Each layer gives its output to the next layer input for classification purposes. Hidden layers are used to perform data extraction purposes by various data mining algorithm techniques such as association rule mining, preprocessing and data cleaning After all the steps are done by this layer the output will be given to the classifiers as a training data set.

E. Convolution Neural Network

Convolution Neural Network is used to classify the input stream of data into different levels and it will be accessed in a hidden layer. After classification, it denotes the data features for preprocessing by data mining techniques. The

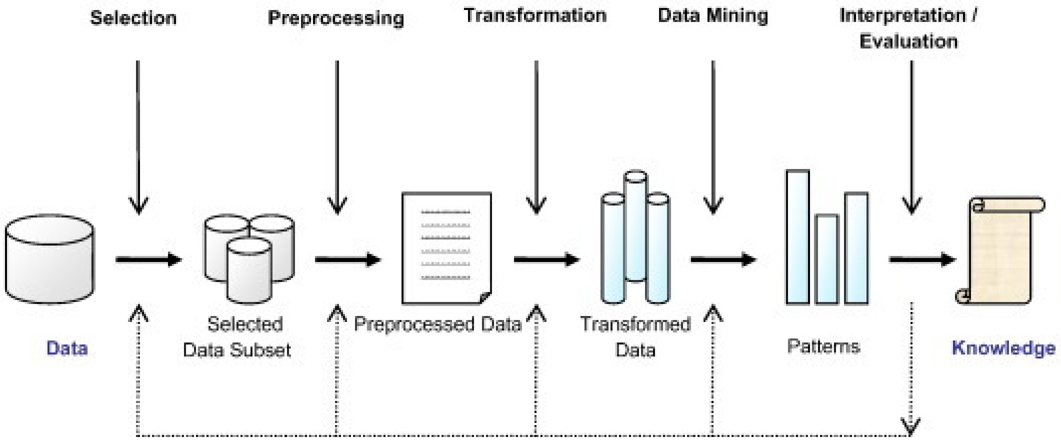

Figure 5. Steps in data extraction

next level RNN is used to rectify at which state the data has modified/changed in the large network. These are separately implemented and compared with time, space complexity for classification accuracy.

F. Recurrent Neural Network

The classic approaches of data mining techniques such as the decision tree and naïve bayes algorithms are used to extract information from web applications at a stipulated time with lower throughput due to its size and geo-location of data. To overcome this problem, the previous location and time of data information have to be stored in every stage with the help of a Recurrent Neural Network (RNN). It has the own internal memory for storing the present state of data and it is worked under a feed-forward neural network concept.

RNN has three layers such an input, an output, and a hidden layer for accessing of data. The main role of RNN is the output of the current input depends on the previous computation. The output will be copied and sent to the recurrent network system for processing. While deciding on this technique like the current input and output have to be learned from the previous input. It is very difficult in RNN to create a training set because of a lot of data states on internet web-based applications. The output of the hidden layer data must be given as input to the same layer iteratively for getting better results.

The figure 6 explains the detailed view of Recurrent Neural Network architecture.

Deep learning algorithms are used to do classify with the help of several text representation techniques like one-dimensional, two/three-dimensional and other format inputs. For one-dimensional format inputs DBN (Deep belief Network), DAE (De-noising auto encoders) and SDA (Stacked De-noising Auto encoders) techniques are used for data representation through data learning from the trained data sets.

In Deep Convolution Neural Network (DCNN) architecture the representation of two / three dimensional input data on the websites has to be classified with the help of Alex Net, VGGNET, and Google NET and ResNet techniques. The figure 7 explains the techniques used for data representation in web content mining.

Long Short-Term Memory (LSTM) is a revised style of RNN, which makes to remember the previous data in the memory. LSTM is used to do the classification of web-based application data and to predict the time series of data from the given time lags of various situations. Back-propagation is used to train this network for data processing. Occasionally, data corruption and duplications also happened because of hackers and intruders on the large network but the geo-location of data is used to identify those fraudulent issues.

The sequential data can be handled by LSTM but the size of data is very high and it may take lot of time to generate the sequences as data states. So Recurrent Neural Networks (RNN) combined with the LSTM model and deep learning concepts to overcome the above-said problem. The major problem raised in mining on web applications is when a huge size of data is created at different places and various times. To find these changes as geo-location of data and generation time of those data has to be done with LSTM, RNN, and Deep learning concepts.

Deep learning is used to learn the network with various trained data sets at a different time from the neural work output layer data. The combined concepts of RNN, LSTM, and deep learning concepts will be implemented in this research to avoid time delay and fake news generation on web-based applications. The working memory of RNN has to be maintained input and output of the processed data whereas LSTM has long-short term memory unit for handling the changes occurred in network. The data sets of geo location based election results tweets and ip address based accessed nodes have taken for experimental purpose.

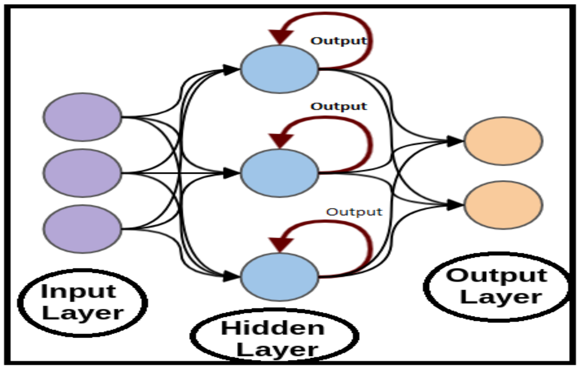

Figure 6. RNN architecture

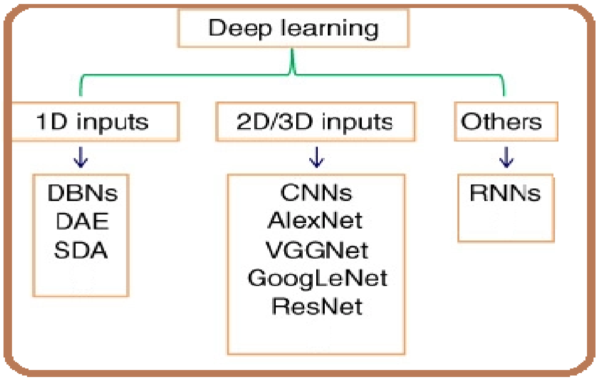

Figure 7. Deep Learning Architectures for Web Mining

The following figure 8 will explain the concept of LSTM based neural network architecture and how it is compared with RNN architecture.

The table I summarizes the algorithms used for web content mining (WCM) used in web mining approaches

Types	Data Types for mining	Modern Tools Utilized	Classification Algorithms Used
WCM	Unstructured, Structured and Semi-Structured Data Multimedia contents	Screen Scaper, Mozenda, Web Information Extrac-tor(WIE), Web Content Extractor, Automation Anywhere7	Decision Tree, Support Vector Machine, Naïve Bayes, Neural Network, CNN, RNN, DCNN, Deep Learning

TABLE I. WCM Summarization Table

The figure 9 summarizes the algorithms used for web content mining (WCM) used in web mining approaches.

4. WEB INFORMATION EXTRACTION (WIE) APPROACHES

Web Information extraction (WIE) plays a major role in we mining where the contents of the various web pages have extracted from different web sites on certain time. The accuracy and latency of those process varied when it was used some approaches. It is classified in to three types namely rule based, feature based and title based[[28] - [30]].

Rule Based WIE is used to extract the information from the websites based on the rules given by the certain methods like hand craft, supervised, semi-supervised and unsupervised methods. But their processing time and cost of the implementation is very high due to the pre described rules were used and large scale annotations. If millions of websites are in the repository then prior knowledge about the constrains are very important using XWRAP approach. These are the disadvantages of the rule based WIE approach. The other methodology W4F, OLERA, WHISK, and WDEPTA is used for information extracting purpose by rule based WIE.

Feature based WIE is working with Feature Extraction and Feature Selection principles with various new methods like DOM based, vision based and blocking based. All are working effectively but their time, memory space complex-

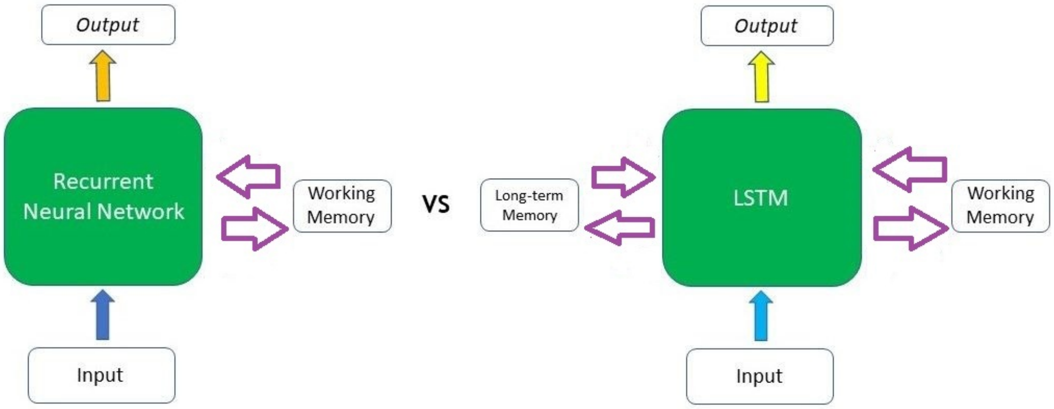

Figure 8. Difference between LSTM with RNN Architecture

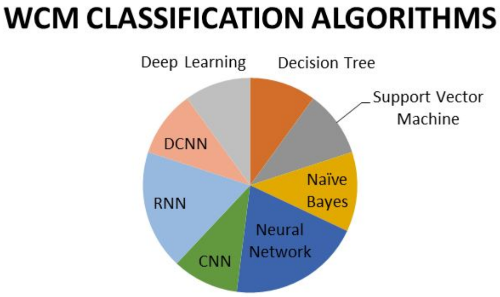

Figure 9. Algorithms used in Web content mining

S No	Classification Type	Methods	WIE Systems used	Disadvantages
1	Rule Based WIE	Hand Crafted WIE	Web-OQL, W4F, WRAP	High processing time and cost
		Supervised WIE	WHISK, DEByE, Soft Mealy, SRV, PPWIE	Large scale annotations
		Semi-Supervised WIE	IEPAD, OLERA, Thresher	Prior knowledge constrains
		Unsupervised WIE	Roadrunner, EXALG, DeLa, DEPTA, NET	More than million websites it is not possible
2	Featured Based WIE	DOM based WIE	HTML tags	High time complexity due to tree traversal
		Vision Based WIE	Page Segmentations	Huge Memory space complexities
		Blocking based WIE	CETR, CEPR, MCSTD	Block distribution has more changing points
		Other WIE	CETD, CLG, CCB	Missing contents due to dynamic changes
3	Title Based WIE	HTML Parsing	Regular Expressions	Noisy pages
		Title Extraction	Hyper Link, HTML,Content,TWCEM, LCS algorithm	Different languages as a titles

TABLE II. Web Information Extraction Techniques

ities with blocking distribution and their dynamic changes of the missing contents are the main disadvantages of the Feature based WIE. CETR and CETD are used to perform blocking the unwanted information while performing WIE.

Last one Title based WIE provides the better results when compared with the previous two approaches because their title extraction have done using HTML parsing which supports exact information driven from the websites. But the noisy pages from the different websites and lot of languages are used in the websites were created wrong information retrieval from the websites. LCS algorithm and TWCEM are also used to perform title extraction.

The table II differentiates various web information extraction methods and their system used with disadvantages.

5. WEB CONTENT MINING ALGORTIHMS APPROACHES USED

Different methodologies and algorithms used in Web content mining approach for performing WIE operations on a larger network websites in internet.

The table III denotes the techniques used for optimization approaches on web content mining taken from the literature survey from section 2. It shows all the methodologies used in WIE from [2] to [35].

A. *AUTHOR'S CONTRIBUTION*

To summarize the contribution for this paper, the authors are explained the web mining optimization approach as

S No	Research Gaps	Methodology Used	Solutions
1	Feature Extraction	Meta architectures, RCNN R FCN, SSD(Single shot Detector)	COCO detection task
2	Recommender System	Data Mining concepts KNN	CBF(Content Based Filtering), CollaborativeFiltering, (CF), hybrid-based (HB) Filtering, and Knowledge based (KB) Filtering
3	Cyber Physical Social Systems (CPSSs)	LCS algorithm , N-Gram algorithm, Nakatsu algorithm.	Page Rank Algorithm
4	Extraction and accuracy	DOM-tree-based Text-density-based MCSTD	Title-based web content extracting model TWCEM
5	Shortest path for more information	Sentimental analysis	Entity-sentiment word pairs extraction Sentiment similarity Direct trust Computing Propagation trust computing
6	Wrong speech, pornographic materials, and phishing websites	Spammer Identification Technology	Elm algorithm Label propagation algorithm (LPA)
7	Web services discovery	TF-IDF, LDA, WE-LDA, Deep mining of hidden information	Wide & Bi-LSTM model
8	Review based analysis	Collaborative Filtering (CF) Matrix factorization (MF) Convolution Neural Network (CNN)	Sentiment-aware deep recommender system with neural attention network (SDRA) Long short term memory (LSTM)
9	Sentimental analysis for extraction of reviews	TF-IDF algorithm and generates weighted word vectors Bidirectional long short term memory (BiLSTM) to capture the context information	Sentiment analysis methods of RNN, CNN, LSTM, and NB
10	Behavior patterns and operational processes. One-to-many and many-to-many relations	Community detection, Information diffusion, Topic detection and monitoring sentimental analysis opinion mining.	Object-centric behavioral constraint models
11	To drive the emotions and attitude on social media platforms	Linguistic terms Natural Language Processing (NLP).	Fuzzy Deep Learning Classifier (FDLC) Convolution Neural Network (CNN) Feed forward Neural Network (FFNN) Mamdani Fuzzy System (MFS) deep learning models
12	ISO/IEC 27001 standards expected growth rate	ISO/IEC 27001	TOE framework Technological Organizational Environmental German firms in the MUP database
13	Extracting unstructured texts from the cyber security entities is a critical and fundamental task	Named Entity Recognition (NER) Bidirectional Long Short-Term Memory with Conditional Random Fields (Bi-LSTM with CRF) to extract security-related concepts and entities from unstructured text	XBiLSTM-CRF, consists of a word-embedding Layer, a bidirectional LSTM layer, and a CRF layer, and concatenates X input with bidirectional LSTM output.
14	Decision-making of Information extraction Technology on news headlines field to assist	VGGNet instead of ZFNet	Bi-LSTM
15	Extracting the features of text sentences for text classification	Feature extraction and algorithms. Bag-of-Words model Principal component analysis(PCA) Random forest Backward feature elimination Forward feature construction	To reduce the dimensions of the feature vector space, combine both feature extraction and feature selection
16	Noisy Question-answer pairs and incorrect information.	High-quality knowledge triples	Medical Knowledge Extraction (MKE) Multiple Linked Truths Noisy input
17	Problem of Long-tail web service recommendation	Deep learning joint auto encoder based feature representation content-usage learning framework	Stacked de-noising auto encoders (SDAE) to perform feature extraction Deep learning for Long-Tail web Service Recommendations (DLTSR).
18	Web videos based on event mining	Multiple Correspondence Analysis (MCA)	Near-Duplicate Key frames (NDKs), and semantic cross-correlation
19	Issue of duplicate questions	Fast Text crawl embedding Google News Vector	MaLSTM ("Ma" for Manhattan distance) Neural Network model 100000 pairs of questions
20	Detect malicious websites	Machine learning-based detection algorithms CNN & RNN SVM Naïve Bayes KNN	CNN constructed complex dataset
21	Classification of irrelevant images	Straightforward approach f-Measure score	AdaBoost Classifier with adding 15 new features
22	To find and read all the news without the user	Fuzzy Systems Rule base textual databases	Instance selection Multi objective Genetic Algorithms

TABLE III. Research Gaps and Solutions

following points:

- Author has taken various points from many recent research papers on the topic of web content mining optimization approach, all papers are discussed about the extraction of the information from the huge data warehouse with high accuracy on time over large networks.
- Author has written this paper about web content mining based on deep learning and neural network algorithms with the examples and have comparisons of all previous methods were used for web mining approach.
- Author has suggested for the optimization of web content mining approach using LSTM method. It will help in future for those who are doing research in web content mining information extraction

6. CONCLUSION AND FUTURE WORK

This paper helps fake news spreading and data duplication on social media, further that can be monitoring based on the avail time and location of the data has generated in the online continuously. The larger network data extractions make delay because of huge volume data can be preprocessed and waiting for a long time to get exact data. Due to classification techniques and the data segregation process done quickly by mining algorithms; delay would be avoided in the cloud, big data, and IoT frameworks. The accuracy of exact data delivery will be proven in web applications for web product items and web news along with web advertisements. Facebook, Twitter, and YouTube are the famous social media platforms and their content distribution will be optimized and secured due to geolocation and time duration of generated data. This paper used to perform web extraction and web mining optimization in real-time applications, and the tracking from different sources can accessed sequentially or dynamically by deep learning concepts. Unauthorized entry and intruders hacking problems also be solved because of the geolocation of the data and its validity time. This paper suggested a methodology for mining huge volume of data from data centers using LSTM in deep learning. Also suggested techniques for fake or duplicate data memory space will be wasted and the cost for configuring the data centers will also be avoided [31]. The Sequential LSTM will be the solution for the continuous monitoring of data status over the large network and will monitor the data position in parallel. It summarizes the optimization techniques along with web content mining for data extraction on the larger networks among all the cluster nodes. The output of the LSTM system will be given as an input on the same layer itself to identify the changes on the status of the data lively. Because of this RNN approach the changes have made on websites monitored immediately and contents have extracted with accurate level. This details has taken from different web information extraction approaches advantages and disadvantages, but it will satisfy the researchers need like accuracy and latency on the network effectively [32].

This discussion helps the social media peoples those who are always in a part of generating and accessing huge volume of data on a larger network can easily post their web advertisement, web product items, and web news in up-to-date status [33]. Huge data creation may cause the problem of spreading fake news to social media and it would lay the cause for so many fraudulent web contents. Suppose any unauthorized data can be merged with original data immediately it will be noted and sent notification to the centralized controller. Continuous monitoring of the geolocation of data with its generation time has helped the social media peoples for identifying the validity and present state of the entire data. Extracting the data from the huge data warehouse has given low latency and high throughput [34]. To keep track of all information continuously configuring huge data centers must be assigned and it will make huge cost. The cost will be reduced using this research work to produce fewer data centers with minimum space requirements. In the future, dynamic and static websites will be taken for optimization on a real-time basis using big data analytics with the help of modern tools such as Hadoop, SPARK. TTL (Time to Live) is the concept used to monitor the process alive time throughout the entire operation. This will be improved using deep learning algorithms for increasing the life time of the process on larger networks. The intermediate memory capacity will be increased in the future to hold the huge volume of data on a large network cluster. Similarly, data status changes are stored in a limited memory which will create a memory problem for storing a huge volume of data on larger networks [35].

References

[1] M. J. H. Mughal, “Data mining: Web data mining techniques, tools and algorithms: An overview,” *Information Retrieval*, vol. 9, no. 6, 2018.

[2] H. S. Maghdid, “Web news mining using new features: A comparative study,” *IEEE Access*, vol. 7, pp. 5626–5641, 2018.

[3] T.-M. Song and J. Song, “Prediction of risk factors of cyberbullying-related words in korea: Application of data mining using social big data,” *Telematics and Informatics*, vol. 58, p. 101524, 2021.

[4] M. Moradi, E. Ghanbari, M. Maeen, and S. Harifi, “An approach based on combination of features for automatic news retrieval,” *arXiv preprint arXiv:2004.11699*, 2020.

[5] A. Mulahuwaish, K. Gyorick, K. Z. Ghafoor, H. S. Maghdid, and D. B. Rawat, “Efficient classification model of web news documents using machine learning algorithms for accurate information,” *Computers & Security*, vol. 98, p. 102006, 2020.

[6] M. Umer, Z. Imtiaz, S. Ullah, A. Mehmood, G. S. Choi, and B.-W. On, “Fake news stance detection using deep learning architecture (cnn-lstm),” *IEEE Access*, vol. 8, pp. 156 695–156 706, 2020.

[7] X. Wang, L. Kou, V. Sugumaran, X. Luo, and H. Zhang, “Emotion correlation mining through deep learning models on natural language text,” *IEEE transactions on cybernetics*, 2020.

[8] L. Hu, C. Li, C. Shi, C. Yang, and C. Shao, "Graph neural news recommendation with long-term and short-term interest modeling," *Information Processing & Management*, vol. 57, no. 2, p. 102142, 2020.

[9] S. Hong, "A study on stock price prediction system based on text mining method using lstm and stock market news," *Journal of Digital Convergence*, vol. 18, no. 7, pp. 223–228, 2020.

[10] Q. Zhang, B. Lian, P. Cao, Y. Sang, W. Huang, and L. Qi, "Multi-source medical data integration and mining for healthcare services," *IEEE Access*, vol. 8, pp. 165 010–165 017, 2020.

[11] S. Shahrivari and S. Jalili, "Efficient distributed k-clique mining for large networks using mapreduce," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[12] C. S. Kumar and R. Santhosh, "Effective information retrieval and feature minimization technique for semantic web data," *Computers & Electrical Engineering*, vol. 81, p. 106518, 2020.

[13] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama *et al.*, "Speed/accuracy trade-offs for modern convolutional object detectors," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7310–7311.

[14] A. Anandhan, L. Shuib, M. A. Ismail, and G. Mujtaba, "Social media recommender systems: review and open research issues," *IEEE Access*, vol. 6, pp. 15 608–15 628, 2018.

[15] Y. Tang, H. Wang, K. Guo, Y. Xiao, and T. Chi, "Relevant feedback based accurate and intelligent retrieval on capturing user intention for personalized websites," *IEEE Access*, vol. 6, pp. 24 239–24 248, 2018.

[16] Z. Tan, C. He, Y. Fang, B. Ge, and W. Xiao, "-based extraction of news contents for text mining," *IEEE Access*, vol. 6, pp. 64 085–64 095, 2018.

[17] S. Zhang and H. Zhong, "Mining users trust from e-commerce reviews based on sentiment similarity analysis," *IEEE Access*, vol. 7, pp. 13 523–13 535, 2019.

[18] H. Si, W. Sun, J. Zhang, J. Wan, N. N. Xiong, L. Zhou, and Y. Ren, "An effective identification technology for online news comment spammers in internet media," *IEEE Access*, vol. 7, pp. 37 792–37 806, 2019.

[19] H. Ye, B. Cao, Z. Peng, T. Chen, Y. Wen, and J. Liu, "Web services classification based on wide & bi-lstm model," *IEEE Access*, vol. 7, pp. 43 697–43 706, 2019.

[20] A. Da'u and N. Salim, "Sentiment-aware deep recommender system with neural attention networks," *IEEE Access*, vol. 7, pp. 45 472–45 484, 2019.

[21] G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu, "Sentiment analysis of comment texts based on bilstm," *Ieee Access*, vol. 7, pp. 51 522–51 532, 2019.

[22] G. Li and R. M. De Carvalho, "Process mining in social media: applying object-centric behavioral constraint models," *IEEE Access*, vol. 7, pp. 84 360–84 373, 2019.

[23] Y. Li, C. Liu, N. Du, W. Fan, Q. Li, J. Gao, C. Zhang, and H. Wu, "Extracting medical knowledge from crowdsourced question answering website," *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 309–321, 2016.

[24] B. Bai, Y. Fan, W. Tan, and J. Zhang, "Dltsr: A deep learning framework for recommendations of long-tail web services," *IEEE Transactions on Services Computing*, vol. 13, no. 1, pp. 73–85, 2017.

[25] C. Zhang, D. Jin, X. Xiao, G. Chen, and M.-L. Shyu, "A novel collaborative optimization framework for web video event mining based on the combination of inaccurate visual similarity detection information and sparse textual information," *IEEE Access*, vol. 8, pp. 10 516–10 527, 2020.

[26] Z. Imtiaz, M. Umer, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, "Duplicate questions pair detection using siamese malstm," *IEEE Access*, vol. 8, pp. 21 932–21 942, 2020.

[27] D. Liu and J.-H. Lee, "Cnn based malicious website detection by invalidating multiple web spams," *IEEE Access*, vol. 8, pp. 97 258–97 266, 2020.

[28] E. Uzun, E. Özhan, H. V. Agun, T. Yerlikaya, and H. N. Buluş, "Automatically discovering relevant images from web pages," *IEEE Access*, vol. 8, pp. 208 910–208 921, 2020.

[29] T. L. Cerqueira, F. C. Bertoni, and M. G. Pires, "Instance genetic selection for fuzzy rule-based systems optimization to opinion classification," *IEEE Latin America Transactions*, vol. 18, no. 07, pp. 1215–1221, 2020.

[30] F. Es-Sabery, A. Hair, J. Qadir, B. Sainz-De-Abajo, B. García-Zapirain, and I. De La Torre-Díez, "Sentence-level classification using parallel fuzzy deep learning classifier," *IEEE Access*, vol. 9, pp. 17 943–17 985, 2021.

[31] M. Mirtsch, J. Kinne, and K. Blind, "Exploring the adoption of the international information security management system standard iso/iec 27001: A web mining-based analysis," *IEEE Transactions on Engineering Management*, vol. 68, no. 1, pp. 87–100, 2020.

[32] P. Ma, B. Jiang, Z. Lu, N. Li, and Z. Jiang, "Cybersecurity named entity recognition using bidirectional long short-term memory with conditional random fields," *Tsinghua Science and Technology*, vol. 26, no. 3, pp. 259–265, 2020.

[33] S. Li and Y. Liu, "News video title extraction algorithm based on deep learning," *IEEE Access*, vol. 9, pp. 12 143–12 157, 2021.

[34] S. Long, X. He, and C. Yao, "Scene text detection and recognition: The deep learning era," *International Journal of Computer Vision*, vol. 129, no. 1, pp. 161–184, 2021.

[35] D. Pandey, B. K. Pandey, and S. Wairya, "Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images," *Soft Computing*, vol. 25, no. 2, pp. 1563–1580, 2021.

Raghavendra R is pursuing his Ph.D. in Computer Applications at Department of MCA, Ramaiah Institute of Technology (Affiliated to Visvesvaraya Technological University, Jnana Sangama, Belgavi) Bangalore, 560054, India. He holds a M.C.A master degree in Computer Science from University of Mysore(2012), a B.Sc. degree in Computer Science from the University of Mysore(2009) India. where he currently holds the position of Assistant Professor in Jain(Deemed-to-be University), Bangalore, India.

Dr. Niranjanamurthy M is hold his Ph.D in Computer Science at JJT University, Rajasthan, India 2016, M.Phil - Master of Philosophy in Computer Science from VM University(2009), Salem, Tamilnadu India, MCA - Master of Computer Application from BMS Institute of Technology,Bangalore, Karnataka affiliated to Visveswaraiah Technological University(2007), Belgaum, India , BCA - Bachelor of Computer Application (BCA) at Kuvempu University(2004), Shimoga, India.