VAWKUM Transactions on Computer Sciences
http://vfast.org/journals/index.php/VTCS@ 2022, ISSN(e): 2308-8168, ISSN(p): 2411-6335
Volume 10, Number 2, July-December 2022 *pp:96-108*

Questgator: A Platform for Content Aggregation and Text Classification

Abdul Latif Shaikh, Fizza Abbas Alvi *, Babar Ali, Ubaidullah Rajput, Hadi Bux

Department of Computer Systems Engineering, QUEST Nawabshah, Pakistan

Keywords: *Content aggregation, Text classification, Metasearch engine, Classification.*

Journal Info:
Submitted:
November 22, 2022
Accepted:
December 21, 2022
Published:
December 31, 2022

Abstract
The Web has witnessed a surge in content over recent years. Content is revolutionizing the way people conduct business, communicate, and make informed decisions. However, the vast amount of data used for communication today is often unstructured and challenging to comprehend. Content aggregators provide a solution to this problem by collecting data from various sources and organizing it into a structured format in one place. This research proposed the content aggregator "Questgator" that extracts content for example news, scholarships, jobs, books, video content, and research papers. In this paper Naive Bayes theorem is used for text classification. Moreover, paper also provides comparison with other platforms to show the efficiency of proposed content aggregator.

***Correspondence Author Email Address:**
fizza_alvi@quest.edu.pk

1 Introduction
From the rise of the Web to the current era, content is being digitized, providing services like aggregators of content in many prominent fields [1]. Bringing data from multiple search engines can provide a more comprehensive and diverse set of results compared to using just one search engine, such as Google. This is because each search engine has its own algorithms, indexing methods, and ranking factors, and therefore may return different results for the same search query. By using multiple search engines, users can get a wider range of results, including those that may not have been returned by a single search engine. This can be especially useful for more specialized or niche searches, where the results from one search engine may not be as comprehensive. The fields include News Media, Research Studies, Business Intelligence, Education, Entertainment, and numerous others. Industries need reliable content to display to their users and content aggregation gives them exactly that. Among those industries are the mainstream web services like Meta, Netflix, Deezer, Reddit, Google, Spotify, etc [2]. The aggregation platform collects content from different sources and exhibits it in one place. Doing so, it reduces the resources used for getting the content and providing updated information to the end-users [3]. The researchers and scientists were producing an abundance of essential research and information, but it wasn't humanly possible for

researchers to comb through this complicated collection of data and develop anything useful [3]. Content aggregators come in handy when the information needs to be versatile.

Various agencies and channels generate and deliver news information in various forms, languages, and formats. Web aggregator services in the news field are thus quite valuable since they may save a significant amount of user time when looking for and obtaining news content from several channels and sources [4]. Not only it saves time, but it can ensure that the news that the user is getting is reliable. Since not all sources can spread fake news at the same time.

Sal studies have shown that none of the various online searches and users have comprehensive coverage of the Web, and it is unlikely that any hem will ever do so [5]. A metasearch engine is a service that gives consolidated access to numerous existing online search engines, increasing WWW coverage. Result aggregation methods are used in a metasearch engine to combine the results from many underlying internet search engines and provide the user with a single aggregated result list [6].

A search engine that provides accurate and specialized information is required to provide job searchers with quick access to recruiting information from a single source [7]. As a result, vacant position classification might be one of the answers for directing job searchers to positions that match their qualifications, preferences, and abilities[8].

This research proposed the content aggregator "Questgator" which provides different types of material from multiple sources. News, Scholarships, Jobs, and Books are aggregated from various sources. To make things more reliable, the search results will also be aggregated from other search engines. That will be done by the Metasearch engine implemented in the platform.

The rest of the paper is as follows. Section 2 highlights the background study. Section 3 discusses the related work. Section 4 provides the proposed approach. Section 5 shows the results and Section 5 concludes the paper.

2 Background

In a world where everything happens in a matter of seconds, users need to do a lot of things in the least time possible. Specially on the internet, it takes a lot of time to go through websites to get the desired content and to analyze it. Users cannot depend on one source only; it may provide false statements or biased information [9]. For that, having multiple sources provides reliable information. That is achieved by our aggregator. Students get distracted nowadays by the distractions around the Web. As a student, getting distracted comes with a high cost. Search engines control the content user see which results in the loss of the search results that users might require. But the proposed meta-search engine provides results without any content control. In this section, a background study is discussed.

2.1 Content

Content is the way of presenting your information to the audience in the form of text, video, and images. It provides information and entertainment that people are actively seeking and willing to pay for with their attention, money or both. It wouldn't be wrong to consider content as a "conversational object" because it helps in initiating social interactions among individuals [9]. It is also an asset that assists someone in doing their task.

2.1.1 Importance of Content

Nowadays, the economy is based on digital assets where data is more precious than ever. It is essential for the appropriate operation of daily tasks, from the authorities to small businesses. Without it, everything would stop working. It is fundamentally changing existing business baselines and introducing newly available potential to the market [10].

Digital media is replacing text and speech as the primary means of communication. People are using images to communicate more than just words. All these developments have resulted in enormously large amounts of data - mining this data and using it to gain an edge is the essence of an organization's success.

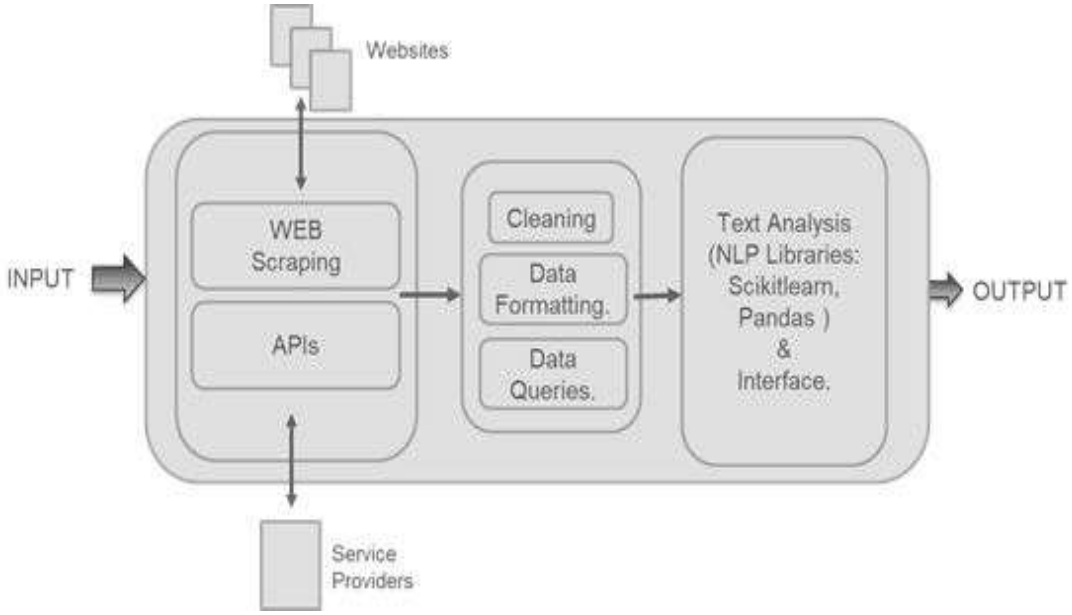

Figure 1. Questgator Architecture

2.1.2 Content Aggregator

A content aggregator is a web/mobile application that gathers web material in one place for simple viewings, such as text, media, and hyperlinks. It is a service that collects and analyzes content based on a single subject with one or more associated keywords [11]. These services aggregate material from a variety of sources into a single place, decreasing transaction costs and offering new information to users. Aggregating material from other individuals complements your unique work while also offering a broader perspective [9].

A content aggregator is an online organization that collects web or media content, applications, or both for reuse and resale. It is a method of content curation. Data will be gathered in real-time and from a variety of reliable platforms in our aggregator, ensuring data integrity and updating. It is how the data will be collected and represented in the future.

There are many applications for example Usepanda.com is a famous content aggregator that collects and shows news stories on the platform, Fastweb.com is a well-known free scholarship aggregator and Indeed.com Indeed.com is the most well-known of the top job search websites [7].

2.2 Techniques of Content Aggregation

The history of content aggregation goes back to the history of data collection for different purposes such as analysis, information, education, and entertainment. Back in 1993, for web-based data collection different techniques were used [7]. Some were manually done, and some were automized by using tools like scrapers but nowadays data collection is somehow easy because of APIs provided by service providers.

2.2.1 Web Scraping

Web scraping is used to transform unstructured data on the web into structured data that can be stored and analyzed in a central local database or spreadsheet. The data is extracted from websites with the help of scraping tools. While web scraping may be done manually, automated methods are usually preferable for scraping web data since they are less expensive and work faster. Automated web scrapers operate in a basic yet sophisticated manner [8]. First, the web scraper will be given one or more URLs to load before scraping. The scraper subsequently loads the full HTML code for the page in question. More sophisticated scrapers will render the whole webpage, including CSS and JavaScript components.

2.2.2 API and Its Uses

An application programming interface, or API, allows businesses to offer up data and functionality of their applications to third-party developers, commercial partners, and internal departments inside their organizations. A defined interface, allows services and products to communicate with one another and use one another's data and capabilities [9]. Developers aren't required to understand how an API works; they just utilize the interface to interact with other products and services. API usage has increased dramatically over the last decade, to the point that many of today's most popular online apps would not be feasible without them [21].

APIs will be used during the extraction of those websites that don't allow us to scrap their content. These websites provide the APIs that let us use the data and their services with a limit.

2.2.3 Meta Search Engine

A metasearch engine, also known as a search aggregator, is an online portal that provides a unique algorithm to aggregate web search results for a phrase or term from several search engines. It merely pulls data from other web search engines. It also allows the user to enter a single query and discover the results from several sources, allowing the user to rapidly receive the best answers from a wide range of information [10].

Metasearch engines provide their results in two formats: single and multiple lists. Most meta-search engines provide results in a single consolidated list that has been purged of duplicate keywords. Others do not integrate numerous engine results, but instead provide them in separate lists, one for each engine. However, this may result in numerous entries.

Using the internet for research or study is advantageous since it provides broad limitless access to knowledge and information. However, comparing different search engine results takes a significant amount of time. Distinct search engines have different crawlers, indexes, algorithms, and many more [23].

2.3 Text Classification

The practice of classifying text into an ordered structure is known as text classification. Text Classification is a supervised machine learning task since a labeled dataset including text articles and their labels is used to train a classifier. Text classifiers can automatically assess text and assign a set of pre-defined tags or categories depending on its content using Natural Language Processing (NLP) [10].

Text classification is becoming a more significant aspect of the business since it enables for easy data analysis and business process automation.

2.3.1 Naïve Bayes Theorem

The simple yet powerful technique used for text classification is a classification technique based on Bayes' Theorem and the predictor independence assumption. In simple terms, a Naive Bayes classifier algorithms the presence of one feature in a class has no bearing on the presence of any other feature.

The Naive Bayes model is easy to build and works well with large data sets. Because of its simplicity, Naive Bayes is known to outperform even the most powerful classification algorithms [11].

3 Related Work

Numerous applications use Metasearch, Scraping, and APIs. In this section, research on this application is discussed.

Web Scraping has had a lot of applications in recent years. In [13] authors examined existing scraping frameworks and tools, noting their extraction capabilities' strengths and limits. The main goal was to demonstrate how simple it is today to build up a data scythe raping pipeline with minimum programming effort and meet a variety of practical applications. They used a biomedical data extraction scenario in which the essential data sources were presented, which are well-known in clinical microbiology and related

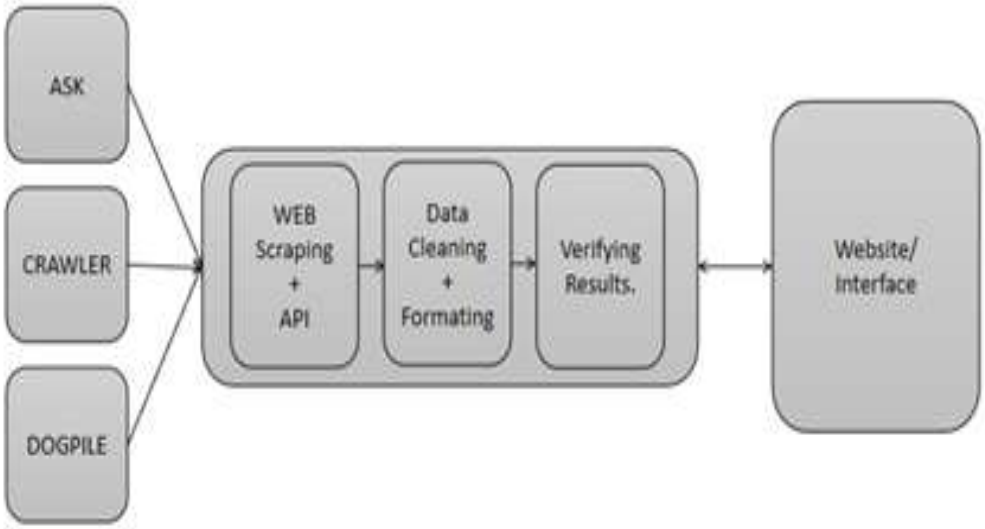

Figure 2. Proposed Meta Search Engine

fields. Furthermore, the paper discussed the working of WhichGenes and PathJam, two bioinformatics meta-servers that deal with gene set enrichment analysis via scraping.

To check the usage of a metasearch engine named Dogpile.com, a research paper was proposed [14]. The study explored 2,465,145 interactions from 534,507 Dogpile.com users on May 6, 2005, and cthe ross-referenced these results to findings from previous Web searching studies.

In [15], authors developed a smart Meta search engine that covers the web more thoroughly by targeting multiple search engines for different search categories. A knowledge base was used to modify the user's search query, and an appropriate search engine was chosen. The results of various search engines were combined and rated. These results were sorted into categories using a classification system, and the results were shown in that order. It was useful to the user who had a vague understanding of the information he or she was looking for [15].

In [16], authors proposed to build a meta-search engine named Helios. Helios was an open-source metasearch engine that was both versatile and efficient. It was meticulously optimized for speed and lightweight, making it suitable for use on low-cost platforms. It worked using 18 search engines, Web, Books, News, and Academic research domains, etc. Those 18 search engines were A9, About, AllTheWeb, Altavista, AOL Search, eSpotting, FindWhat, Gigablast, Google, LookSmart, Mozdex, Msn, Overture, Ask/Teoma, Yahoo!, Google News, Google Scholar, Yahoo News. Although, other search engines may be simply added [16]. In [17], researchers introduced the SocIoS platform and attempted to solve issues such as data diversity and the absence of a standardized manner of accessing it. This was accomplished by offering tools that run on top of a variety of prominent social networks, giving consistent access to their data. It offers a standardized entry point for gathering network data and functionality, as well as a collection of analytical tools for exploitation. The SocIoS API was proposed in the study as an abstraction layer on top of social networks that exposes operations that encapsulate the functionality of respective APIs. The platform supports seven social networks as of now, but it is scalable [17].

In [18], authors proposed a news aggregator which provided a user interface different from existing news aggregators. It also demonstrated the capability of analyzing text and URLs. The main purpose of developing this application was to provide a news aggregator application with a user interface, different from existing news aggregator applications in Indonesia. The application had implementations of gamification and the capability of analyzing text and URLs. The authors also conducted a survey that probed their preferences for a news aggregator [18].

The variety of jobs due to the availability of work from many sources creates a complicated puzzle in the selection process, resulting in inefficient runtime. For this issue authors offered a simple strategy for simplifying job searching through the development and collaboration of a web scraping technology and categorization using Nave Bayes on a search engine [19]. This research is yielding an effective and efficient application for users to search for possible jobs that match their preferences [19].

Text classification is used to classify documents into defined categories. It is extremely useful in Web content management, search engines, email screening, and other applications. It is a challenging assignment since the feature vector is high-dimensional and contains noisy and irrelevant characteristics. Several feature reduction strategies have been presented for removing unnecessary features and lowering the dimension of the feature vector. For improved classification outcomes, the machine learning model employs a relevant and reduced feature vector.[20] article discusses numerous approaches to text classification that use machine learning techniques and feature selection strategies to reduce the high-dimensional feature vector.

The above studies highlight that managing a large amount of data is challenging. People tend to get confused and distracted by this huge data. Therefore, a platform is proposed where users can find all the necessary content related to education. Another motivation was that users didn't want to depend on one source. For that purpose, the proposed platform could retrieve information from multiple trusted sources.

4 Proposed Approach

This section provides details of the proposed "Questgator". This research has two phases, first phase is to develop the backend portion of the project which involves python programming for scraping, API usage, and machine learning for classification, and then secondly, cleaning that data for the website development part.

4.1 Proposed Architecture

The proposed architecture is followed the following steps and is shown in Figure 1.

The first step is that the user will input a query into the search bar or will be given a default input. The aggregator will take that input and go into the data collection stage. The aggregator checks the query and uses one of the two methods to request the data. It may directly go through website pages and scrape them with BS4 (A python library used for scraping the web content directly) Although, sometimes this approach does not work as some websites do not allow their web pages to be scrapped. In that case, some website owners provide access to their website content through the help of APIs. The aggregator uses these APIs to access their content by sending the API request to a service provider with the queries.

Data goes through the pre-processing stage where it is cleansed of any errors and unnecessary results. Here, BS4 does all the cleaning by reducing all the useless HTML code that does not concern the user query and other unnecessary data. The unnecessary data is removed using simple python string functions and regex for regular expression cleaning for converting data into a JSON format for further processes.

Next is the processing stage, the cleansed data is categorized accordingly using the Natural Language Processing algorithms such as Scikit-learn and Naive Bayes theorem for text classification. The final step sends the data back to the interface which displays the data that the user sees on the screen using the website page (An interface by which the user can access our tools and see the results) which is built by using the Django framework of python for website development.

4.2 Meta Search Engine

The meta search engine is designed by scraping and using the API available free on internet from multiple search engines such as metacrawler.com, ask.com, dogpile.com the main reason for selecting these is that these are also metasearch engine, and we are building our meta search engine on these search engines. That will make our search engine a "Super Meta Search Engine" as shown in the diagram below

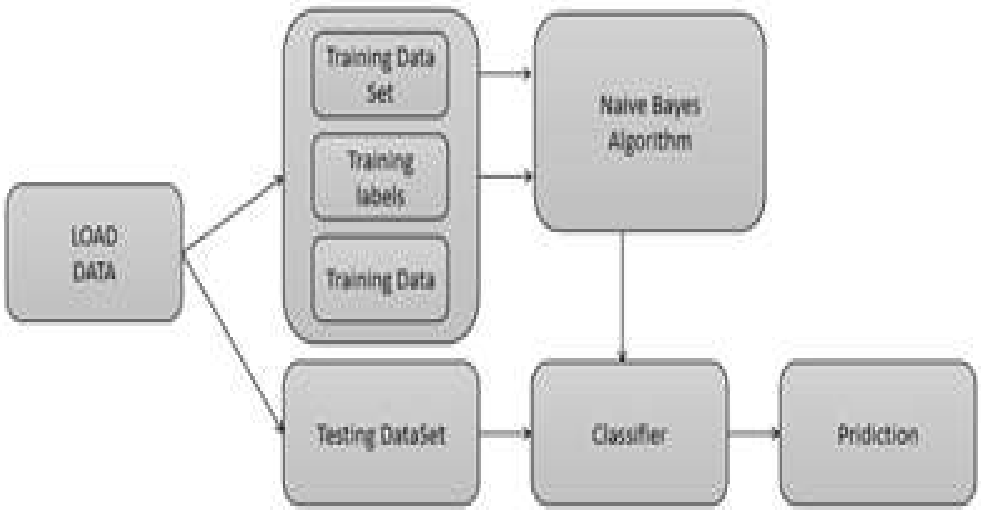

Figure 3. Overall working flow of text classification

4.3 News Aggregator

The news aggregator is designed by scraping the trusted news sources of Pakistan for now in this project we have scraped DAWN News, ARY News, SAMMA News, and BBC News. After collecting all the data using the Bayes theorem, we categorized our news.

4.4 Scholar with a Timeline Graph

This feature of the proposed approach built a copy of google scholar by scraping that scholar because for now, it is not providing any API for that after collecting that data and then cleaning that data in lists, we have plotted a graph using “matplotlb” library for plotting different types of plots/graphs.

4.5 Text Classification Using Naive Bayes Theorem

Web content classification differs from normal text classification because of its uncontrollable nature. The naive Bayes classifier assumes that all the features are unrelated to each other. The presence or absence of a feature does not influence the presence or absence of any other feature. For implementing the Bayes theorem scikit learn which is the python library used for text classification will help here to build a Naive Bayes model in Python. The dataset used in the proposed project is fetch_20newsgroups from the Scikit learn library this is the pre-installed data set on which we can train our model. The steps are as follows:

- Importing all the necessary libraries.
- Importing the data set and dividing it into train and test data sets.
- Vectorizing the data sets and targeted categories (assigning the weights because the machine can only understand numbers not text. Value is assigned according to the keyword importance)
- Training the data through pipelining as shown in the code.
- Then predicting the new data(query) to check the accuracy.

The overall working flow of text classification is shown in Figure 3.

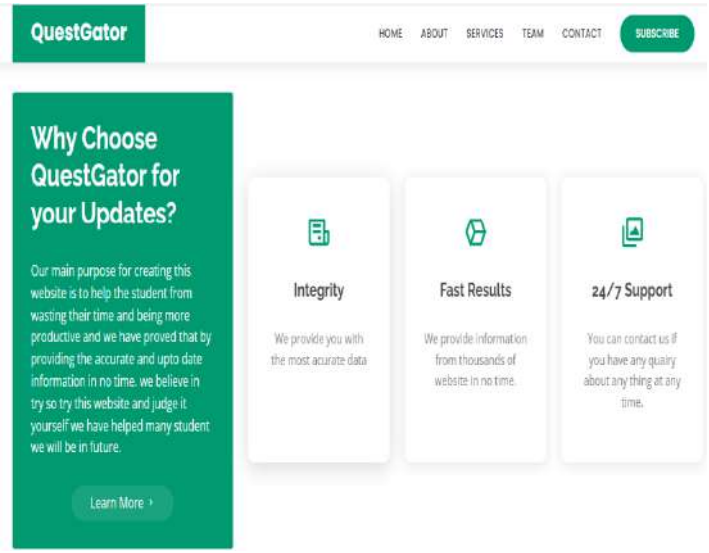

Figure 4. Landing Page of the Content Aggregator

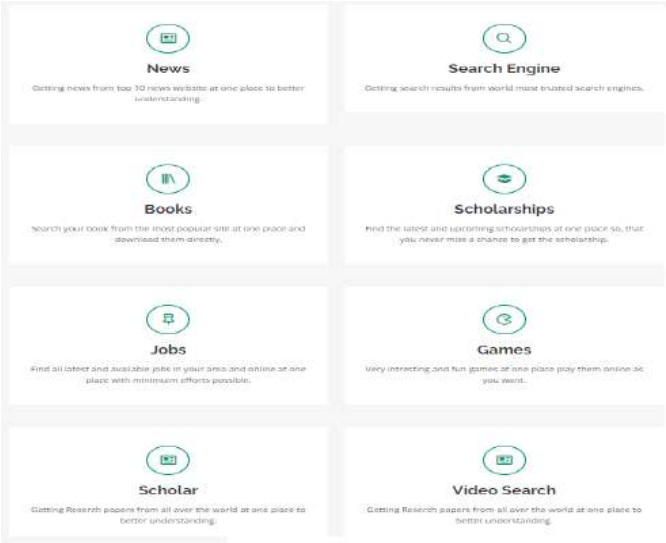

Figure 5. Services provided by Proposed platform

4.6 Website Development Phase

The website of the proposed research project is built by using the Django framework of python for backend development of the website, bootstrap for website designing is used because it reduces the code and makes it easy to build the website in not much time which was very helpful. The languages used for that are HTML, CSS, JavaScript, and Python.

The proposed research project requires some packages that need to be installed for it to work properly. It includes Python for starters. Python is very popular among data scientists because it has a wide range of packages that can be used to perform a variety of operations from data wrangling to visualization. It is a great general-purpose programming language and with the help of these powerful packages i.e. NumPy, SciPy, and Matplotlib, it becomes a powerful environment for scientific computing.

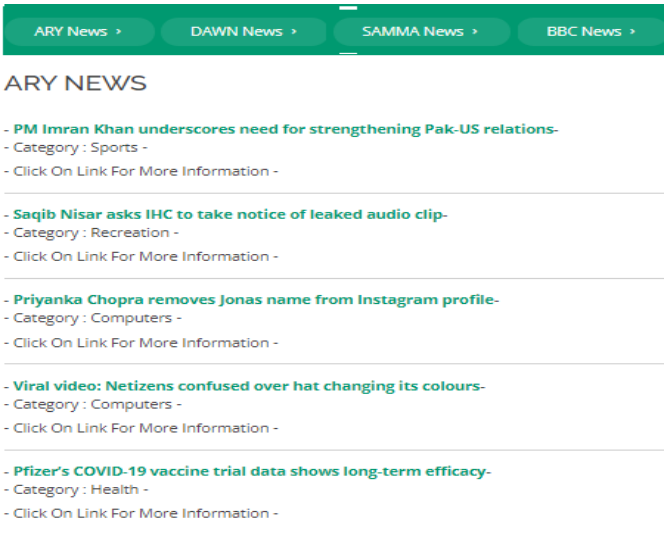

Figure 6. News Aggregation

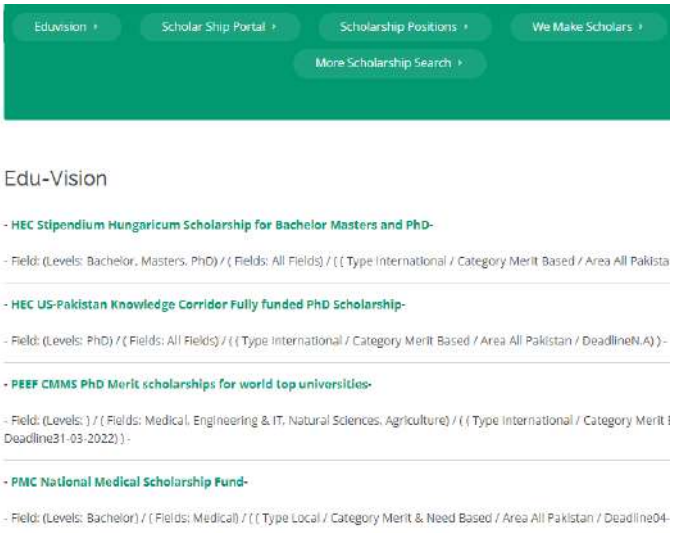

Figure 7. Scholarships Aggregation

5 Results and Discussion

This section shows the results of the proposed approach. It also contains the figures for the interfaces of the website.

5.1 Landing Page of the Content Aggregator

In Figure 4, the interface is shown where the user will arrive when visiting the website. It contains Navigational Buttons. We have used bootstrap and Django for designing the website interface.

5.2 Services provided by Proposed platform

Figure 5 shows the services provided by the platform such as News, Metasearch Engine, Books, Scholarships, Jobs, Games, Scholarship, and Video Search. Each of these services provides different content that is aggregated from multiple sources. Clicking/Tapping on a category sends the user to a page where they can choose from a variety of sources.

5.3 News Aggregation

Figure 6 shows the News aggregation interface. This is the interface provided when a user tries to access the new aggregator in which there are multiple channels for getting news and the user can see according to his/her demand. Such as ARY News, DAWN News, SAMMA News, and BBC News.

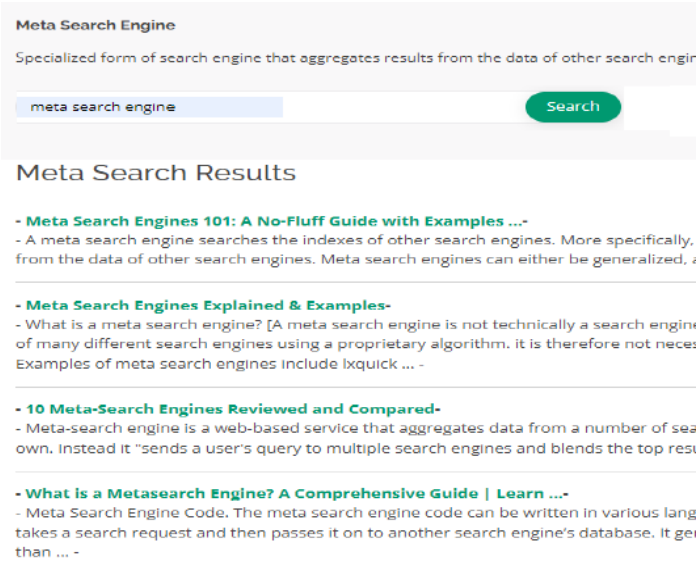

Figure 8. Metasearch Engine

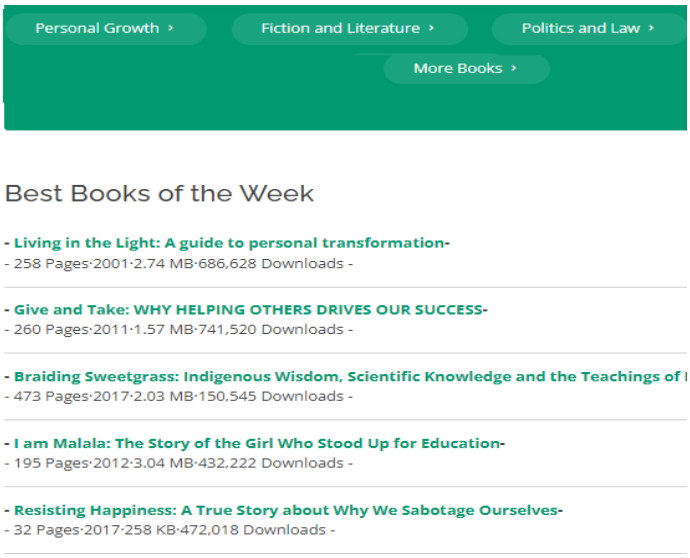

Figure 9. Books Aggregation Interface

In Figure 7, the interface belongs to scholarships in which users can get information about scholarships from different platforms such as 'Edu-Vision', 'Scholar Ship Portal', 'Scholarship Positions', and "we make scholars".

5.5 Metasearch Engine

Figure 8 shows all about the search engine you can search anything users want, and it will fetch results from multiple different search engines at one place on the run time, so it is fast, efficient, and up to date.

5.6 Books Aggregation

Figure 9 shows the page where users can find all the books that user want to download in one place which is a very powerful service to have. Users will get direct links to download or the link to a website that sells the book online.

5.7 Jobs aggregation

The interface belongs to jobs in which you can get information about all the latest jobs in one place from jobs.pk, job alert.pk, job.com.pk and pakistanjobbank.pk is shown in Figure 10.

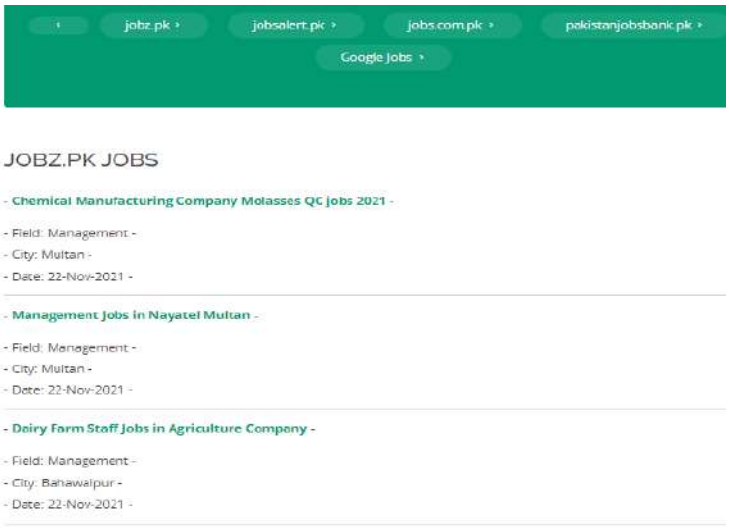

Figure 10. Jobs Aggregator Interface

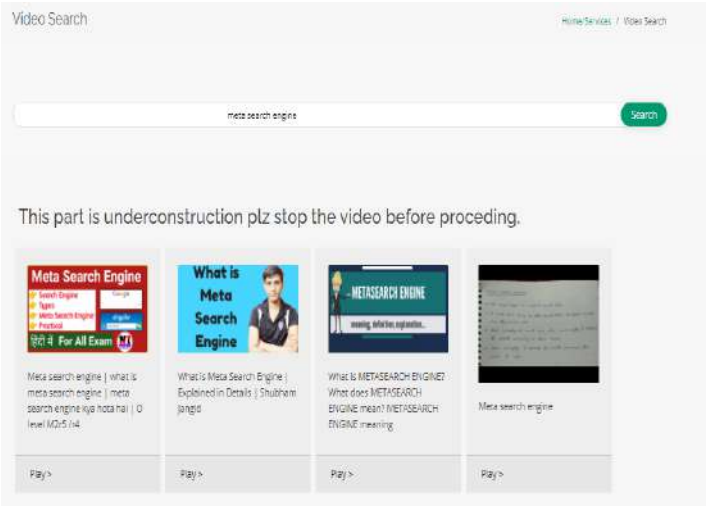

Figure 11. Video Search Engine

5.8 Video Search Engine

Figure 11 shows the Interface of the Video Search Engine. This is the search engine for searching videos from multiple platforms in one place with a very user-friendly environment.

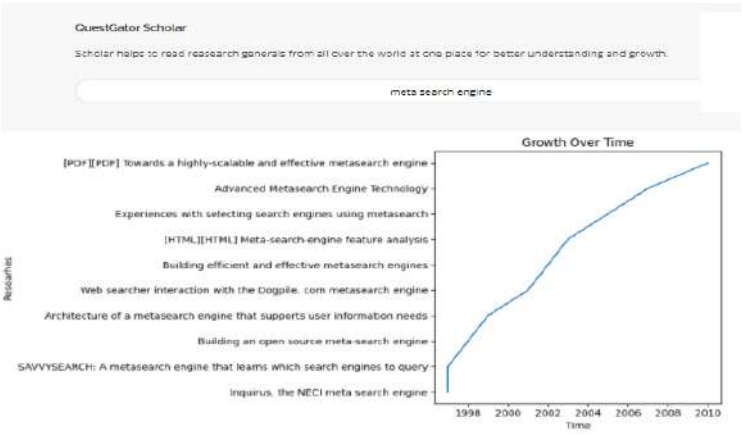

Figure 12. Research Paper-Time Graph

Figure 13. Research Paper Search Engine

5.9 Research Paper-Time Graph

The interface of the project provides users the Time Graph of any topic in which any research is conducted from start to latest research is shown in Figure 12. Users can see all the research by any professor with

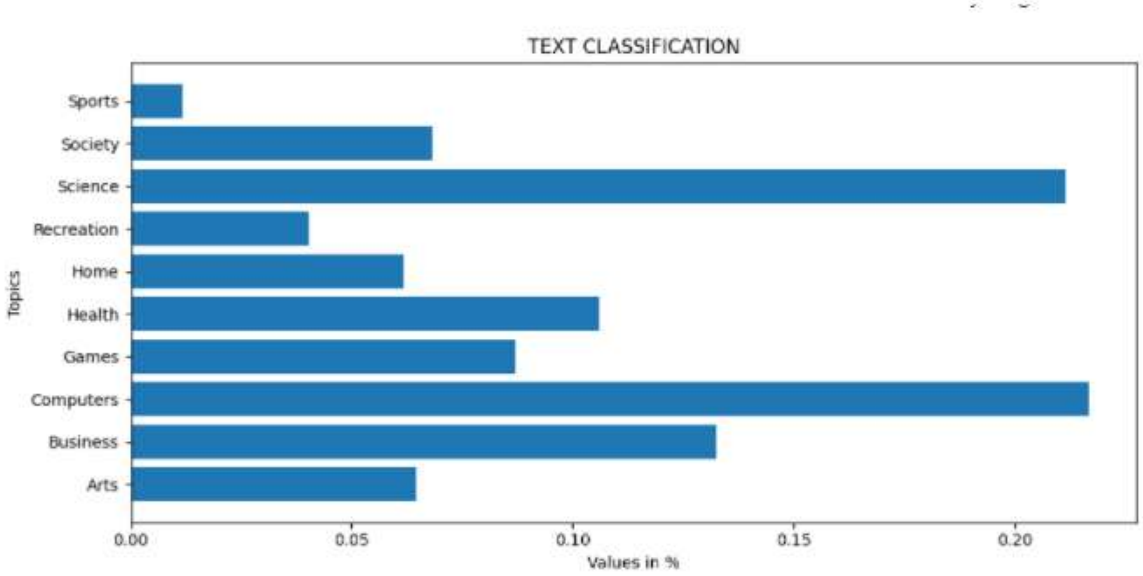

Figure 14. The Interface of Text Classification

Table 1. Comparison Among Existing Aggregators

Sites	NEWS	JOBS	SCHOLARSHIPS	Scholars	Games	Super Meta Search Engine
Use Panda	Yes	No	No	No	No	No
Indeed	No	Yes	Yes	No	No	No
Fastweb	No	No	Yes	No	No	No
Ask	Yes	Yes	Yes	Yes	Yes	No
Questgator	Yes	Yes	Yes	Yes	Yes	Yes

the help of a simple time diagram.

5.10 Research Paper Search Engine

Figure 13 shows the research paper search engine. This is a search engine only specific to research just like google scholar but with the addition of graphs that help understand it in a much better and easy way.

5.11 Text Classification

Figure 14 shows the interface of the Text classification based on machine learning in which the user can provide it with any paragraph, and it will guess the category of that paragraph, this part is also used in news to categorize the news.

5.12 Comparison with other platforms

In this section, we will discuss the results of our proposed solution and compare it with the existing systems. The existing systems are not generally like our content aggregator. We have tried to merge different options for users so that they never miss an important update or information. It has NEWS Aggregator, Scholarships Aggregator, Books Aggregator, and Jobs Aggregator. we have also built a text classifier that can identify the topic of any given text. Table 1 presents the comparison of some of the sites that use scraping to aggregate the different available sites

6 Conclusion

In this paper, a platform named "Questgator" is presented. The proposed content aggregator platform is designed for students. By using the proposed platform students can get information regarding news, jobs, scholarships, books, Meta Search Engines, and Research papers that save time and effort. This platform is an initiative in making a stand for making platforms that are distraction-free and help the student and others in achieving excellence. Moreover, a comparison with an already available aggregator is also provided that shows the efficiency of the proposed platform. In the future, we will add more and more features such as 'Text Analysis', 'Paraphrasing', 'Image Text Converted', and free and Paid Courses', for students.

Author Contributions

Abdul Latif Shaikh : Conceptualization, Methodology, Software. **Fizza Abbas Alvi**: Supervision, Data curation, Writing- Original draft preparation. **Babar Ali**: Visualization, Investigation. **Ubaidullah Rajput**: Software, Validation. **Hadi Bux**: Writing- Reviewing and Editing.

Compliance with Ethical Standards

It is declared that all authors don't have any conflict of interest. It is also declare that this article does not contain any studies with human participants or animals performed by any of the authors. Furthermore, informed consent was obtained from all individual participants included in the study.

References

[1] L. Chiou and C. Tucker, "Nber working paper series content Aggregation by platforms: The case of the News media," Nber.org, 2015. [Online]. Available: https://www.nber.org/system/files/working_papers/w21404/w21404.pdf.

[2] A. J. Bottomley, "Giant pools of content: Theorizing aggregation in online media distribution," JCMS: Journal of Cinema and Media Studies, vol. 59, no. 1, pp. 149–156, 2019.

[3] G. Lloren, "Content Aggregation and Context Curation: Redefining the Future of Online Advertising Through Information Syndication", Asian Congress for Media and Communication, 2014.

[4] Chowdhury, S., and Landoni, M., "News aggregator services: user expectations and experience", Online Information Review, Vol. 30 No. 2, pp. 100-115, 2006.

[5] R. M. Losee, "When information retrieval measures agree about the relative quality of document rankings," J. Am. Soc. Inf. Sci., vol. 51, no. 9, pp. 834–840, 2000.

[6] N. Dimri, H. Kaul, and D. Gupta, "MetaXplorer: an intelligent and adaptable metasearch engine using a novel ordered weighted averaging operator," Int. J. Syst. Assur. Eng. Manag., vol. 9, no. 6, pp. 1315–1325, 2018.

[7] A. Pranav and S. Chauhan, "Efficient focused web crawling approach for search engine," Ijcsmc.com. [Online]. Available: https://www.ijcsmc.com/docs/papers/May2015/V4I5201599a17.pdf.

[8] V. Singrodia, A. Mitra, and S. Paul, "A Review on Web Scrapping and its Applications," in 2019 International Conference on Computer Communication and Informatics (ICCCI), 2019.

[9] O. Castrillo-Fernández, "Web scraping: Applications and tools," Europa.eu. [Online]. Available: https://data.europa.eu/sites/default/files/report/2015_web_scraping_applications_and_tools.pdf, 2015.

[10] X. Chen, Q. Li, Y. Lin and B. Zhou, "A synthesized method of result merging in meta-search engine," 10th International Conference on Human System Interactions (HSI), 2017, pp. 206-211, 2017

[11] J. Chen, H. Huang, S. Tian, Y. Qu, “Feature selection for text classification with Naïve Bayes”, Expert Systems with Applications, Volume 36, Issue 3, Part 1, Pages 5432-5435, 2009.

[12] W. Zhang, F. Gao, “An Improvement to Naive Bayes for Text Classification, Procedia Engineering”, Volume 15, Pages 2160-2164, 2011.

[13] D. Glez-Peña, A. Lourenço, H. López-Fernández, M. Reboiro-Jato, F. Fdez-Riverola, “Web scraping technologies in an API world”, Briefings in Bioinformatics, Volume 15, Issue 5 Pages 788–797, September 2014.

[14] B. J. Jansen, A. Spink, S. Koshman, "Web searcher interaction with the Dogpile.com metasearch engine”, Journal of the American Society for Information Science and Technology Volume 58, Issue 5, Pages 744-755, 2007.

[15] A. Modi, A. Bhandari, K. Desai, N. Shah, “Smart Search Engine Using Artificial Intelligence” ICWET ’11: Proceedings of the International Conference & Workshop on Emerging Trends in Technology, February 2011 Pages 707–710, 2011.

[16] A. Gulli , A. Signorini, “Building an open source meta-search” WWW ’05 Pages 1004–1005, 2005.

[17] M. Kardara, V. Kalogirou, A. Papaoikonomou, T. Varvarigou, and K. Tserpes, “SocIoS API: A data aggregator for accessing user generated content from online social networks,” in Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 93–104.

[18] R. Haryanto, J. J. Lamug Martinez, “Creating Mobile And Web Filtered News Aggregator Application,” Test Engineering and Management, 83 (5), 1-1, 2020

[19] C Slamet, R Andrian, D S Maylawati, Suhendar, W Darmalaksana, and M.A Ramdhani, “Web Scraping And Naïve Bayes Classification For Job Search Engine,”, IOP Conference Series: Materials Science And Engineering, Volume 288. 2019

[20] B. Agarwal and N. Mittal, “Text classification using machine learning methods-A survey,” in Advances in Intelligent Systems and Computing, New Delhi: Springer India, 2014, pp. 701–709.

[21] J. Ofoeda, R. Boateng, and J. Effah, “Application programming interface (API) research: A review of the past to inform the future,” Int. J. Enterp. Inf. Syst., vol. 15, no. 3, pp. 76–95, 2019.

[22] J. Canavilhas and I. Satuf, “Who brings the news?: Exploring the aggregators apps for mobile devices,” in Advances in Multimedia and Interactive Technologies, Hershey, PA: IGI Global, 2016, pp. 220–238.

[23] R. S. Crema, G. N. N. Neto, and P. Nohama, “Metasearch: A web-based application to perform systematic reviews,” in Proceedings of the Future Technologies Conference (FTC) 2022, Volume 3, Cham: Springer International Publishing, 2023, pp. 775–785.