***Analysis of Full-Text Databases and Keyword Search in Academic Databases***

|  |  |
| --- | --- |
| **Hemanth Reddy Yanamala** | **20111855** |
| **CMP7200 - Masters Project Dissertation** | |
| **Supervisor:** | **Yuvgeniya Kovalchuk** |

# Abstract

The study highlights the details and topics associated with Analysis of Full-Text Databases and Keyword Search in Academic Databases. It is very necessary to input the correct form in the search box to obtain the desired result. Most of the time when the input entered does not meet the requirement or the preferences of the stored data then the results are not presented. It could also be taken as when the keywords entered are different from the data in the academic databases then the results shows are wrong. Spelling mistakes and the use of synonyms are some of the major issues that do not let the database give the proper and desired result. The research has been completed with the use of secondary data, and information regarding full text and keyword search in databases has been collected through previous articles, journals, and reports. The secondary data collection technique helps get the insights of the topic as the study is done with the perspective of different authors citing views on full-text databases and keyword searches in academic databases. The study also signifies the roles with proper comparison and topics portrayed in data analysis. Furthermore, the conclusion and recommendation segment depict the advancements that could help the users in the future while accessing academic databases.

Contents

[Abstract 2](#_Toc75298548)

[Table of Figures 4](#_Toc75298549)

[1. INTRODUCTION 5](#_Toc75298550)

[1.1. Background 5](#_Toc75298551)

[1.2. Problem statement 6](#_Toc75298552)

[1.3. Research Aim 7](#_Toc75298553)

[1.4. Research Objective 7](#_Toc75298554)

[1.5. Research Questions 7](#_Toc75298555)

[1.5.1. Primary Questions 7](#_Toc75298556)

[1.5.2. Secondary Questions 7](#_Toc75298557)

[1.6. Research Rationale 7](#_Toc75298558)

[1.7. Significance of research 8](#_Toc75298559)

[1.8. Research Gap 8](#_Toc75298560)

[1.9. Research Structure 9](#_Toc75298561)

[2. LITERATURE REVIEW 10](#_Toc75298562)

[2.1. Concept of databases and identify the result on the Search Engines 10](#_Toc75298563)

[2.2. Importance of the full-text database and keyword-based search on academic databases 13](#_Toc75298564)

[2.3. Utility of Full-text database and Full-text search 14](#_Toc75298565)

[2.4. Challenges Associated with Full-Text Search in Academic Database 17](#_Toc75298566)

[2.5. Comparison of Keywords based Search and Full-text Database Research 20](#_Toc75298567)

[3. RESEARCH METHODOLOGY 22](#_Toc75298568)

[3.1. Introduction 22](#_Toc75298569)

[3.2. Research Procedure 22](#_Toc75298570)

[3.3. Research Method 23](#_Toc75298571)

[3.4. Research Philosophy 23](#_Toc75298572)

[3.5. Research Strategy 24](#_Toc75298573)

[3.6. Research Approach 24](#_Toc75298574)

[3.7. Data collection 24](#_Toc75298575)

[3.8. Data Analysis 25](#_Toc75298576)

[3.9. Research Limitation 25](#_Toc75298577)

[3.10. Ethical Consideration 26](#_Toc75298578)

[4. DATA ANALYSIS 26](#_Toc75298579)

[4.1. Introduction 26](#_Toc75298580)

[4.2. Challenges associated with full-text database and search 26](#_Toc75298581)

[4.3. Proposed framework 27](#_Toc75298582)

[4.3.1 Experimental Aim 28](#_Toc75298583)

[4.3.2 Experimental Design 28](#_Toc75298584)

[4.4. Comparison of full-text searches and abstract search 32](#_Toc75298585)

[4.5. Discussion 34](#_Toc75298586)

[5. CONCLUSION AND RECOMMENDATIONS 35](#_Toc75298587)

[5.1. Conclusion 35](#_Toc75298588)

[5.2. Recommendations 36](#_Toc75298589)

[References 38](#_Toc75298590)

# Table of Figures

[Figure 1 - Overview of Full-text database (ScienceDirect, 2021) 11](#_Toc75298540)

[Figure 2 - Full-text Searching (SQL Server)(Ashraf, 2019) 15](#_Toc75298541)

[Figure 3 - Challenging factors of Full-text Database Searching 19](#_Toc75298542)

[Figure 4 - Distribution of Data Instances across each Output class 29](#_Toc75298543)

[Figure 5 - Confusion Matrix for Multinomial Naive Bayes 30](#_Toc75298544)

[Figure 6 - Confusion Matrix for Stochastic Gradient Descent 31](#_Toc75298545)

[Figure 7 - Confusion Matrix for Random Forest Classifier 31](#_Toc75298546)

[Figure 8 - Confusion Matrix for Gradient Boosting 32](#_Toc75298547)

# INTRODUCTION

## Background

Databases are used to store large amounts of data that can be easily accessed and are easy to add new data, update existing data, or delete old data. A full-text database is like an anthology that provides documents or some other information in the form of databases. A full-text database holds the entire text of journals, reports, newspapers, and articles. Databases help to keep greater data integrity and manage data efficiently. A full-text database is helpful when an individual searches for a keyword and retrieves all information related to that. With the help of full-text databases, people can analyse more relevant information and articles (Adams, 2018).

A full-text database is a compilation of documents and other important information in the form of a database where complete data is available for online analysis, printing, reading, or downloading. Apart from the documents full of texts, the full-text database also includes images, graphs, maps, and diagrams. Generally, when a full-text database views, it occurs in ASCII format or can say .txt extension but if the file is in. In pdf format, it can be concluded that it is a scanned copy of the original document. It is preferably appropriate for online study courses where the students are given study material by downloading it from the internet. Besides being used in educational institutions, the full-text database is also used in some law offices, government offices, or some agencies. For any registered individual or personnel who pays a specified fee to the viewed text, it is normally restricted to access these databases (Westergaard, et al., 2018).

With the increasing availability of full-text articles on the internet, it has become easy for everyone to search relevant articles. It does provide full-text articles rather than small abstracts due to which searching for a full-text article requires more computational resources. A full-text database is different from a bibliography as a bibliography provides an explanatory record of an item but does not provide the article itself whereas a full-text database provides full-text of a publication (Jamali and Nabavi, 2015).

The “search engines” is a mechanism of several interlinked connections that work together to recognize the piece of the web content based on the collection of word types in a search bar. Search Engines Optimization is used to enhance the content of the web pages. The three basic mechanisms for the search engines are web crawler, search index, and search algorithm. The web crawler gathers the data required to index a page correctly. Search Index is used to record the availability of the online content on the web pages. It allows an association between keyword terms and the content of the page. The search algorithm is used to calculate the grade and the quality of the web page. Keyword research is the method of finding and analysing the search terms that are entered by the people into the search engines to gather specific information. The importance of the keyword search increases as it provides valuable insights with the queries to attract the target audience. Keyword searches talk about the topics for which people concentrate about. Before starting the keyword search, it is necessary to sort out the list of topics for which research is required (Harikumar, 2018).

Another type of full-text database is the full-text journal database made available online by organizations, publishers, and groups. It is mostly used for subject searching or searching publications about a topic. A full-text journal database is largely helpful when one wants to search a large database of a very interesting journal article. A full-text journal database is different from bibliographic databases. The content used in the full-text journal database is not selected to bring all the journals together to form effective data. Also, there are very limited search facilities for full-text journal databases.

## Problem statement

Keywords play the most significant role while searching any data on the search engines. **Improper keywords, navigation speed, trouble with broken links, retrieval of irrelevant documents, and keyword stuffing** are some of the major problems that affect full-text databases and various search engines. With the help of proper and accurate keywords on different search engine platforms, better content can be created to hit the audience and the ranking of the full-text database will also increase. The search engines can take more time when entire research questions are entered on different SEO platforms including academic databases. **It will not provide the exact data and it negatively impacts the database library of the search engines** (Bhandari, et al., 2018)**.**

An improper keyword will consume more time and will not provide accurate output for the research on the search engines. Data Loading and the speed of the navigation is also the major problem that can be created while giving the improper keywords to the search engines. The broken links trouble the visitor and give a negative impact on the databases of the search engines. The improper keyword can provide an irrelevant document that can create major problems while working upon the available databases (Bhandari, et al., 2018). The duplicate content is also one of the major problems which degrade the ranking of the keyword and impact negatively the availability of the full-text databases.

## Research Aim

The research aims to explore the prowess and effectiveness of Machine Learning in corpus Text-mining for large corpuses. To this extent, the following objectives and sought after.

## Research Objective

* To finalize a dataset that is relevant to the research.
* To perform data pre-processing on the dataset to get it ready for Machine Learning and to also get insightful metadata from the dataset.
* To select appropriate Machine Learning algorithms based on the dataset observation for the experiment to prove the effectiveness of Machine Learning in corpus Text-Mining.
* To quantify and measure the effectiveness of Machine Learning for corpus Text-Mining using performance metrics to measure the accuracy of the algorithms implemented in the experiment.

## Research Questions

### Primary Research Questions

The questions that make the basis of the motivation behind this research are as follows:

1. What are effective sources of large data collections?
2. How to best make use of huge data available in corpuses?
3. Can Machine Learning be an effective tool for Corpus Text-Mining?

### Secondary Research Questions

1. What are the major issues associated with the full-text database and the keyword-related search in academic databases?
2. How to improve the keyword search performance on search engines?

## Research Rationale

The research of the search engine helps to identify the impact of full-text databases and the significant role of the keywords. The proper keyword research used in the search engines targets the audience and prioritizes the time accordingly. It also helps in the ranking of the available database in the library of the search engines. The search engines work on keywords, therefore the use of specific keywords is necessary to analyze the correct result and outcomes. The short and proper keywords give accurate results from the database library of the search engines. The understanding of the proper search engines plays the most crucial role in the daily life of the people to elaborate and identify the exact outcome. The full-text database helps in examining the words stored in every single document (Voit, et al., 2017).

One of the top useful search engines is google scholar which helps to recognize the importance of the full-text database in the search engine library. Different types of techniques and various platforms help to analyse the working criteria of several database libraries on the search engine platforms (Bensalem, 2017). The proper search helps to increase the traffic and the navigation speed of the available database. Improper keywords can increase the effective load on the search engine due to which accurate research will hamper. The research contains major issues of the keywords stuffing, which reduce the traffic and ranking of the search engine.

## Significance of research

Research is the driving force of mankind. It’s fuelled by curiosity humans become interested, ask questions, and dive into everything that is needed to know. Learning is prosperous. Without interest and study, progress would slow down to a halt and the lives would be different as it shows. Research is required for all practitioners and not just for students and academics. It is also relevant both offline and online for budding and veteran authors. Finding an interesting subject for discussion and/or writing should go beyond personal experience among professionals and scribes. Determining what the public may want to know or what researchers want others to do may be a reason to carry out research (Cai, et al.,2020). The Brain Science Trust is aware that research is important. There is no question that it is important to find out how to avoid potential cures for diseases as well as for every new finding. A study is therefore needed to determine if one’s ideas are supported.

## Research Gap

The research gap helps highlight the aspects of the research that have been valued in the present research and were lagging in the previous articles. The research depicts the concepts associated with searching into the academic database through the use of proper keywords. The methods through which the analysis could be done of full-text databases are also portrayed in the study. Furthermore, the concepts associated with the utility of full-text database and full-text search are also valued as a topic of consideration in the research. The research also helps gain a modern grip over the concepts and updated vision for future researches.

## Research Structure

The research framework helps to describe the step-by-step mechanism used to conduct systemic and sequential research. This research framework comprises five chapters, which are: Introduction, Literature Review, Research Methodology, Data Analysis and Findings, and Conclusion and recommendations.

***Introduction***: This is the first chapter which offers a concise outline of the investigation. It provides the history, goal, purpose, and study questions of the research field. This chapter further identifies the rationale and importance of the thesis to explain the cause and the need for analysis. This section also provides a comprehensive description of the study structure, gaps, and limitations.

***Literature review***: This part provides the theoretical basis for the study to establish basic research expertise. An analysis of current ideas and hypotheses is carried out to establish previous insights and test new results with the help of the literature analysed. This literature review is based on secondary data gathered. Research papers, magazines, books, and case studies written as secondary sources

***Research Methodology***: This is the third chapter of the study and includes the strategies and methods used for research. This segment focuses on the research methods chosen. The section describes data collection techniques and processes as well as analysis philosophy, approach, procedure, and strategy awareness. This section selects and elaborates on the technique used in data analysis (Popova, et al.,2017).

***Data Analysis and Findings***: This is the most important research section because it relates to the assessment of research results. This section uses the chosen data analysis approach to perform the data analysis. The findings of the research construct the results of the data analysis and interpretation. This chapter is designed to achieve the objective set out in the first chapter. For efficient understanding, the results of data analysis are graphically displayed.

***Conclusion and recommendations***: This is the final part of the report which provides a comprehensive overview of research results. This chapter examines whether or not the research goals have been achieved. The conclusion generalises the results of the study for the broader population. In this chapter, the recommendations are also proposed based on the results of the data analysis and of different research constraints and assumptions.

A detailed summary of the analytical findings is provided in the final chapter of the paper. The recommendations are also proposed in this chapter based on the results of data analyses and various research restrictions and conclusions.

# LITERATURE REVIEW

## Concept of databases and identify the result on the Search Engines

Under the reflection of the concepts proposed by Rovira, et al., (2018), keywords help in ranking the business and other platforms up in the search engine. These are the set of certain words that are written together are injected into the content which is available over the internet. The keywords could also be considered as the words or phrases which are used by the public most commonly to look for the relevant content. The selection of suitable keywords can help in a certain manner, considering a scenario when it is required to design a website or to update the old website. Keyword research helps in editing the architecture of the present website and also in making suitable copyrights for the new website. In most cases, the primary requirement of designing a website is to get traffic to the business, so the selection of the keywords is based on the content and the locality.

SEO (Search Engine Optimisation) helps the companies to get the website ranked above the competitors. The process that involves research for keywords has a first level, where the architecture of the site is made. Through this step, the user experience of the people visiting the website is enhanced. Furthermore, discovering new and better keywords helps in separating the services or the products that overlap. A business must have a website that can get ranked with certain keywords. There must be a focus on dedicated words, because if the ranking is done with multiple keywords, then it could lead to the overlapping of similar websites. Selection of the keywords and getting ranked depending upon a specific niche is very necessary for the website to stand with identity. According to Yang. Et al., (2018), the keyword research should also make an influence over the navigation of the website, content, and URL naming for each page. For every primary page name, navigation title, the analysis of the keyword helps in understanding the terms people search for. It is the practice that is considered under the best use of SEO so that appropriate keywords could be selected and then selected keywords could be placed in the content.

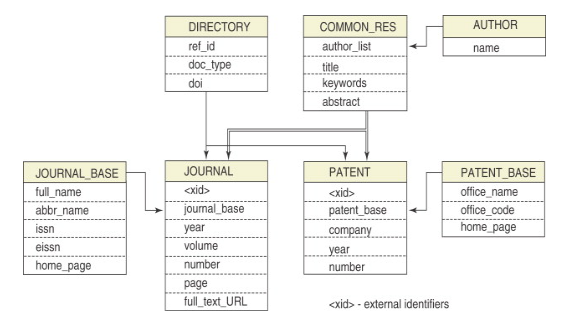


Figure 1 - Overview of Full-text database (ScienceDirect, 2021)

Content optimisation is another major factor that helps justify the importance of keywords, SEO requires a constant look after in the form of maintaining and it is an ongoing process. According to certain reports, it is also observed that the algorithms of google are changed multiple times a year, the report depicts a number that the algorithm is changed around 400-500 times in a year. Most of the changes that are adapted in the algorithm are minor and do not require major involvement, so the impact on the ranking of the website of content is not done at a major level. Optimisation also defines the requirement that with the help of routine check and maintenance in the website the content always ranks above in the search results. These practices help in understanding the need of looking after the results so that the content could be levelled above the competitors in Google results. It could further be related that content optimisation for the keywords is one of the most important parts of practising SEO on the website. It is also recommended to maintain the optimisation of the content for every site regularly, but if it could not be done periodically then it should be done annually. Optimisation also helps understand that the common terms and phrases which are used to denote the products and services could be completely different from the phrases or keywords by which they are referred over the internet. Furthermore, the companies dealing with the optimisation of the keywords also need to consider that SEO should be done with changing description and meta titles.

Under the reflection of Heersmink, (2018), the study of a database is a crucial approach to find out the keyword research. It is observed that a full-text database provides a variety of benefits for the users and organizations to demonstrate the knowledge of keyword research. The author demonstrates that a full-text database is a collection that contains the full transcript of each linked article for internet streaming, copying, or downloading. Pictures, such as charts, maps, pictures, and infographics, are frequently provided in addition to text documents. The databases help the researcher by providing full knowledge about the research topic on a search engine. Full-text search algorithms excel in analysing enormous amounts of text fast and efficiently. This could be highly unstructured, such as a Word document, or semi-structured material, such as HTML web pages, which have some shape and information but are largely text.

The author shows that full-text databases become common among researchers and other experts to provide a systematic solution of the research topic on search engines. In the respective study, the author describes the role of databases to identify the results on a search engine. The use of databases can be done over the search engine of a web portal which is an online tool for the user’s search query (Naqvi, et al., 2017). A full-text database searches and analyses all the keywords throughout every recorded document to match research criteria. With the help of databases of research, the users can access search many things over search engines such as Information created by governments, companies, groups, and individuals. The databases also provide the knowledge of Freely accessible data released by a government agency and freely available articles written by a credible media organization.

In the words of Bates, et al., (2017), database searching and search engine both are capable to handle the queries of academic experts. A database search engine searches information contained in a digital database. In full text searching the data is written on the paper sheet which may be divided into different folders the concept of databases provides systematic knowledge about the searching of the research. Simple words and phrases, as well as various variants of a word or phrase, can be included in full-text inquiries of the databases. It also demonstrates much functionality in the field of keywords searching for academic experts. The author describes that the searching of full-text databases supports the collections of different documents and functions. It also provides advantages for the users to expands the digital processing.

## Importance of the full-text database and keyword-based search on academic databases

According to Boffa, et al., (2017), Users who browse for papers in full text are much more likely to discover important articles than those who just search for abstracts. This result underpins the importance of full-text databases for document classification and serves as a springboard for potential research into algorithms that take advantage of rapidly expanding digital archives. A full-text database covers the full text of the authenticated journal, book, study, or news stories. The author describes that a full-text database knows the entire article which can also include the knowledge of bibliographic information. The full-text databases have complete scanning of law papers and full-text servers in Lexis, Westlaw, and Hein, online. A full-text database is a system that maintains a large collection of texts and it also provides fast and accurate access. These two goals of the research are relatively orthogonal and have critical profit from the collection of texts.

The author demonstrates that a full-text database of research can get the entire article immediately within a database which is used by users. Full-text databases only provide the information of citation, abstract, and keywords (Fraser, 2020). According to the author, the database of the full text may contain peer-reviewed articles which are written by trustable authors such as journalists, researchers, and experts. Users can locate the knowledge they need more easily because databases offer powerful and flexible tools for filtering results.

Under the reflection of Tripathi, et al., (2018), the term keywords are commonly known as search results. The author highlights keywords as a word that can be entered into the search boxes of the database. Keyword or research is indeed the terms in use in ordinary routine to define the subject and reflect the main concepts of the proposed study. Without keywords researcher and experts feels more trouble for the searching of a research project. The concept of keywords is important in research that provides effective results for the user’s benefits. Choosing the most appropriate nouns from such a research question/thesis is a fast and simple method to separate keywords; all other words are insignificant. When searching for keywords, users will still get more results than if they search for phrases or sentences. The main purpose of using keywords related to the research’s topic is becoming crucial for the user's benefit. The article's keywords define the area, subfield, subject, research problem, and so on. The article's keywords define the area, subfield, subject, research problem, and so on. The author demonstrates that the use of keywords is very necessary and important for the research outcomes. It can speed up experts' ability by reducing the workload of researchers. It is obvious that without having keywords a user gets more difficult for the finding of the research. It is less time-consuming for the users.

The full-text database has many types but one of them is most common and easy to access and it is Google Scholar. Google scholar is the simplest way to search for scholarly articles and literature widely. An individual can explore works like, articles, or scholar which are related to each other. When comes to citations, personnel can check who is using its citation and can create an author profile on Google scholar. The format in which Google scholar arranges the documents are based on the most searched items, it weighs the full text of the document, its publication date, and other information, and how recently it has been cited by different authors in their citations. Accordingly, texts based on magazine articles, newspaper articles, book reviews, or editorial is not suited for Google scholar, it always links to technical reports, conference presentations, and journal articles (Tóth and Kovács, 2017).

Apart from these two full-text databases, the JSTOR database exists which contains more than 1900 academic journals of full-text. JSTOR is a non-profit organization that used the technique of the digital journal archive for its current publications. JSTOR is a very large database that contains current issues of journals and it even consists of e-books. According to sources, it is to be noted that JSTOR does not contain the current volume of journals; JSTOR has 3-5 years old journals in the name of current journals. The searches can be made by using full text or abstracts. An individual can also limit its search to particular publications (Ansari and Raza, 2018).

## Utility of Full-text database and Full-text search

As compared to a basic abstract search, searching the full text of an article will reveal a lot more detail. More content is indexed than just the findings and discussion; this information can be used to find publications that are linked in ways other than the research subject. The following example shows how full-text searches will uncover a wide variety of useful posts.it can be said according to the work of Haddaway, et. al., (2015). In terms of usability, academic databases are far more superior in provided. Online articles and records are increasingly available in the full-text archive. This study aims to explain the most significant effects of full-text searches on a search engine. Full-text searching increases research efficiency of research by directly increase search efficiency.

According to the author, full-text databases used a special and specialized vocabulary to build and use them. The full-text search system enables the researchers to find any relevant information relatively easily. The internet now has more information than any other medium of information. While Google Scholar, ERIC, PubMed useful tool for locating relevant literature, they cannot substitute conventional scholarly citation databases or manual searches for grey literature. The author further states that compared to the abstract search system full text will enable the researchers to specify certain elements of research they want to reach, where abstract search follows a more complicated path. Hence it can be recommended according to the article that improves full-text search will improve the search effectiveness. The issues like the limited amount of search results shown and the advanced search facility's incomplete Boolean process can be remedied with the help of full-text database systems. Furthermore, since the academic literature misses a large percentage of grey literature which reflects that it is essential to adapt better protocols (ZHANG, et. al., 2015).

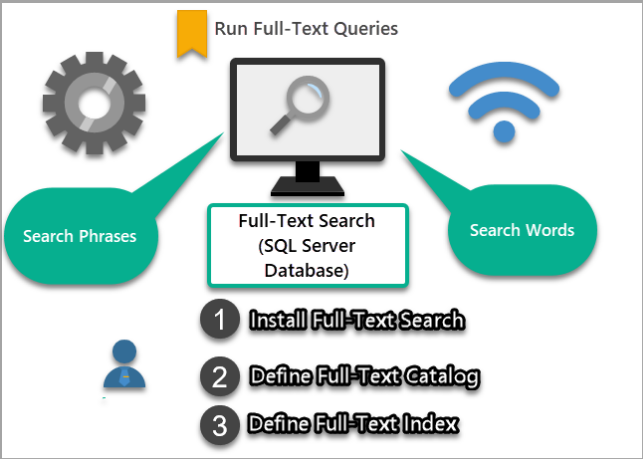


Figure 2 - Full-text Searching (SQL Server)(Ashraf, 2019)

The work of Erb and Sica, (2015) can be used to understand the effectiveness of full-text search in comparison to abstract search. The author explains how a full-text analysis archive can instantly retrieve the whole article inside a database that is used by users. Full-text databases only provide the information of citation, abstract, and keywords. According to the source, the full-text index could include peer-reviewed publications written by reputable writers such as journalists, scholars, and experts. Databases have efficient and scalable tools for filtering results, making it easier for users to find the information they need. The full-text databases offer short and reliable access to a vast selection of text material. Full-text search engines are equipped relevance ranking system that enables the determination of the best result for any search query. Another advantage of full-text search engines is that data can be stored and retrieved in original form. The full-text search engines work with an indexing approach for a high volume of textual data. One of the main reasons for using a full-text database is that it allows queries to span over multiple source repositories. When applications require highly flexible search queries, a full-text search engine delivers better performance and search results. A full-text archive helps users looking for the complete text and enables them to find important posts rather than just abstracts. This aspect of a full-text archive affirms the importance and provides a basis for potential research scholars with an algorithm that benefits from the ever-increasing digital archives. The author further illustrates the essential need to improve algorithms and structures for distributed text retrieval. This is highly useful since full-text papers are considerably longer than abstract articles and can demand more computing resources.

The research conducted by Jamali and Nabavi, (2015), elaborates on the utility of a full-text archive. Google Scholar is an international platform for scholars and researchers to store and find relevant scientific literature. Every individual who uses the search engine get the most relevant and complete information available and this must be where the utility of full-text search comes into the picture. Since Google scholar is a text database every search query must inspect complete text data to find the most relevant results. The author further states that full-text search engines grant the ability to search for the content beyond just abstracts. Finding the right information at right time can be very useful for researchers. The query tools used in full-text search engines are more flexible and feature-rich which are sophisticatedly designed ranking capabilities to fetch the best document. Using elements like prior interaction and queries, full-text search engines allow relevancy boost in real-time (ZHANG, et. al., 2015).

Under the reflection of the research conducted by Zientek, et. al., (2018), it can be understood that a full-text archive may be more accurate when searching for a large database or a database often with an extensive range of topics. The content of the full-text databases does not enable the disassembly of important journals in the field. A full-text search database will not only be limited to the bibliographic database. In this system, information cannot be searched using historical periods and relevant options Another important aspect here to notice is that full-text searches do not utilise the machismo of regulated and controlled vocabulary. Since full-text articles are getting faster acceptance on the internet and on websites, it has now become relatively easier to access full-text material. The author explains how it has become faster and effective to get to the most important articles than to look for abstracts only. More computer power can be derived with full-text databases; however, full texts are considerably longer than abstract documents. The theoretical disadvantage of scanning the whole text is that there can be the phenomenon of noise.

According to Halevi, et. al., (2017), it can be learned that full-text databases are very effective in mitigating the major issues faced in the regular search process. The increasing demand for faster and improved search engines has contributed to the demand for full-text databases. The main issue faced by search engine users was that their limit reaches only the abstracts of reports. A full-text search engine with full-text databases allows researchers to use overall existing data in the database. Containing the collection of statistics and data in the form of storage is a known identifier of full-text databases. Such a data set includes the full-text guide which helps users to browse the available material to download, print, and view. The content available on Google scholar is almost completely text and the archive also includes photographs, text records, charts, graphs, and diagrams (ZHANG, et. al., 2015). Search results in ASCII formats such as.jpg extension and.net text file extension can be shown while the user can also scan documents from the full-text database.

The author reflects on the high utility of a full-text database and how it can be useful to incorporate the technology in the field of research and analysis. The faster and effective availability of critical articles is an important part of every research, being a collective source of all scholarly and research work Google scholar requires delivering the best services to the users.

## Challenges Associated with Full-Text Search in Academic Database

According to Lemos and Joshi, (2017), it could be stated that different factors lead to the issue and challenges in the full-text search in the academic database. In consideration of some of the most impactful issues, the synonym issue is at the top. This issue takes place because different people can express their words or feelings in different words. It could lead to a situation or scenario where the system finds it hard to understand the requirement of the user and then the results which are portrayed do not meet the requirement of the person. The synonym issue is further divided into different aspects, like a true synonym. It is taken two or more two words that deliver the same meaning. Depending upon the origin or the practices which are followed in the public the selection of the word is done. It is a completely natural process as people enter the word in the search bar which is very common to their knowledge.

In full-text searching when the person enters a phrase to get the results then if the word or the sentence meets the data available over the internet the server helps the person with the results. But if the things are not scaled and the details entered by the person do not meet the data available over the internet then the things do not go as predicted and the results shown to the person could be different or the system could show no results. The author has depicted the information through an example, if any person enters “leprosy” then all the documents available over the internet with the word or phrase will be shown to the person. But if the person enters the same thing but with a different name like, “Hansen’s disease” then the results shown to the person will be limited. The reason behind this thing is that leprosy is a common word that is known to all, while the documents which will have Hansen’s disease will only be shown to the person if the entered details are Hansen’s disease. So, according to the author, it could be understood that things are reported as the words and details are entered into the search option.

As per the understandings of Gujari, et al., (2020), words that have the same meaning or which sound in the same manner could also be spelt differently. Most of this issue takes place in the context of US English and in UK English. It could be seen that the spellings of the word differ but have the same meanings. The author has highlighted the example as, “labour” and “labour” are two words that have the same meaning and have different spellings. The first word is spelt in US English while the second word is spelt in UK English. Both the words have the same meaning but have different spelling, so this is another issue that hinders the results and the details obtained from the search bar. Most of the time people search the thing with one result while the keywords or the phrase entered in the system are with different spellings. So, the results obtained are not adequate. Furthermore, there are many other scenarios as well where some spellings are common in one form of English. The author has highlighted, the in US English some of the words are spelt differently. “Donut” and “Doughnut” both mean the same and both the spellings are appreciated in US English. These issues lead to the failure of a result obtained and the results obtained could not be scaled as per the requirement.

The author has further discussed the details in the form of a shortened form of results, where the details entered by the person are different. There could be many scenarios where the person enters the detail have a different form or it is entered differently. In consideration of an example, PETA is the name of an organisation that works for animal rights. People could enter PETA in the search results while some of the people the full form as, People for Ethical Treatment of Animals. This could create an issue in the searching and depiction of the results.

Figure 3 - Challenging factors of Full-text Database Searching

In the words of Stansfield, et al., (2017), a full-text search for the research and documents provides information regarding satisfying the query and options for the users. It provides a variety of solutions for the user but faces some challenges during the time of the searching procedure. The author describes that to understand the techniques and concepts of full-text databases users must need to understand the challenges. The full-text searching is a distributed database that impacts the performance during the time of research. The author demonstrates that the challenge of variant spelling is one of the biggest challenges of full-text data search. Words that represent the same thing can be spelt differently, as in the case of British and American spelling variations. So, this is the major concern of full-text search in the academic database. Many websites develop better solutions to mitigate such issues of variant spelling. Apart from this the shortened form of terms also shows challenges for the full-text data search. In the presence of shortened form, the researcher is not able to research the exact document as per the need of the research’s topic.

Under the reflection of Iversen, et al., (2017), Full-text searching occurs when a computer compares the words in a search engine with terms in different files in a database. The search via full-text research classifies the research’s results, either by relevancy or by some feature of the individual objects returned, such as publication date. In the respective study, the author described all the chief challenges of full-text searching. The full-text searching process is unable in the proper searching of research’s facts. Sometimes some specific searchers are needed for the attribute and online searching but the full-text searching process is not able to deliver the results for a similar process so it is one of the major challenges of full-text research. The searching of the full text can also be affected due to the use of figurative language. According to the author in full-text research use of figurative language is one of the major barriers to the collection of effective information. Figurative language is defined as language that does not employ words in their literal dictionary definitions. The challenges of full-text searching are becoming major issues in the proper search of information and databases.

According to Walters, (2013), full-text searching becoming popular in between the field of academic and research. Full-text searching is involving the functioning to search online content with the help of keywords. Keyword searching is the most common benefit of full text searching with the benefits full-text searching also has many challenges which can directly impact the overall performance of the research process. Full-text searching has the challenge associated with searching through the image, the author describes that it is not able to provide better solutions for the image searching. Apart from this the full-text searching also has challenges of re-indexing which provide new headings of the research. Full-text searching has shown the challenges to shows accurate fact about the research. The keywords searching of the research shows the inability to provide searching with facts.

## 2.5. Comparison of Keywords based Search and Full-text Database Research

According to Gusenbauer, (2019), the primary goal of full-text searching is to allow users to search and obtain web documents using queries to obtain the required information and knowledge. On the other hand, the searching of keyword-based searching is completely different from the full-text research of the topic. The author described that the keywords-based search engines are extremely useful for locating information on the internet and delivering results quickly, but they suffer from a lack of understanding of the keywords and expressions used in websites, as well as the relationships between them. In the respective study the comparison of full-text databases and keywords bases research is provided The author shows that both keywords-based searchings are full-text databases are important for gaining better information over the internet but the searching is based on full-text databases is considered more effective as compared to the keywords based searching.

According to the author keyword bases research do not includes the actual meaning of keywords user are unable to provide appropriate search results based on keyword search. Full-text searching databases are considered more advanced versions of today’s web, with the use of such databases users, can provide a meaningful representation of information for both machines and humans (AlRyalat, et al., 2019). The main components of full-text searching are the use of synonyms. Keywords-based searching can be considered as the traditional way of research which provided the results in the demanding contest on the other hand the full text-based research is more useful for providing accurate and relevant information about the research topic. In keyword searching the page ranking algorithms can produce span results which may be problematic in the future while in full-text research the information is retrieved independently.

In the words of Westergaard, et al., (2018), Full-text database search provides more structural results and advantages over keyword-based databases. The base of full-text research provides more authentic results for the users. Searching based on a full-text database guarantees that the user will receive more accurate and relevant results of interpretation of the user’s information. The author described that users will get more relevant results by the use of a full-text database, on the other hand, the use of keywords based research becomes a traditional way of a research topic over the search engine. The respective study demonstrates the comparison between keyword bases searching and full-text database searching. The comparison between both research approaches can be understood by comparing their working operations. The author describes that keyword bases research does not focus on the stop words like is, and, or, how.

This is the main reason that users will not get accurate results. On the other hand, the research-based in full-text databases focuses on the stop words and punctuation marks which affect the search results on high priority. The keywords-based searching process delivers the results with the help of web pages which deliver less satisfying results for the users. Searching based on full-text databases provides the answers only to resolve the user's query. The usage of keywords does not bring up any terms or phrases that help obtain accurate results. While the Sentences or words that respond to the user's question are highlighted in the full-text database (Schmucker, et al., 2017). The author describes that both full-text databases and keywords-based research use different languages of programming to runs the operation. According to the author, keyword searching uses HTML, XML languages for metadata while full-text database research use OWL and RDF for the creation of metadata.

# RESEARCH METHODOLOGY

## Introduction

This part of the research provides the knowledge of all methods and techniques used in the research. The study of research methodology is one of the most crucial approaches of the research. This section of research discusses the analysis and collection of data and the research design. This chapter of the research covers many segments such as the research process, design of the research, research method, approach of the research, research philosophy, data collection and analysis, and research limitation (Mohajan, 2018). These all segments are necessary for the systematically processing of data and information. The methods used in this research help to answer the primary and secondary research questions.

## Research Procedure

The procedure of the research defines the step-by-step plan that is followed in the research to attain desired outcomes from the research. The procedure of this research is given below.

* Selection of Research Topic
* Formation of Aim and objectives
* Formulation of Research Questions
* Formation of Literature Review
* Selection of research methods
* Data Collection
* Analysis of Collected Data
* Formation of Conclusion and Recommendations

## Research Method

The study of the research method is a crucial approach of this chapter. It refers to all the strategies used in the research to collect and evaluate data to provide a better understanding and factual information. This section also provides in-depth knowledge for a better understanding of the research topic for the identification and addressing of the research problem. Generally, the research experts used two types of research methods to evaluate the findings of the research. These methods are known as Qualitative and Quantitative research methods (Basias and Pollalis, 2018). The qualitative approach of the research involves the collection and processing of non-numerical data to understand ideas, thoughts, views, and experiences of the research topic. The qualitative approach of the research also provides in-depth insights into the research problem. While on the other hand the quantitative research is built for collecting and analysing numerical data, it provides the data results by using mathematical approximations. The quantitative research method is mostly used to find pattern, averages, develop statistical assumptions, and create predictions based on mathematical facts. In the respective research **qualitative research method** is used for evaluating the findings of the research. The qualitative research method is selected because it is considered a more appropriate approach for delivering effective results for the generalization of secondary data.

## Research Philosophy

The philosophy of the research provides the evaluating of knowledge concerning assumption and trust. Selection of research philosophy can be considered a crucial approach because it shows whether the research employs a suitable approach, strategy, and process in its completion. Generally, numerous types of research philosophies are available for the successful conduction of the research. These philosophies are pragmatism, postmodernism positivism, interpretivism, and critical realism (Ryan, 2018). The research requirements enable the selection of suitable research philosophy. In the respective study, Interpretivism research philosophy is used to conduct the research systematically. Interpretivism research philosophy is selected because it develops new ideas for the understanding of the social world. According to the interpretivism research philosophy, the social reality is subjective, and the research experts include certain roles for the proper observations. The interpretivism philosophy is used in this study to incorporate the interest of the research expert into the research topic. The theory of interpretivism philosophy described that the research in the respective study is based on research experts. With the help of this method, research experts can select meaningful data sources for the data collection. Interpretivism philosophy also emphasizes the creative qualities of science in many ways.

## Research Strategy

The strategy of the research also described knowledge about the facts of the research topic. The strategy of research provides a plan or path which is used by the research to examine the answers to research questions presented in the introduction chapter, as well as the process for reaching the research's objectives. In general, the strategy of research is considered as the connecting bridge of methodological choice and research philosophy. A variety of research strategies are available such as Grounded theory, Experiment, Archival research, Action research, and Ethnography (Chun Tie, et al., 2019). In the respective study, Grounded Theory is used as the research strategy based on the requirement. Grounded theory is selected as the research strategy in the respective research because it is considered as a qualitative research strategy for analysing the findings of the respective research. It also refers to a collection of systematic inductive approaches. In the research, the grounded theory will provide focuses, conceptual, and abstractive theories in combination with the qualitative research method. The grounded theory of the research also produces detailed descriptions for the futures contradiction and conflict.

## Research Approach

In this research, the ***inductive research approach*** is utilised. The inductive research approach helps to attain the research aim with the help of research questions and attempts to develop the new theory instead of concluding on the existing theories. In the inductive approach, the observations and assumptions are analysed at the initial stage, and at the end, the observations are concluded as the theory. As the research is conducted with a qualitative research method along with interpretivism research philosophy, an inductive approach is most appropriate for this research.

## Data collection

Data collection is a process of maintaining the required information from relevant sources to identify the solution of the research. The collection of the data is the most significant and mandatory part of any research. The whole process is based on the collection techniques and its analysis. The primary data is a source collected from different interviews, experiments, and surveys. In the given research**, the secondary data** collection method is used to gather the information. The secondary data collection is a source gathered from articles, journals, and case studies. The secondary data collection method also provides a theoretical and practical concept that helps in the evaluation of the fundamental content. With the help of data collection methods, new and realistic information can be handled that helps to understand the topic of the research. Most of the organizations used the data collection tools to make assumptions and future probabilities for the market growth and business trends. Data collection techniques help to recognize the relevant sources for the research. Data collection also helps to recognize the possible outcomes of the research (Trinh, 2018). The secondary data used in the given research plays an important role to conduct the entire research.

## Data Analysis

Data Analysis is the method of cleaning, modifying, and examining the theoretical data. It is a systematic arrangement of analyzing theoretical and logical techniques that support the inspection and transformation of the data. The data analysis practice helps to find better outcomes of the research. Various processes and techniques are available that can analyze the quantity and quality of the research**. In the given research, the content analysis method** is used to explain the findings of the research. The content analysis is a research tool that plays a crucial role to identify the presence of certain words, themes, and concepts of giving research. **The main focus of content analysis is to identify the qualitative data of the research.** The qualitative data research of the content helps to offer an objective evaluation that provides an accurate method to analyze the data. The content analysis method requires a systematic collection of the data from interviews, articles, speech, journals, and content on the web pages. The content analysis data techniques offer a satisfactory result that provides the outcomes and findings of the research (Roller, 2019). The practice of analyzing the data helps to find the positive outcomes for the research work.

## Research Limitation

The limitations create the boundary for the research to conduct properly. The limitations of the secondary data develop a barrier for the understanding of the primary data collection methods. it creates an obstacle that hampers the presence of the primary data. the presence of the authentic data can be disturbed using the secondary sources for collecting the data. the qualitative research is composed based on assumptions, that can impact negatively the research work. The research is supposed to be finished in a limited time, which can restrict the people to manage the authentic source for collecting the data. The findings and results are based on secondary sources, which cannot easily describe the possible findings and every single research can be directly dependent upon the assumptions. The research of the secondary source will work as a barrier for taking strong- decisions (Queirós, et al., 2017). The research consists of a minimum collection of the data, which is another limitation that can disturb the possible outcomes.

## Ethical Consideration

The research is conducted with the consideration of research ethics such as data integrity, confidentiality, authenticity, privacy, independent review, and scientific validity. The collected data is not manipulated during the research process to acquire the desired results. The data is collected from the academic databases and only Peer reviewed, verified, and published journals, articles, and research papers are considered as data so that the validity and reliability of the research can be maintained. The research is conducted without any prejudice and partisanship shows that the findings of the research can be generalized for the wider population.

# EXPERIMENTAL METHODOLOGY

## Introduction

The following section of the report will focus on presenting the analysis of data collected during the research. The main purpose of this segment is a clear understanding of vulnerabilities and improvement in the system. Full-text databases and full-text searches are two important aspects of academic literature stored online. A suitable framework will be identified and present in this section. A comparative analysis of abstract search or keyword with the full-text search will also be included in the section.

## Challenges associated with full-text database and search

With help of data collected, it has been identified that full-text databases are not perfect in every aspect and there are several areas where the framework lacks. The major advantage of the full-text searches is that it allows the user to extend the search beyond just abstracts (Aye and Thein, 2012). Having the instant result to a vast text database is a complicated task, below are the challenges associated with the full-text database search.

* Scalability: The amount of academic data stored online is considerably high and also increasing exponentially which means any particular storage capacity can’t be enough. On the other hand, increasing the capacity of data storage requires higher computational power for searching.
* Proximity: Proximity refers to the relevance of search results present by the search engine. An enormous text database will have many repetitive scenarios and the result might not come as expected. Word vectors play an important role success and performance of a text database.
* Abbreviations, synonym, and homonyms: Text searches are very sensitive towards the use of keywords and recent development in the lingual structure like increased use of abbreviations and common names play an important role. Identification of the most relevant document regardless of abbreviation and synonyms use is a challenge for full-text databases(Xie, et al., 2019).
* Reverse image search: Reverse image search refers to the user using an image as a search query to trace it back to the relevant document. Full-text databases effectively deliver text results but reverse image searches are a challenging element.
* Mistypes and updated data: A full-text database offers effective means for text search but in the case of misspelt words the results can be very different and irrelevant. A full-text database requires updating and all data stored must be effectively stored. Maintaining spelling mistakes and updates is a challenge for full-text databases and searches.

## Proposed framework

The scope of this research is vast and cannot be captured in just single research. However, with the experiments planned and executed in this section, the author hopes to capture and convey the prowess of text mining when it comes to full-text databases and corpora. The aim of this research as mentioned previously is to analyse keywords and full-text databases related searches in academic databases and identify challenges associated with full-text search. The process of analysing text in large text databases and corpora and deriving information from it is referred to as Text Mining. The generalized aim of this research can be specified as the analysis of Text Mining techniques, i.e., the procedures, the techniques involved, the methodology and its limitations. To this extent, this research will conduct an experiment involving text-mining, Machine Learning and Natural Language Processing to express the prowess of text-mining by predicting the ratings of products on Amazon by analysing the Amazon Product Review Corpus.

### 4.3.1 Experimental Aim

This experiment aims to support the main research aim by shedding light on the prowess of text mining through the prediction of product ratings on Amazon. To this extent, the following objectives are fulfilled:

* Dataset selection
* Dataset pre-processing
* Machine Learning algorithm selection
* Performance Metrics
* Result comparison

### 4.3.2 Experimental Design

The experiment consists of the following modules:

#### Data Pre-processing

The pre-processing of data in the context of this research involves the removal of characters in the corpus that are not English alphabets, normalization to lowercase, removing the stopwords and stemming. Stopwords can be considered as useless words that introduce unnecessary computation expense and affect prediction results. These words can be helping verbs, articles, parts of speech, etc. Stemming is the process of reducing the word to its basic form i.e. if the word is “changing”, it will be reduced to “change” which is its basic form. The corpus is extremely large i.e., the number of data instances in the corpus is 413,000. Since the experiment serves as a proof-of-concept made to support the research aim, only 20,000 random data instances will be employed from the total corpus to execute the code faster. The dataset is also split into a training dataset and test dataset for the same purposes.

TF-IDF transformation is applied to the dataset. TF-IDF or term frequency-inverse document frequency is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. It is a weight assigning technique that makes it easier for the algorithms to identify patterns that lead to a specific output class. The best features are filtered using the chi-squared statistical metric and selected for the implementation of Machine Learning algorithms for the prediction of rating.

#### Data visualization

There are five ratings in the Amazon rating system, i.e., five output classes. This section is to observe the distribution of the finalized 20,000 data instances for each output class. This is represented graphically as follows:

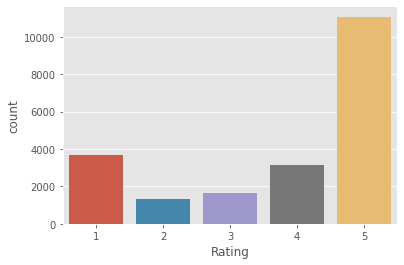


Figure 4 - Distribution of Data Instances across each Output class

The graph above shows that the trend of the distribution of data instances for each output class resembles an inverse parabolic behaviour i.e., the data instances are more concentrated towards the extreme ends of the output spectrum when compared to the middle section. Also, the concentration trend of the data instances appears to be a little right-skewed. The output class associated with a 5-star rating has the most data instances associated with it.

#### Machine Learning

The paradigm to be followed in Machine Learning i.e., whether the Machine Learning algorithms will belong to the regression paradigm or classification paradigm is decided based on the dataset and its metadata. Upon observing the dataset, it can be deduced that the dataset has five output classes and all the features correspond to those five classes. This means that the paradigm for this research experiment will be the multiclass classification paradigm. There are five algorithms finalized for this purpose. The algorithms are selected in such a way that each aspect of the classification paradigm algorithms. i.e., probabilistic, statistical and deep learning algorithms can be covered to observe the full range of Machine Learning at work. The algorithms selected are as follows:

* + Multinomial Naïve Bayes
  + Stochastic Gradient Descent
  + Random Forest Classifier
  + Gradient Boosting
  + Deep Dumb Neural Network

#### Performance Metrics

The performance metrics used for this research experiment are the confusion matrices, the f1-scores and the accuracy scores for the test dataset. The results for each algorithm are as follows:

* + Multinomial Naïve Bayes

Accuracy Score: 66.09%

F1-score: 66%

Confusion matrix:

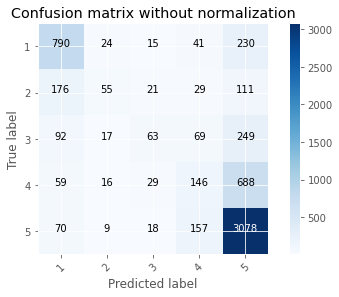


Figure 5 - Confusion Matrix for Multinomial Naive Bayes

* + Stochastic Gradient Descent

Accuracy Score: 68.69%

F1- Score: 69%

Confusion matrix:

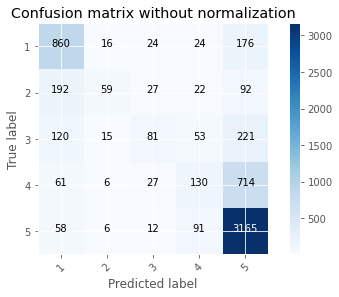


Figure 6 - Confusion Matrix for Stochastic Gradient Descent

* + Random Forest Classifier

Accuracy Score: 69.12%

F1-Score: 69%

Confusion matrix:

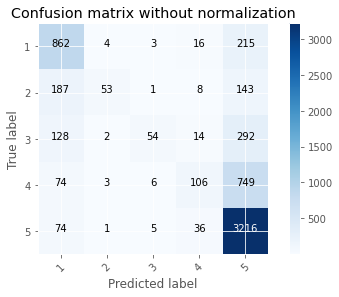


Figure 7 - Confusion Matrix for Random Forest Classifier

* + Gradient Boosting

Accuracy Score: 64.61%

F1-Score: 65%

Confusion matrix:

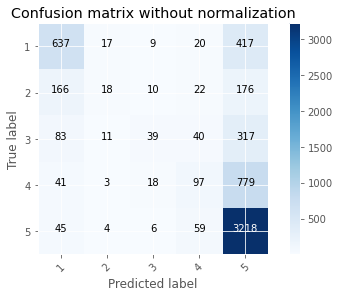


Figure 8 - Confusion Matrix for Gradient Boosting

* + Deep Dumb Neural Network

Accuracy Score: 68.1%

## Comparison of full-text searches and abstract search

The most noticeable difference is between a complete and an abstract article is available data for processing. Abstract databases only have limited storage requirements compared to full-text databases. A number of the methods for information search and retrieval nowadays based on the “group of words” approach. The text is stored as a weighted feature vector with each term as a characteristic, distribution of the probability of terms used for language modelling techniques. The main differentiating factor is that these models draw features from statistical features like a recurrence of the terms and frequency of relevant documents.

However, other aspects are also vital to take into account for the comparison. An unnecessary amount of data is not effective for users and that signifies the importance of improved search results. The best way to enable users to determine what articles they want to read is to obtain a more thorough article and image access, rather than answer user inquiries directly with inadequate results. Instead, the information produced by the system must be considered when text is extracted. Complicated information demands such as the ones encountered by scientists and many times image results are not up to mark. One of the main issues associated with full-text databases is the issue of scalability due to its exponentially increase storage requirements (Bhavani's Portfolio, 2021). The abstract databases are smaller in size and scalability is not much of an issue in the process. Every search query initiated in full-text search will have to go through a lot of algorithmic processing and it means more data will be used. The major aspect of full-text search is that it enables the researcher to effectively locate critical documents with the best possible search precision, whereas the abstract search is more of a specified and generalised process.

The implementation of random query generators and indexers increases the utility and effectiveness of the full-text databases. The relevance of the search result is another differentiating factor due to the smaller database and the keyword-focused search abstract database provides relevant results on top. The full-text databases are different as they require additional computation to shortlist the recommendations. In comparison to keyword-based databases, full-text database search yields more structural findings and benefits. The whole-text research base provides users with more authentic findings. A complete text search ensures that the user receives more precise and appropriate data interpretation results. On the other hand, the usage of keyword-based research is becoming a typical approach for study topics through the search engine and consumers will obtain more relevant results by using a complete text database.

The data analysis describes that keyword searches are incredibly effective in the location and delivery of data on the Internet, but they do suffer from incomprehension and a lack of relative understanding of the terms and phrases used on websites. Both keyword searches are full-text databases, which are vital for better online information but are more successful than keyword-based searches in a full-text database (Elsayed, et al.,2019). The comparison of full-text search and abstract search indicates that both approaches have their potential benefits but it is very critical that synchronization of algorithms should be taking place. To search a huge database is a complex task since language and words vary according to personal and regional requirements. This indicates that the amount of processing and pre-defining of queries required by a full-text search will be much more than abstract searches. The use of synonyms and homonyms is always a crucial matter for text search and queries, abstract searches are compact and usually not troubled by the issue as much as full-text searches. It is vital to blend these types to make the search engine effects. Moreover, abbreviated words or abbreviations are often a huge challenge while looking for whole material (Cebollero, et al., 2015). Each user and investigator cannot exactly seek a specified document if the text filter and the analyser are not provided in any specific way

## Discussion

With aid to the details proposed in the study and the concepts highlighted it could be stated that full-text databases are more efficient in helping the learner and yielding better results than a keyword search. In most cases, the details are gauged depending upon the content availed through search. Keyword search in academic databases helps with the results but it makes it difficult for the learner to get the relevant information. If the learner wants to get information on human anatomy and searches with some word that is not directly present in the content but has a similar meaning to it. In this condition, the learner will not be able to get proper results, and then it would make things difficult. While, if the learner enters a keyword and the same is mentioned in the research paper, article, journal then the results are presented (Gusenbauer, 2019).

Several other challenges are faced in the keyword search in academic databases, most of the time the person does not correctly enter the details. Spelling mistakes could be an example of the issue, in most cases, the results are delivered through proper searches. While, when the input is fed with improper spelling then the system does not give proper and desired results. There are many different academic databases, but not all could help with proper results when entered the wrong spelling. There are many different platforms like IEEE Xplore and ScienceDirect but the algorithms of all the databases are not similar to Google Scholar. Google Scholar has advanced working algorithms that help in showing the results if the spellings are improper, while other databases fail to do so. While full-text databases help the learner in getting the results easily (Shafiq and Wani, 2018). Finding out results becomes efficient when academic databases are looked into with proper text phrases.

# CONCLUSION AND RECOMMENDATIONS

## Conclusion

**Concerning objective 1**, it could be concluded that academic databases store a huge pile of information and to get the best results out for the purpose it is required to input the right keyword. In consideration of full-text search and keyword-based search, it could be stated that both types of methods help extract the results from the database. When the keyword is entered in the search bar then the results which are shown to the individual are associated with the work and have the same keyword. While for the full-text keyword search it could be considered that the exact data needs to be present in the article or journal then it reflects the desired outcome. For better results the individuals need to possess good knowledge of the keywords, there could be scenarios where the information entered does not match the keyword and the desired results could not be gained by the individual.

**Concerning objective 2**, it could be concluded that different academic databases help the individual in understanding the topics and gaining insights into the topics. Some of the most preferred academic databases which are used by the individuals and learners are Google Scholar, ScienceDirect, IEEE Xplore, and many more. With aid to the concepts proposed in the study, it could also be brought to the conclusion that Google Scholar has been rated as the most used academic database. It possesses knowledge of all the topics and the learner could find different articles, journals, papers, and other things on the portal. The algorithm that helps the learner in yielding the output is also finely designed, the search box could predict the best words to complete the incomplete keywords entered by the learners.

**Concerning objective 3**, different issues are associated with full-text search options. It could be concluded that some of the most hindering issues are synonyms, abbreviations, and issues in spellings. These issues could also be considered to impact the complete search profile of the learner. Sometimes, a single spelling issue could lead to portray a different result. Moreover, there are also chances that the results which are desired could not be fetched because the phrase entered in the search bar has a different word while the journal has the same content but with a synonym. These are some of the most impactful issues which do not let the result yield proper results to the learner. It could also be concluded that the research needs to be done with the help of different keywords, so the learner could also get the results depending upon different keywords entered.

**Concerning objective 4**, it could be concluded that the issues which are faced in getting proper and desired results while searching for the content in the academic database could be eradicated with the help of proper skills and knowledge of concepts. The learners should know that the results could be cultivated out from the databases if the use of different keywords is done. If proper keywords are not used then the results gained could not be taken as sustainable. Moreover, the use of proper keywords should also be known to the learners. When the results could not be gained with the use of different words and different keywords then the learner should also consider looking for different platforms or different academic databases. It is a fair practice to look for the proper journals, and if the results are not gained as desired then the input parameters should be changed by the learners. Use of different keywords, use of proper phrases, and getting results in proper context is also very important.

## Recommendations

In the academic database, full-text database search and keyword search have brilliant potential to provide significant search results. These type of keyword searches consumes less amount of time, energy and computing. As databases hold a large amount of information, hence it is necessary to use a filter and keyword finders. Although using inappropriate keywords, navigational speed, irrelevant words with broken links may affect search results badly (Bramer, et al., 2017). To avoid these and inappropriate search results which are affected by keywords search. Important measurements should be taken to enhance the searching efficiency by using proper indexing of text, choosing the correct set of keywords, using synonyms, by providing relevant words according to associated data matches. Full-text database searches and keyword searches have a significant role in time saving and online indexing. The search results should be **organic** and should not be influenced by any paid or non-paid advertising network. They should be ranked according to their relevance to the search term. To avoid any negative impact on the academic data library, proper search engine optimisation is needed. By using relevant terms and appropriate algorithms, one can avoid **data loading** and **navigational speed**. By eliminating the use of **duplicate content** and maintaining the uniqueness of academic data, degradation of keyword ranking could be maintained. By organising the content in a **topic cluster** will help to develop a mindful strategy that limits duplication of content. Having a general look at the **taxonomy** of academic sites or data is also a viable option. **Canonical tags,** **meta-tagging**, **parameter handling**, **duplicate URLs**, and **redirects** should be used to avoid any possibility of data duplication of academic content (Ahsan, et al., 2017). By using appropriate fonts, complete URLs, avoiding punctuations, and proper testing, the problem of broken links can be eliminated easily. To avoid the problem of **website and content navigation** following steps should be taken.

* By using descriptive labels.
* Limiting the number of list items to seven.
* By mentioning the important stuff at the beginning of the page.
* Avoiding the drop-down menus
* By following the best practices of web navigation.
* Checking analytics.

By avoiding data loading, text navigational speed blockages, and usage of inappropriate keywords. The efficiency of keyword search can be improved significantly.

# References

Adams, S., 2018. Is the full text the full answer? –considerations of database quality. *World Patent Information*, *54*, pp. S66-S71.

Ahsan, M.M., Chowdhury, F.Z., Sabilah, M., Wahab, A.W.B.A. and Idris, M.Y.I.B., 2017, July. An efficient fuzzy keyword matching technique for searching through encrypted cloud data. In *2017 International Conference on Research and Innovation in Information Systems (ICRIIS)* (pp. 1-5). IEEE.

AlRyalat, S.A.S., Malkawi, L.W. and Momani, S.M., 2019. Comparing bibliometric analysis using PubMed, Scopus, and Web of Science databases. JoVE (Journal of Visualized Experiments), (152), p.e58494.

Ansari, N.A. and Raza, M.M., 2018. Usage of JSTOR Database Among Research Scholars in the Faculty of Social Science, Aligarh Muslim University. DESIDOC Journal of Library & Information Technology, 38(3).

Aruleba, K.D., Aremu, D.R., Oriogun, P.K., Agbele, K.K. and Agho, A.O., 2015. Evaluation of Full-Text Search Retrieval System.

Ashraf, H. 2019. Implementing Full-Text Search in SQL Server. [ONLINE]. Available at: https://codingsight.com/implementing-full-text-search-in-sql-server-2016-for-beginners/ [Accessed On: 6-06-2021].

Aye, K.N. and Thein, N.L., 2012, January. Efficient indexing and searching framework for unstructured data. In Fourth International Conference on Machine Vision (ICMV 2011): Machine Vision, Image Processing, and Pattern Analysis (Vol. 8349, p. 83493F). International Society for Optics and Photonics.

Basias, N. and Pollalis, Y., 2018. Quantitative and qualitative research in business & technology: Justifying a suitable research methodology. Review of Integrative Business and Economics Research, 7, pp.91-105.

Bates, J., Best, P., McQuilkin, J. and Taylor, B., 2017. Will web search engines replace bibliographic databases in the systematic identification of research? The Journal of Academic Librarianship, 43(1), pp.8-17.

Bensalem, E., 2017. The impact of keyword and full video captioning on listening comprehension. *Journal of Teaching English for Specific and Academic Purposes*, *4*(3), pp.453-463.

Bhandari, R.S. and Bansal, A., 2018. Impact of search engine optimization as a marketing tool. *Jindal Journal of Business Research*, *7*(1), pp.23-36.

Bhandarkar, S. and BN, N., 2020. A Full-Text-Based Search Algorithm vs Elasticsearch. Studies in Indian Place Names, UGC Care Journal, 40(74), pp.2168-2171.

Bhavani's Portfolio, 2021.Challenges in full-text search systems. [ONLINE]. Available at: https://bhavaniravi.com/blog/challenges-in-full-text-search/ [Accessed: 7 June 2021].

Boffa, D.J., Rosen, J.E., Mallin, K., Loomis, A., Gay, G., Palis, B., Thoburn, K., Gress, D., McKellar, D.P., Shulman, L.N. and Facktor, M.A., 2017. Using the National Cancer Database for outcomes research: a review. JAMA oncology, 3(12), pp.1722-1728.

Bramer, W.M., Rethlefsen, M.L., Kleijnen, J. and Franco, O.H., 2017. Optimal database combinations for literature searches in systematic reviews: a prospective exploratory study. *Systematic reviews*, *6*(1), pp.1-12.

Cai, J., Hwang, S., Hiebert, J., Hohensee, C., Morris, A. and Robison, V., 2020. Communicating the significance of research questions: Insights from peer review at a flagship journal. *International Journal of Science and Mathematics Education*, pp.1-14.

Cebollero, M., Natarajan, J. and Coles, M., 2015. Full-Text Search. In Pro T-SQL Programmer's Guide (pp. 317-345). Apress, Berkeley, CA.

Chun Tie, Y., Birks, M. and Francis, K., 2019. Grounded theory research: A design framework for novice researchers. SAGE open medicine, 7, p.2050312118822927.

Elsayed, A., Hazman, M. and Ellakwa, S.F., 2019, December. Integrating the expert system with a full-text search to solve growers' problems. In 2019 15th International Computer Engineering Conference (ICENCO)

Erb, B. and Sica, R., 2015. Google Scholar: Flagship Database for Literature Searching or Helpful Auxiliary? *The Charleston Advisor*, *17*(2), pp.47-50.

Fraser, L.H., 2020. TRY—A plant trait database of databases. Global change biology, 26(1), pp.189-190.

Gujari, R., Sahithi, C. and Parihar, N., 2020. Academic search engine optimization. *Journal of Critical Reviews*, *7*(13), pp.1373-1376.

Gusenbauer, M., 2019. Google Scholar to overshadow them all? Comparing the sizes of 12 academic search engines and bibliographic databases. Scientometrics, 118(1), pp.177-214.

Haddaway, N.R., Collins, A.M., Coughlin, D. and Kirk, S., 2015. The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS one*, *10*(9), p.e0138237.

Halevi, G., Moed, H. and Bar-Ilan, J., 2017. Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation—Review of the literature. *Journal of informetrics*, *11*(3), pp.823-834.

Harikumar, C., 2018. Keyword analysis: Using text analysis for search engine optimization. International Journal of Pure and Applied Mathematics, 118(5), pp.565-571.

Heersmink, R., 2018. A virtue epistemology of the Internet: Search engines, intellectual virtues and education. Social Epistemology, 32(1), pp.1-12.

Iversen, C.M., McCormack, M.L., Powell, A.S., Blackwood, C.B., Freshet, G.T., Kattge, J., Roumet, C., Stover, D.B., Soudzilovskaia, N.A., Valverde‐Barrantes, O.J. and van Bodegom, P.M., 2017. A global Fine‐Root Ecology Database to address belowground challenges in plant ecology. New Phytologist, 215(1), pp.15-26.

Jamali, H.R. and Nabavi, M., 2015. Open access and sources of full-text articles in Google Scholar in different subject fields. Scientometrics, 105(3), pp.1635-1651.

Lemos, J.Y. and Joshi, A.R., 2017, February. Search engine optimization to enhance user interaction. In *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)* (pp. 398-402). IEEE.

Mahajan, H.K., 2018. Qualitative research methodology in social sciences and related subjects. Journal of Economic Development, Environment and People, 7(1), pp.23-48.

Naqvi, S.N.Z., Yfantidou, S. and Zimányi, E., 2017. Time series databases and influxdb. Studienarbeit, Université Libre de Bruxelles, p.12.

Patil, N.S., Kiran, P., Kiran, N.P. and KM, N.P., 2018. A survey on graph database management techniques for huge unstructured data. International Journal of Electrical and Computer Engineering, 8(2), p.1140.

Petkovic, D., 2018, September. Full-text search extensions for JSON documents: Design goals and implementations. In International Conference: Beyond Databases, Architectures and Structures (pp. 283-293). Springer, Cham.

Popova, O.B., Romanov, D.A., Loyko, V.I. and Evseeva, M.A., 2017. Methodical significance of activities research results. *Sociological Studies*, *3*(3), pp.79-87.

Queirós, A., Faria, D. and Almeida, F., 2017. Strengths and limitations of qualitative and quantitative research methods. *European Journal of Education Studies*.

Roller, M.R., 2019, September. A quality approach to qualitative content analysis: Similarities and differences compared to other qualitative methods. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 20, No. 3).

Rovira, C., Guerrero-Solé, F. and Codina, L., 2018. Received citations as the main SEO factor of Google Scholar results ranking. *El profesional de la información*, *27*(3), pp.559-569.

Ryan, G., 2018. Introduction to positivism, interpretivism and critical theory. Nurse researcher, 25(4), pp.41-49.

Schmucker, C.M., Blümle, A., Schell, L.K., Schwarzer, G., Oeller, P., Cabrera, L., von Elm, E., Briel, M., Meerpohl, J.J. and OPEN consortium, 2017. Systematic review finds that study data not published in full-text articles have an unclear impact on meta-analyses results in medical research. PloS one, 12(4), p.e0176210.

Sciencedirect, 2021. Full-Text Database - an overview. [ONLINE]. Available at: https://www.sciencedirect.com/topics/nursing-and-health-professions/full-text-database [Accessed on: 6-06-2021].

Shafiq, H. and Wani, Z.A., 2018, February. Assessment of Search Interface of Information Retrieval Systems A Case Study of Select Academic Databases. In *2018 5th International Symposium on Emerging Trends and Technologies in Libraries and Information Services (ETTLIS)* (pp. 45-53). IEEE.

Stansfield, C., O'Mara‐Eves, A. and Thomas, J., 2017. Text mining for search term development in systematic reviewing: A discussion of some methods and challenges. Research synthesis methods, 8(3), pp.355-365.

Tóth, E. and Kovács, B.L., 2017. Technical relevance of keyword searches in full-text databases. Qualitative and Quantitative Methods in Libraries, 3(2), pp.477-484.

Trinh, Q.D., 2018, April. Understanding the impact and challenges of secondary data analysis. In *Urologic Oncology: Seminars and Original investigations* (Vol. 36, No. 4, pp. 163-164). Elsevier.  
Tripathi, M., Kumar, S., Sonker, S.K. and Babbar, P., 2018. Occurrence of author keywords and keywords plus in social sciences and humanities research: A preliminary study. COLLNET Journal of Scientometrics and Information Management, 12(2), pp.215-232.

Veretennikov, A.B., 2018. Using Additional Indexes for Fast Full-Text Search of Phrases That Contain Frequently Used Words. arXiv preprint arXiv:1801.09079.

Voit, A., Stankus, A., Magomedov, S. and Ivanova, I., 2017. Big data processing for full-text search and visualization with elastic search. *International journal of advanced computer science and applications*, *8*(12), p.18.

Walters, W.H., 2013. E-books in academic libraries: challenges for discovery and access. Serials Review, 39(2), pp.97-104.

Westergaard, D., Stærfeldt, H.H., Tønsberg, C., Jensen, L.J. and Brunak, S., 2018. A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. *PLoS computational biology*, *14*(2), p.e1005962.

Xie, J., Qian, C., Guo, D., Wang, M., Shi, S. and Chen, H., 2019, April. Efficient indexing mechanism for unstructured data sharing systems in edge computing. In IEEE INFOCOM 2019-IEEE Conference on Computer Communications (pp. 820-828). IEEE.

Yang, L., Li, K. and Huang, H., 2018. A new network model for extracting text keywords. *Scientometrics*, *116*(1), pp.339-361.

ZHANG, N.N., WANG, Y.S. and Kun, Z.H.U., 2017. Design and Implementation of Full-Text Search Engine Based on Lucene. *DEStech Transactions on Computer Science and Engineering*, (CST).

Zientek, L.R., Werner, J.M., Campuzano, M.V. and Nimon, K., 2018. The use of Google Scholar for research and research dissemination. *New Horizons in Adult Education and Human Resource Development*, *30*(1), pp.39-46.