Journal of Positive School Psychology http://journalppw.com
2022, Vol. 6, No. 5, 3360-3380

Survey On Electronic Medical Record Search Engine In Healthcare

Dr. M.Thangaraj[1] , M.Karthika Devi[2]

[1]M.Tech. Ph.D,Professor & Head, Department of Computer Science, School of Information Technology, Madurai Kamaraj University, Madurai-21.thangarajmku@yahoo.com
[2]M.C.A. M.Phil,Research Scholar, Department of Computer Science, School of Information Technology, Madurai Kamaraj University, Madurai-21. karthikamku2020@gmail.com

ABSTRACT

Data at a high rate are produced at a higher rate by healthcare organisations. This leads to many benefits, but at the same time, it has drawbacks too. There is a rapid growth of free-text clinical documents in health care sectors that are generating alarge amount of EHRs data. Every patient has his medical chart. A patient's basic clinical data and medical details, which includes crucial signs, medications, demographics, examinations, plans of treatment, improvement notes, threats, dates of immunisation, allergies, radiology images and laboratory and test results, are maintained as a complete record named as a medical chart. Data from various sources, including medical Information, patient surveys, and administrative databases used to pay bills or manage care, are collected. An accumulation of patients' records saved in a computer is a medical database.

Health care providers termed it a computer-stored patient record, an electronic patient record, an electronic health record, or an electronic medical record. Accessing Information from the medical and healthcare domain is a challenging task while considering the time is taken and the accuracy of the retrieved information. Health records are private, so we can only access some patients records if we are authorised. So we have collected plenty of journals related to retrieving Information from an electronic health record. So we have discussed different information retrieval techniques, methods, and search engines for accessing health data from the medical database. Several limitations and future issues are explained.

Keywords: Clinical Data, Search engine, Electronic Patient Record, Electronic Health Record, Electronic Medical Record.

A. INTRODUCTION

A search engine is a tool based on the web, and it assists the users in identifying information on the world wide web. Normally, the search engine is used by people for three purposes: shopping, research and entertainment. There are mainly three types of search engines, found out during multiple projects on research. They include navigational, informational, and transactional.All major search engines try to identify which one the user has chosen by analysing the query and discovering the intent of each client to get a better insight into people's search needs.

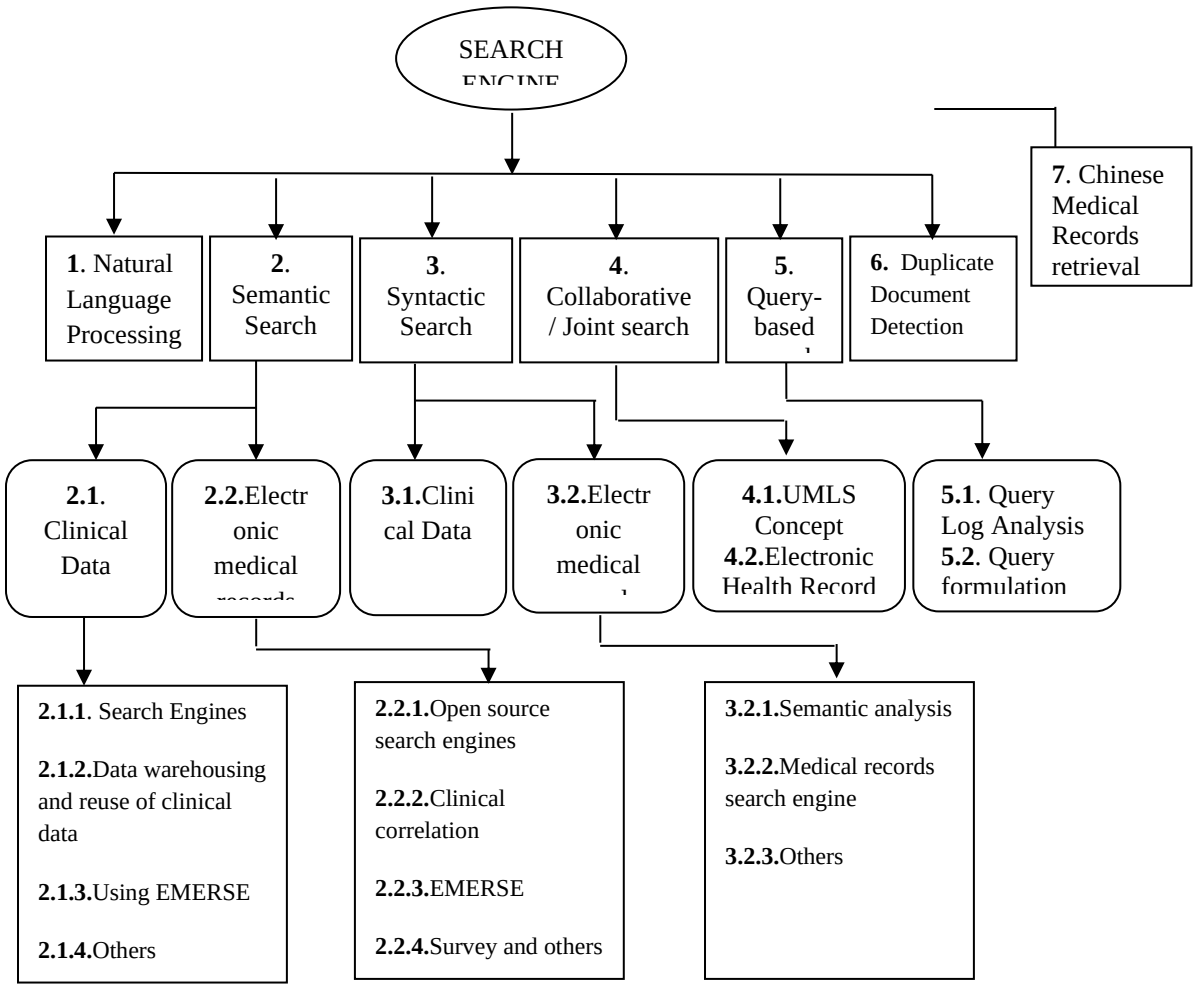
We classify the research papers into different categories as follows,

1. Natural Language Processing
2. Syntactic search
   - 2.1 Clinical Data
     - 2.1.1 Search Engines
     - 2.1.2 Data warehousing and reuse of clinical data
     - 2.1.3 Using EMERSE
     - 2.1.4 Others
   - 2.2 Electronic Health Record (EHR)
     - 2.2.1 Open source search engines

2.2.2 Clinical correlation
2.2.3 EMERSE
2.2.4 Survey and others
3. Semantic search
3.1 Clinical Data
3.2 Electronic Health Record (EHR)
3.2.1 Semantic analysis
3.2.2 Medical records search engine
3.2.3 Others
4. Collaborative search engines
4.1 UMLS Concept
4.2 Electronic Health Record
5. Query-based search
5.1 Query Log Analysis
5.2 Query formulation
6. Duplicate Document Detection
7. Chinese Medical Records retrieval

B SURVEY OF LITERATURES

Many journals related to searching for patient Information from an electronic health record database are reviewed and explain various information retrieval methods and search engines for accessing accurate health data from the medical database. Current practices for handling medication data in clinical trials have emerged from the requirements and limitations of paper-based data collection. Still, there are now many electronic tools to enable the collection and analysis of medication data. First, natural language processing (NLP) is a field of Artificial Intelligence that gives the ability of a computer program to understand human language as it is spoken. Second, Syntactic matching is matching search queries to keywords based on the searcher's actual words into the engine. This would be exact and phrase match. Third, Semantic matching is matching search queries to keywords based upon the intent of what the searcher typed into the engine. Then fourth, collaborative search engines (CSE) are Web search engines and enterprise searches within company intranets that let users combine their efforts in information retrieval (IR) activities, share information resources collaboratively using knowledge tags, and allow experts to guide less experienced.



Above figure 1 shows the hierarchy representation for different categorised research papers.

Figure.1 – Research Hierarchy

people through their searches. Next is a query-based search is a query based on a specific search term that a user enters into a web search engine to satisfy his or her information needs. Finally, Chinese Medical Records retrieval is to extract clinical information from free-text narratives in Chinese EMRs. Figure 1 shows the research hierarchy for search engine classification.

1. NATURAL LANGUAGE PROCESSING

This paper explains the retrieval of structured data related to glaucoma diagnosis and development from ophthalmologists' visit reports entered as "notational text" throughout patient interactions [1]. They compared different text transaction processing: GDP (Glaucoma Dedicated Parser), a local template matching system, and MedLEE, a substantiated natural language processing scheme routinely used at the NewYork-Presbyterian hospital's Columbia-Presbyterian Center to encode insights from mammogram reports and chest radiographs. Raw query data may be filtered and separated into usable analytical kinds using natural language processing technologies like this study.These data may be retrieved and analysed to learn more about the actual content that consumers seek and their degree of comprehension and complexity when they visit the website [2]. If natural language search engines were available on a health-information portal, for example, it might increase consumer access to needed information, especially for customers who are less familiar with the topic or language. Future queries will be subjected to further in-depth analysis. EHRs have been found in studies to assist doctors in reducing drug mistakes, improving patient outcomes, and increasing care efficiency. The specific benefits of the EHR are described by illustrating how natural language processing extended to electronic data might assist doctors in following complex occurrences after surgery, according to several articles evaluated [3]. Clinical notes must be manually examined to retrieve Information, limiting physicians' and researchers' capacity to study vast numbers of clinical encounters swiftly and efficiently. Natural language processing examines the context of phrases and words in medical records before allowing them access for high-performance computing, resulting in the capacity to understand EHRs manually.Natural language processing can be used in various ways, from automated quality evaluation to comparative effectiveness studies. They offer two NLP-assisted IR models, POS-BoW and POS-MRF, in which automated POS-based term weighting algorithms are included in bag-of-word (BoW) and Markov Random Field (MRF) IR models, respectively [4]. The findings support their prediction that NLP approaches, notably POS tagging, might increase the recognition rate in the biomedical sector by enhancing IR models.

Additional tests are carried out to confirm the efficiency of the proposed machine learning and the POS category selection. These considerable gains demonstrate the utility of using POS tagging for biological IR tasks. The suggested model's first restriction is that it is dependent on POS tagging performance; they would like to examine the impact of alternative POS taggers. The authors investigate the use of natural language processing (NLP) to select patients with non-alcoholic fatty liver disease, analyse disease progression trends, and detect gaps in care caused by a breakdown in usually expressed [5].They show that NLP-based techniques outperform ICD/text search-based approaches in diagnosing non-alcoholic fatty liver disease (NAFLD) in electronic health records. According to their explanation, NLP-based techniques can uncover lost insights in EHR data that require further investigation. For these investigations, they only employed one NLP tool. Future research should try a thorough review and comparison of various NLP tools for various use cases. Because they only examined data from one medical institution, the relevance of their findings to other situations is unknown.

2. SYNTACTIC SEARCH

Under syntactic search, we have classified journals into

2.1. Clinical Data

2.2. Electronic Health Record

The Information related to health correlated with routine patient care or part of a clinical trial

program is referred to as medical or clinical data. The clinical details of defined patient populations are collected and tracked with the help of patient/disease registries. The standardised accumulation of electronically-stored patient health details in a digital format is called an electronic health record.

2.1. CLINICAL DATA

Under clinical data, journals are further divided into Search engines, Data warehousing and reusing clinical data, Using EMERSE, and others. A search engine is a website through which users can search for internet content. The process of formulating and using a data repository is termed data warehousing. A data warehouse is constructed by coordinating data from various heterogeneous sources that assist in analytical reporting, structured and ad hoc queries, and decision making. The reuse of data is a technique that includes the use of research data for research activity or other than it was meant. Using EMERSE is a spontaneous, powerful search engine for free-text documents in EMRs. Other papers related to this topic were also discussed.

2.1.1. SEARCH ENGINES

The manual review identifies more than 90% of all registry patients at the University of Michigan Health System. This time-consuming and labour-intensive activity might be accurately performed by an automated computer system [6]. They developed a technology that can manually scan free-text medical records for cancer-related phrases. They created customized lists with over 2,500 words and phrases, as well as 800 SNOMED codes. The Case Finding Engine (CaFE) analyses text and highlights pertinent terms for inspection by a registrar before being utilised to create the registration database. The caFE is both accurate and efficient and has been well accepted by their registry team. Suggestions by the registrars have already resulted in substantial improvements. Additional work on specific areas could continue to improve its reliability and acceptability. The authors mention the method and the usage of the Star Tracker clinical search engine. The ultimate goal is an integrated searching capability that allows users to classify patient groups based upon demographic, clinical, and diagnostic data [7]. As possible, developing an enterprise-level data warehouse is the best solution but requires more resources and time. An alternative solution that works with multiple legacy systems without requiring prior integration of those databases may simply and inexpensively provide significant additional value. The Star Tracker search engine offers a complete population-level search capability using existing hardware and database systems.

They want to use technological advances to handle enormous volumes of unstructured textual data to swiftly identify and obtain information in this study. Fast availability of Information is especially vital in the medical field [8]. More effective access to vital Information concealed inpatient data indicated by millions of patient records is required for next-generation patient care. They showed the first biomedical hybrid NLP search/navigation engine process of forming the idea index for Electronic Health Records (EHR) utilising NLP for a large clinical text. Some enhancements to the way keywords are picked from the PHR might be made, such as employing IR methods to get relevant terms from explanations of every item on the record, rather than only utilising the title of every item.

2.1.2.DATA WAREHOUSING AND REUSE OF CLINICAL DATA

There has never been a comprehensive value analysis of PHRs, and one is necessary to assess which PHR functions provide the most value to PHR stakeholders. They propose a framework that may be used to determine the value of PHR functionalities and aid in creating PHRs. The paradigm is unique to personal health records, and the authors have subsequently adapted the associated assessment approach to other healthcare technology [9].The PHR framework presented and the associated approach should produce a complete assessment of PHR value. The author reviews the approach to coding medication data in multi-site research contexts. The author proposed a framework for classifying, reporting, and analysing medication data[10]. The framework can develop tools for classifying medications in coded data sets to support context-appropriate, explicit, and reproducible data analyses by researchers and secondary users in all clinical research domains. Future work is needed to better integrate medication classification into data management and analysis workflows and to assess the utility, validity, reproducibility, and scalability of different subsets of the

classification.A major impediment for the reuse is the presence of data in legacy patterns and the increased regularity when saved in an electronic medical record system. For this reason, techniques are necessary to normalise, accumulate and query data covered in the EHRs, to approve their reuse at any time needed[11].This work proposes a DW (DataWarehouse) environment based on EHR standards to allow the interoperability, agile aggregation of data sets, and data reuse in different scenarios, e.g., clinical research, CDS, surveillance, etc.

The tools and processes required for modelling, transformation, integration, standardisation, and aggregation of data flowing from the EHR to reuse it has been discussed in this paper. Recognising the value of clinical information modelling procedures, not just for ordinary health care delivery but also for clinical data reuse. The work of managing healthcare data is tough, and retrieving the patient's Information takes longer.Based on the new approaches, they offered a data warehouse architecture [12] for a medical information system. The suggested architecture maintains clinical data and assists data analysts and clinical managers in mining and analysing data stored in the warehouse. The data warehouse's major downside is the increased maintenance required due to the vast amount of data. As a result, if they use an intelligent system to save data in the future, human efforts should be decreased.

2.1.3. USING EMERSE

The main purpose of this work is to describe the occurrence and usual experiences of aneurysms among the recipients of abdominal transplants. It also focuses on finding patients with an arterial aneurysm in the medical record of those who have been experiencing liver or kidney transplants for over 11 years[13].The University of Michigan Institutional Review Board permitted to conduct this research. They could only check the electronic medical records of patients who used EMERSE at their institution. Some patients' Information may have been documented on non-electronic systems or at other hospitals throughout the state. The results of this study suggest that taking antacids regularly may have considerable therapeutic advantages in HNSCC patients.The causes for this link are still being researched, and they might lead to the development of novel treatment and preventative strategies using medicines with low toxicity [14]. Chart review and data abstraction from CareWeb using the EMERSE programme were used to identify patient medication usage.

They developed complicated yet accurate search queries using this custom-designed system to detect medicines taken and when (pre-or post-treatment), clinical, baseline demographics, and histopathological data in this cohort. A sequence of clinical studies will be required to further examine the anticancer potential of antacids in healthcare situations, with the ultimate objective of improving the prognosis of patients with squamous carcinoma of the head and neck (HNSCC). Reduction in HCV (Hepatitis C virus) disease burden will require the development of treatment regimens targeted towards patients in the current Era and improvement in early diagnosis and referral of infected patients to appropriate centres for treatment [15]. And also a reduction in costs of newly approved DAAs (Direct-acting antiviral agents), otherwise implementation of screening programs and availability of highly effective treatment regimens will have little impact on disease burden.

Medical records were reviewed using the electronic medical record search engine (EMERSE) developed at Michigan. Their institutional review board approved the study protocol. The authors explain the University of Michigan's nine-year involvement in supporting and using a search engine based on full text planned to give support to information retrieval from documents saved in electronic health records called Electronic Medical Record Search Engine (EMERSE). The performance of medical IR is increased by this method. It also affirms quality in search and results from flexibility despite the user’s training stage, medical history, or degree of technical proficiency [16]. In many studies of assessment, increased levels of sensitivity and specificity are established. Also, the efficiency of chart review is increased to a great extent. But, it is difficult to build IR functionalities into electronic health records. It also enables a better understanding of the various needs of the individuals and groups who want to mine Information from historical clinical documents.

2.1.4. OTHERS

This work describes the CER Hub, a web-based informatics platform for developing and

conducting research studies that combine complete electronic clinical data from multiple healthcare organisations [17].CER Hub involves systematic and scalable techniques to extract, consolidate, and analyse large, multi-institutional clinical data. CER Hub provides a solution to the issues of conducting comparative effectiveness research across many institutions. While HL7 standards and methodology guided their methods in constructing the CER Hub infrastructure, some reviewers noted they were not doing enough to define the "one global standard" that is unquestionably required for the local and global exchange of clinical data for research purposes. They just touched on a few of the numerous concerns in governance, organisational policy, data security, and other critical areas to operations in this study.The authors developed the Melanoma Rapid Learning Utility (MRLU), a component of the RLS, providing an analytical engine and user interface that enables physicians to gain clinical insights by rapidly identifying and analysing cohorts of patients similar to their own [18]. The MRLU enables physician-driven cohort selection and stratified survival analysis.

The system successfully identified several known clinical trends in melanoma, including frequency of BRAF mutations, the survival rate of patients with BRAF mutant tumours in response to BRAF inhibitor therapy, and sex-based trends in prevalence and survival. Further research is necessary to evaluate when and how to best use this functionality within the EMR clinical workflow to guide clinical decision-making.The authors compare MedWISE - a novel EHR that supports user-composable displays with a conventional EHR regarding the number of repeat views of data elements for patient case review [19, 20].

2.2. ELECTRONIC MEDICAL RECORDS (EMR)

Under EMR, journals are further classified into the following categories. First, Open source search engines simply mean that the source code (programming) is available to anyone to use and modify as they desire. The second one is a clinical correlation; it is a medical process physicians use to help them make a diagnosis on a patient to treat his or her condition. Next, EMERSE (The Electronic Medical Record Search Engine) is a powerful search engine for free-text documents in the electronic medical record. Finally, survey papers related to EMRs are discussed.

2.2.1. OPEN SOURCE SEARCH ENGINES

Web search engines provide access to all kinds of Information; additional tools are required to automatically link Information retrieved from these engines to specific biomedical applications. CDAPubMed is a platform-independent tool that facilitates literature searching usingkeywords contained in specific EHRs [21]. CDAPubMed is visually integrated, as an extension of a webbrowser, within the standard CDAPubMed interface. CDAPubMed is an open-source tool with a modular architecture; future integrations, code reuse, and contributions from the biomedical informatics community will be feasible. It can be freely used for non-profit purposes and integrated with other existing systems.A full-text search tool was introduced into the daily practice of Leon Berard Center (France), a health care facility devoted to the treatment of cancer. To describe the development and various uses of a tool for full-text search of computerised patient records, using an open-source search engine Solr[22]which indexes content sources, processes query requests, and returns search results in electronic health records, this project demonstrates that the evolution of computer technology in the health sector has the potential to significantly change daily medical practice inpatient care.

A new application based on open-source (GastrOS) has been refined to identify whether it is easier to manage a clinical data system produced with the help of open HER model-driven development versus the methods based on the mainstream. This application follows open EHR's multi-level modelling approach [23]. Normally, it took half the time to carry out fluctuations in GastrOS.The use of openEHR model-driven development can result in better software maintainability. One limitation of the study design is that only one developer implemented changes on each application, and neither was unknown to the nature of the experiment. However, the second author, who developed GastrOS and implemented the CR (Change Request), was neither familiar with the application domain nor had any open EHR implementation experience before the study.This work introduced the new open-source BIO Medical Search Engine Framework

for the fast and lightweight development of domain-specific search engines [24]. The BIO Medical Search Engine Framework supports the development of domain-specific search engines. The key strengths of the framework are modularity and extensibility in terms of software design, the use of open-source consolidated Web technologies, and the ability to integrate any number of biomedical text mining tools and information resources.

2.2.2. CLINICAL CORRELATION

The authors concentrated on integrating various comprehensive data sets (e.g. laboratory results, vital statistics, drugs) with partly unstructured medical data such as diagnostic reports, discharge letters, clinical notes, and a research database [25]. This sort of knowledge-based system gives doctors a practical tool for analysing medical data and making decisions. They created a user interface for faceted search that the Solr Engine powers. In the future, we will examine data in a temporal context by connecting textual patient information with elements such as diagnoses, prescriptions, and test values.Exfoliation syndrome (XFS) was detected using an algorithm that searched both structured and unstructured (free-text) EHR data [26]. The technique was used to calculate an XFS likelihood score for every patient based on their EHR data. The method created, tested, and validated in this work appears to be more effective than the traditional technique of using EHR data to investigate individuals with ocular illnesses in detecting the presence or absence of XFS. The first limitation is,additional validation is required to determine how well the algorithmidentifies XFS among patients receiving care in other settings and with other EHR systems. Second, it would not detect the condition if a patient received a misdiagnosisor the clinician did not document the characteristicfindings of XFS. Third, with more EHRs to train the algorithm, it may be possible to achieve even greater sensitivity and specificity for detecting this condition.

2.2.3. EMERSE: THE ELECTRONIC MEDICAL RECORD SEARCH ENGINE

EMERSE is a powerful search engine for free-text documents in the electronic medical record.The Electronic Medical Record Search Engine (EMERSE) was developed to address the need for searching the medical record for research and data abstraction. It offers various choices for creating complex search queries however has an interface that is easy to be used by those with minimal computer experience. EMERSE [27] is perfect for traditional chart reviews and data abstraction and may be possible for clinical care.EMERSE presently searches only those patients specified by a user, even though the program could be modified to search all patients in the health system. This could be useful for finding patients who are affected by a drug recall. Additional privacy issues would need to be addressed. In addition, EMERSE has applicability in the direct patient care situation where clinicians are increasingly pressed for time. A quick method for reviewing a patient's history for mainactions of interest would be welcome.

2.2.3.1. USING EMERSE

The following four papers use EMERSE for their research. So they are listed under using EMERSE.The authors compare the clinical accuracy and speed of eligibility chart reviews performed using EMERSE with those done manually through the EMR. Based on their experience with this tool in several studies, they hypothesised that using EMERSE would be faster than manual chart reviews while maintaining clinical accuracy. EMERSE is not from a more superior and complex search algorithm but a user interface designed clearly to assist users to search the entire medical record and protect health information [28]. The majority of cases where raters showed disagreement with the gold standard resulted from falsely classifying a patient as eligible. When screening for study inclusion, these false positives will have less impact on targeted enrolment as there are typically several occasions in the process to reconfirm true eligibility. This work evaluated the occurrence, clinical features, risk factors, and outcomes of Engraftment syndrome (ES) in children and adults undergoing first-time allogeneic hematopoietic cell transplantation. A patient with engraftment syndrome is found out with the help of 82 search terms [29].EMERSE, which delivers software tools to thoroughly scan all clinical papers from their electronic health record (EHR) system for keywords and phrases to guarantee that even seldom mentioned occurrences are discovered, was used to enable data abstraction. Every patient whose chart has been highlighted by EMERSE

is subjected to a more thorough evaluation. To improve future investigations, prospective collections of clinical characteristics should be integrated with correlative laboratory analysis.

2.2.4. SURVEY AND OTHERS

The centralised system design philosophy is hampering their capacity to be flexible and adaptive to shifting demands. These demands result from the context in which the techniques are employed and changing user needs [30]. Their customisable information retrieval and access service architecture may be utilised to allow high-level personalised requests while also adapting and utilising available technology in the medical and healthcare sectors.Create a prototype and test the idea utilising real-life scenarios in the future. The insignificant utilisation of electronic health record data is contingent on getting precise and complete information about specific patient populations [31]. The medical records track at the Text Retrieval Conference 2011 was used to retrieve patients useful for clinical investigations from a database of de-identified medical records grouped by patient visit. Their study highlights natural areas that will increase the capacity to acquire and use healthcare information more efficiently for secondary purposes like research and surveillance.

Noteworthy chart references to a term, variation in terminology and spelling, lack of distinction among various situations with high similarity, lack of difference among past, present, and specific pattern or procedures, and failure on IR systems to differentiate in between existence or denial of a symptom or condition were among the most prevalent causes of retrieval error, according to their research. Future advances in EHR systems and retrieval abilities will have to address these potential causes of mistakes.To fully understand the value of the important Information recorded in the EHR, more effort is required. Clinical studies take time and necessitate a strong focus on data quality. The expanding volume of electronically available patient data allows medical informatics to help two crucial activities: recruiting patients into studies and documenting trial data [32].

They proved that automatic data reuse from the ICU's electronic medical record could save time while collecting research data. Data quality might be harmed as a result of the higher risk of typing mistakes. They want to urge academics who create multi-centre data gathering systems for future research projects to always allow batch data input from current electronic data sources and web-based data entry interfaces. The clinical narrative text is treated as a document series using a framework that models both the textual and temporal aspects simultaneously.In terms of structure and generalisation, the [33] framework demonstrates to be adaptable. They want to dig further into the temporal structure of EHRs in the future and discover a more expressive and efficient representation to capture the temporal aspect of EHRs. Another component of their future work will be developing a more complex way to merge temporal and textual similarities. Medical record and patient outcomes have improved as a result of EHR implementation. Depending on the form of the study topic, the EHR can supply the requisite current data for beneficial results and the clinical effect of the change [34].All present techniques for leveraging EHR data to help research have advantages and disadvantages. Collaboration is required to determine the optimum technique for incorporating research-oriented data collecting into routine paediatric urologic clinical practice that best matches the institution. Due to privacy concerns, the small size of data sets is frequently cited as a cause for visual retrieval in the medical arena. However, there has been a lot of progress in medical visual information retrieval because of the rising availability of big data sets and scientific difficulties.There are various limitations to this research [35], including the fact that the amount of evaluated texts from 2011 to 2017 is restricted to only six years. The problem with the domain is that the particles are dispersed throughout several study topics, making it difficult to be systematic and comprehensive. It is critical to incorporate multimodal data wherever feasible since it appears hard to evaluate visual data without all of the data that impacts the visuals. As a result, fusion techniques and methods for combining data from disparate sources will be necessary for the future. Another issue to address is the limited clinical utilisation.

3. SEMANTIC SEARCH

As same as syntactic search, the semantic search also classify journals into

3.1 Clinical Data

3.2 Electronic Health Record (EHR)

3.1. CLINICAL DATA

BioPatentMiner is a system that uses a combination of algorithms to find Information about biomedical patents. The algorithm locates and analyses physiologically relevant phrases in patents and establishes relationships amongst them [36]. They want to perform user tests with domain experts to verify the efficacy of their methodologies and increase the recall of their Connection Annotator by introducing new templates for detecting other relation patterns in phrases. Sustainability is one of their main issues. One alternative is to create a distributed Web and change its algorithms to retrieve data from different sources. To allow automatic reasoning on biological ideas, distributed Web servers would record the "meaning" of ideas and sets of predefined rules in biomedical ontologies. Researchers will be able to get all relevant information on a biological topic using a single semantic search. The requirement for prompt care of patients in life-threatening situations.They offer a framework [37] that encourages people to use tailored mobile services and integrated medical systems to get access to medical systems. The proposed mobile medical approach incorporates current medical systems and enables high connectivity via service composition and QoS adaption, using suitable standards and structure on available services. Mobile on-demand home healthcare services would be provided via safe and reliable wireless communication between medical specialists, patients, and hospital medical data. Various web-related semantic similarity metrics have been used in the study. However, determining semantic similarity between the two phrases is a difficult issue. Both ideas must have dwelt in the same ontology tree, according to traditional ontology-based techniques (s).The internet is a never-ending and massive growth machine. As a result, using the page counts of two biological ideas provided by the Google AJAX internet search engine, a method of evaluating semantic relatedness is suggested. The study [38] lays the groundwork for investigating and implementing a real-time application for extracting similarity measure keywords from completely relevant data in electronic, free-text medical records, such as radiology reports and departure notes.

3.2. ELECTRONIC HEALTH RECORDS

Journals are further divided into the following categories under EHR. First, by combining semantic analysis with text mining techniques, it is feasible to get better outcomes in terms of information retrieval by utilising as much database information as possible. The second is a medical records search engine created exclusively for looking up medical information on the internet. Finally, several studies about electronic health records are examined.

3.2.1. SEMANTIC ANALYSIS

The volume of digital information has increased dramatically in the last two decades, and the health industry is no exception. Medical data may be retrieved to acquire useful Information regarding treatments and the progression of clinical problems, which can help diagnosis new patients faster. The primary goal of this study was to contribute to information extraction using semantic analysis, which yielded positive findings [39]. Furthermore, this paradigm may be used in any language as long as the database dictionaries are given relevant information. They want to work on complex words in the future, including spelling correction, detecting negative phrases and analysing probabilistic and hypothetical statements.Integrating medical knowledge sources can help with information retrieval, but it is not an easy job. They suggested a unique medical information retrieval system with a two-stage query expansion technique that can effectively model and utilise latent semantic correlations to increase performance in this study [40]. In the future, they plan to (a) compare the proposed system's performance to that of existing medical IR datasets. (b) Using a Bayesian approach, investigate the possibility of combining the proposed tensor-based latent semantic relevance model with the probabilistic tensor decomposition framework to improve the performance of the medical IR system. (c) collaborate with clinicians and decision-makers at local hospitals to apply the system in real-world medical decision-support applications.

3.2.2. MEDICAL RECORDS SEARCH ENGINE

Traditional search engines are unaware of the context in which the user is searching. When consumers are looking for health-related Information, this becomes a major concern. They describe a [41]search tool that assists

users in finding relevant health information on the internet by leveraging data from personal health records as context to give results relevant to the user's current conditions. They tested this instrument on 18 people, and the findings of that test are discussed. The way the context keywords are picked from the PHR may be improved by employing IR methods to get relevant terms from descriptions of each item on the record, rather than only using each item's title. Information retrieval operations in various fields rely on prominent web search engines like Google, Yahoo!, and Live Search, among others, to gather preliminary data. In the eyes of the search engine user, these retrieved information items may not be related to the search objective. This work aims to research and create a dynamic, question-and-answer kind of search engine [42] that allows for more precise and relevant information in the Electronic Medical Record (EMR) area by searching by characteristics. They provide the results of one case from 13 physicians who used MedWISE [43], a widget-based electronic health record (EHR) interface, to familiarise themselves with actual patient situations and express their opinion and strategy. Multiple methodologies were utilised in MedWISE to investigate use patterns, time spent, new feature employed, user-created interfaces, diagnosis, and human-computer interaction processes.

All physicians rapidly mastered MedWISE, used more than half of the new features, and deemed MedWISE to be simple to use and valuable. The solitary patient instance, the limited number of users, and the fact that the study was a laboratory study that may not adequately mirror field circumstances are all study limitations. Although this may limit its applicability, utilising real patient cases in a simple system for a realistic goal is a strength. The purpose of this study was to explore if employing a semantic search will help a physician perform better on an information-gathering job. Semantic search, as used in this study, is a cloud-based tool [44] that allows doctors to conduct structured searches of unstructured text anywhere in a patient's file using clinical vocabularies.The semantic search challenge required participants to use the semantic search interface to locate similar Information from a separate patient record. It removes the obstacles to physician performance related to information overload, allowing physicians to make more correct decisions. Because the study was done in only one EHR setting, the outcomes may alter if participants used a different EHR. The search tasks were completed in the same sequence by all participants rather than in random order. A brief video example was used to teach users the semantic search capabilities. Because the movie defined the purpose of the semantic search tool and its intended usage, this introduction probably impacted the time on task or number of clicks measures. They advise utilising a bigger sample size for measuring the efficacy of the semantic search tool.

3.2.3. OTHERS

They use semantic web-based technologies to draft attached data from medical information systems and build a heterogeneous data integration model [45] that integrates the Mediator/Wrapper architecture with ontology and uses OGSA-DAI as data accessing middleware. This model aims to resolve semantic heterogeneity induced by a heterogeneous medical information system, support integration and sharing of heterogeneous EPR (Electronic Patient Record) and HIS (Hospital Information System) data, provide data model process schema to support the exchange of Information among heterogeneous systems in a grid environment, and demonstrate uniform views to users. The actual construction and verification of the model to integrate disparate health information systems will be part of their future effort.

Participants (N=10) were given two real-life scenarios in which they had to look up patient information. In the first instance, participants answered properly to a patient-specific information request. Participants were provided with a semantic search tool that identified phrases inside a patient EHR in the second scenario. Following that, participants were asked questions on their current use of the EHR in a semi-structured interview.

According to this study, semantic search capabilities might be a useful strategy to lessen the cognitive burden in clinical settings for comparable patient-specific information demands [46]. It's feasible that giving a description or list of the EHR's searchable items would have increased trust and consequently accuracy perception. Determining how perceived and real accuracy increases over time would be fascinating research.

They outline their attempts to develop a context-based EHR that uses biological ontologies and (graphical) illness models as domain knowledge sources to find relevant record elements to show in this paper. They provide a system [47] that collects and mines results and characteristics from free-text clinical reports, maps findings to concepts in existing knowledge sources, and creates a personalised record presentation depending on the user's information needs.

Current search engines have two major drawbacks: user interaction with the list of returned resources is limited, and there is no explanation for their relevance to the query. Users have no notion how to fine-tune their searches such that the answers are as expected. This study [48] proposes an information retrieval system that employs a graphical explanation of query results to favour user interactions and relies on domain ontology to find a collection of relevant resources. Reformulation poses various optimisation and mathematical concerns, but it also raises critical difficulties with user input to continue comprehending and successfully interacting with the IRS (Information Retrieval System).

4. COLLABORATIVE SEARCH

UMLS and electronic health data were used to categorise the collaborative search. The UMLS, or Unified Medical Language System, is a collection of files and software that combines a variety of health and biomedical vocabularies and standards to allow computer systems and EHRs to communicate with one another.

4.1. UMLSCONCEPT

Using a human-computer collaborative method, find Common Data Elements (CDEs) in the eligibility criteria of several clinical trials researching the same illness. To find disease-specific eligibility criteria CDEs, a set of free-text eligibility criteria from clinical trials on two representative illnesses, breast cancer and cardiovascular disorders, was collected. A semantic annotator [49] is utilised in this study to recognise phrases from the Unified Medical Language Systems (UMLS) inside the eligibility criterion text. Using a human-computer collaboration strategy to improve domain experts' ability to identify disease-specific CDEs from free-text clinical trial eligibility criteria is viable and time-saving. This method has two significant drawbacks. Some false positives were caused by verbs (for example, "arrange" and "repair"). Second, several UMLS phrases may be found in multi-term CDEs such as "[patient's] age at diagnosis" and "[patient's] age at death." The authors describe a revolutionary electronic health record (EMR) retrieval system that aids in the reduction of manual effort in the healthcare industry.Their approach [50] makes good use of medical domain knowledge to improve retrieval speed. They propose a joint searching framework for developing an EMR search system to flexibly apply medical domain knowledge. They demonstrate this property by introducing an effective way to enrich domain knowledge using MetaMap and UMLS and a novel approach for external notion space expansion.

4.2. ELECTRONIC HEALTH RECORD

Users' lack of search knowledge and/or medical domain knowledge is a key obstacle to using such full-text search engine solutions efficiently. The authors evaluated a 'collaborative search' feature [51] using a home-grown EHR search engine that allows users to preserve their search expertise and share it with others to alleviate the problem. They developed and tested a 'collaborative search' tool for user involvement and collaboration to preserve, jointly refine, and disseminate EHR search expertise across individuals and domains. As a result, they recommend practitioners and academics to investigate using this and maybe other social information-foraging strategies extensively utilised on the internet to improve the quality and efficacy of healthcare information retrieval. Each record in clinical data has a timestamp, which is the date the record was created. Within a certain period, the timestamps are taken from a collection of records (a patient's profile).In this study [52], they offer a strategy to exploit the link between EHR temporal distributions and implement it into the IR system to improve EHR search performance. One component of their future effort will be to use more advanced combination approaches. The other part will take temporal information analysis to a higher level, such as incorporating temporal Information into medical ontology analysis to improve EHR search performance even further. One of the most important tasks in ensuring proper use and comprehension of accessible medical data is information retrieval. This study

[53] established a ranking algorithm for determining the ordering of results based on multiple dimensions and forms of intent from users' searches. Their idea included annotation of meanings with intention aspects in EHR repository textual materials.

5. QUERY-BASED SEARCH

Journals are further categorised into the following categories in query-based search. The analysis of user searches to investigate information-seeking behaviour, system functionality, and search subject trends is known as first query log analysis. Query formulation (QF) is the interactive information access process in which a user converts an information requirement into a query and sends it to an information access system such as a search engine (e.g., Google).

5.1. QUERY LOG ANALYSIS

They assess the impact of UMLS knowledge exploration in medical domain information retrieval by mapping domain. Expanding searches and documents automatically based on semantic relations in the UMLS hierarchy and large text of collection ImageCLEFMcd to UMLS concepts [54]. They investigated the use of the UMLS Metathesaurus, an external knowledge base, in medical information retrieval. With brief and topic-oriented texts, document expansion proved to be the most efficient strategy for retrieval on the domain-specific corpus. To approach a more semantically rich IR system, more effective investigation of semantic relations and other semantic Information in this knowledge source. In this effort, they needed to understand the user's wants as revealed by their search queries. Their research found that a range of user types used an EHR-based search engine and that physician queries in the EHR are mostly informative and focused on laboratory results[55].While log analysis is a quick approach to learning about a user's behaviours, it does not provide information about the user's background, which is necessary for searching. Log analysis must be augmented by observational and survey research to identify what people are looking for and the usability of a search utility. The semantic classification of questions is another study's drawback. The only way to solve this problem is to use a semi-manual technique. A trained classifier algorithm categorises a random sample of queries, which are then manually assessed. This type of evaluation would only happen once in a while.The authors provide a detailed understanding of the inquiries, an organisation of the information needs revealed by the queries, and physical patterns of the users' information-seeking behaviour. The findings [56] suggest that the information needs in the medical sector are far more sophisticated than those that general-purpose online search engines must serve. As a result, there is a significant challenge and area to give intelligent query recommendations to aid the retrieval of information from electronic health data. This study sheds light on how to create an effective information retrieval system for electronic medical records.

5.2. QUERY FORMULATION

The author reviews state-of-the-art techniques as well as major methodological issues in [57]. They modified a previously published conceptual framework for interactive information retrieval, which identifies three entities: user, channel, and source, by elaborating on query formulation channels in the context of allowing end-users to query EHR data. Their findings demonstrate that allowing biomedical researchers to conduct reference interviews for EHR data is a potential route for increasing user autonomy during EHR data interrogation. They created search parameters that were based on pre-determined subjects. This technique may favour self-selected subjects and exclude topics related to this review but not searchable using the query obtained from the pre-selected topics. They may have missed significant works in the area from the past due to their concentration on the most recent four years of literature. They feel, however, that their comprehensive citation search should have essentially resolved the issue. They offer a comprehensive description of the issues experienced in the communication area in this paper. Second, they find synergies across domains researching human-to-human and human-machine communication that might help researchers better understand biomedical data query mediation. They devise a mixed-initiative system [58] that may operate as an intelligent intermediary between clinical data and clinical researchers, guiding them through an effective and well-organised query formulation process step by step. In studies of communication in clinical research query

mediation, there is a significant research gap. To give relevant characterisations of dialogue behaviour in human-to-human and human-computer conversation, rigorous and systematic investigations are required.

6. DUPLICATE DOCUMENT DETECTION

Document Cleaning is the pre-processing step used to clean the noise and irrelevant Information in the huge volume of text documents. In the Document Cleaning phase, the Duplicate Document Detection framework is developed, which helps identify duplicate documents based on the content similarity of the documents. Document Transformation is the process of transforming the documents into a specific format suitable for further analysis. This research work has developed the Automatic and Appropriate File Name Generation framework for Document Transformation. The primary goal of this framework is to generate the appropriate file names automatically based on their contents.

From time to time, due to human errors and various reasons, the creation of duplicate documents may happen in the personal computer or servers. Documents may have the same name, same contents or same size. Documents are considered duplicates if they contain identical content. The inclusion of such duplicate documents in the search results degrades the search quality. Duplicate document detection is the process in which multiple documents with identical data can be identified and prevented during the generation of search results. The content of every document is scanned for detection of duplicates using the indexing process during global analysis. When Information is gathered from multiple sources, algorithms play a major role in the detection of duplicates. On deletion of duplicate documents, the search precision is improved, and runtime is reduced. Searching and Indexing are accelerated by duplicate document detection.

In [59] proposed the significance of identifying duplicate documents and preventing multiple records of the identity document in the database. The character Shape Coding (CSC) technique was used for mapping the character images based on their location and shape corresponding to the text lines. This method was cost-efficient and robust. Based on the size of the constituents of the page, the page image was transformed into an array of connected and bound within boxes. Large components are assumed to be non-textual Information, and those were removed. Only the textual Information was taken into consideration for comparison. This method was efficient in databases with up to 100,000 documents.

In [60] recommended a technique for identification of near-duplicate copies with sentence-level detection algorithm. This method offered a higher recall rate, precision and efficiency. It consumed lesser time for Indexing and storage. The degree of duplication was measured at the sentence level only, ignoring the non-informative details and capturing only the required Information. This method was greatly used in the detection of plagiarism and the identification of possible citations. A sliding window was used initially for analysing the word count in sentences. Further, document fingerprinting and shingling methods were used to compare their efficacy and proficiency with word-level features.

This paper [61] introduced a solution for the duplicate document detection issue through a framework for clarification and formalisation. The author has presented four discrete models with a complementary algorithm for string matching. The models were full-layout duplicates, full-content duplicates, partial-layout duplicates and partial-content duplicates. A set of sample data was used, and the system was tested, and its robustness was authenticated. This research work overcame the drawbacks of the traditional Optical Character Recognition (OCR) Model.

In [62]**,** used fingerprint technique to identify similar and nearly similar documents by signature stability evaluation. This work focused on determining the degree and duplication varieties. They had experimented with 50 million documents to identify duplicate documents. The authors had examined the utility of document signatures for addressing duplicate documents.

[63] used I-Match collection statistics to identify duplicate documents with sample datasets ranging from 30 MB to 2 GB. The solutions were provided accurate results and reduced execution time for up to $1/5^{th}$ of the

existing state of the art methods like syntactic filtration, shingles, digital syntactic clustering and its super shingle (DSC-SS). Uncorrelated OCR outputs were used for duplicate detection in databases of document images.

In [64] used five different types of documents, which includes printed, faxed, third generation, light, dark and annotated, are analysed for OCR accuracy**.** The noise sources and their effects in the real-world environment are studied. Five experiments were conducted, and each focuses on specific reasons like detecting duplicates under realistic, moderate noise conditions, analysing the influence of using a degraded input and examining duplicate detection when there are difficulties in reading order. It also determined the empirical relationship between duplicate models and comparison measures and the detection of duplicates in databases containing several false duplicates.

In [65] introduced a technique for evaluation of information retrieval and fingerprinting methods. Authors have analysed the effectiveness for accurate identification of documents that are co-derived. In [66] developed a novel method using the trie-tree structure for duplicate document detection. This structure was used to store 64-bit fingerprints of documents that were collected from the website. The proposed method was applied to detect spam mails with higher accuracy.

Paper [67] has proposed an efficient model for automatically renaming the files based on the content analysis. This model consists of three important stages: file name classification, semantic proof of file names, and content analysis. For each stage, they have proposed new algorithms. Overall, the proposed model has been used to rename the files based on their content efficiently.

7. CHINESE ELECTRONIC MEDICAL RECORDS

The study of symptom name recognition (SNR) in the principal complaints, which is one of the major jobs in structuring TCM (Traditional Chinese Medicine) FCRs (Free-text Clinical Records), is detailed in this paper. SNR in principal objections is tackled as a sequence labelling problem in this study [68]. For various practical reasons, the basic sequence labelling technique is properly tailored for the SNR job based on the domain-specific properties of TCM FCRs. Furthermore, three commonly supervised classifiers are examined, and their specialisations for the SNR problem are carefully analysed. Treatments, time information, and medications are among the different entities that must be examined. The bigger the number of entities considered, the greater the obstacles. To fully use TCM's FCRs, many sorts of named entities will need to be recognised.

In electronic medical records (EMRs), various rule-based and machine-learning approaches are accessible; however, only a few studies have looked into a hybrid technique for retrieving Information from Chinese EMRs. Automatically extracting information [69] from unstructured Chinese clinical writings is a difficult challenge that differs significantly from clinical literature in English. They suggested a new hybrid technique for extracting clinical Information from free-text narratives in Chinese EMRs that combines the BiLSTM-CRF Model with heuristic criteria.Their approach can help physicians obtain clinical data that could be useful in clinical investigations more efficiently. They offer the Cloud Medical Record Retrieval System, an efficient and robust cloud-based system for large-scale TCMR retrieval (C-MRRS).

They want to employ a distributed TCMR retrieval system [70] in this study to improve the secondary usage of large-scale TCMRs in the cloud. They propose a multi-indexing model to ensure that the appropriate TCMRs are organised and managed to recover in real-time, as well as semantics development and the multi-factor ranking model, which are used to improve retrieval quality. They also present a template-based visualisation method to improve generality and usability, where the medical notes are showcased via a friendly web interface. The proposed system can handle only TCMRs. Working in non-TCM situations is tough since they must develop and maintain many CDA document templates. This is a challenging and demanding job. For performance evaluation, they only give 2 and 4 keyword queries. The study on electronic medical records has certain hurdles due to pre-processing data issues, time-consuming and arduous data labelling in Chinese electronic medical records, and various electronic medical record data storage. The semi-supervised

learning approach in machine learning is applied to electronic data in this article [71].The approach may also be used to analyse data from other electronic medical records. In diagnosis and treatment, the study presented in this paper aids decision-making for follow-up doctors and other medical personnel. For the selection of limitations, there is still some research room. Furthermore, the data from electronic medical records that are now being studied is incomplete in terms of time and place. The effect and reliability of auxiliary diagnosis will improve with the sharing and fusion of relevant data.

Table 1.The key contributions and drawbacks of secure data storage blockchain-based EHR systems.

Reference	Technologies	Contributions	Limitations
[72]	PKE	❖ Maintain responsibility for confidential healthcare information and Integrity ❖ Cryptographic functions can secure patient data. ❖ Return the right to access private data back to From patients ❖ The real identity can be covered using pseudonymity from the patient	❖ Cost-effective PKE computing ❖ Key Management is difficult ❖ Chance of (ID/password PWD user's) and leakage of data
[73]	PKE	❖ It is possible to use the specifics of nail photos for Control of identity and assistance in further study of Disease and wellness. ❖ To easily use SVM and the random forest tree algorithm, Biometric verification, and precise. ❖ Protecting the confidentiality and privacy of confidential data Blockchain Use	❖ There could be bottlenecks in the IoT devices with resource constraints ❖ The possibility of the leakage of nail image data in the Public blockchain ledger for public usage
[74]	AES (SKE)	❖ The standard of data can be accessed from wearable devices enhanced using methods of machine learning. ❖ For large volumes, an off-chain computing database is used for Datasets ❖ The Shamir secret strategy of sharing is used to Enhancing data protection and privacy ❖ Users hold the freedom to regulate their rights. Health records and can safely exchange them.	❖ On intent or unintentionally, data leakage customers who decrypted the demanded details
[75]	MPC	❖ In the private blockchain, healthcare data has kept the cloud against threats to secrecy and honesty. ❖ Integration of healthcare data is scalable and simple. Easy indicator-centred schema to be used as storage the model ❖ The MPC can be used to do calculations without encrypted data among untrusted entities leakage of data.	❖ High-cost computing with MPCs ❖ Data replicas to requestors can cause data replications without the exploitation or leakage of records, Permission from the owner.

		❖ This encourages patients to handle their info via their data gateways safely.	
[76]	MA-ABS	❖ No patient identity or characteristics for clear speech signature argument for privacy preservation ❖ Unforgeability of the verifier ❖ Resist attack of collusion	❖ Computing for high-cost ❖ The general nonmonotone does not helpPredicates
[77]	SKE & CES	❖ Enable patients to exchange signed ones selectively by their dreams, medical details ❖ Using various public keys for various transactions protecting the true identities of consumers ❖ Transaction for anonymous and charitable patients ❖ Malicious applicants may be monitored.	❖ Have a clear transaction impact processing because it takes a long time to deliver a fresh Block
[78]	SKE (AES/3DES)	❖ The cost of preservation for encryption keys significantly reduces blockchain. ❖ Improve the protection of physiological data substantially in using distinct buttons, the block ❖ The rival, after decrypting ciphertexts, has no ability to relevant Symmetric Keys.	❖ The possibility of public ledger data leakage ❖ If the information is revealed, all the data will be exposed. ❖ Lost the related symmetric key

C. DISCUSSION

Electronic health records, or EHRs, include a lot of information about a patient's health, such as test results, diagnoses, and treatments. Searching for EHR data can be hard for researchers to use because there are many ways to describe the same thing. For example, cancer may be called Carcinoma in the data. The EHR may also include abbreviations or spelling mistakes. Search engines offer users large amounts of relevant information. So researchers want to build an efficient search engine for EHR data for search, and it is also used for research and health organisations.

Factors leading to this intricacy include the clinician's constant use of compatible terms, acronyms and abbreviations, and contrary and avoided phrases. Further, a shortfall of typical grammar and punctuation usage results in uncertainties. It also results in fundamental complexities for the computer systems to deal with context-sensitive meanings. EMERSE has a limitation as additional privacy problems need to be concentrated. A speedy technique for checking the history of the patient for important actions of interest would be expected. EMERSE is a free-text EHR search engine. It acts as a relevant tool to aid the researchers and practitioners recover the details from electronic health records more efficiently and facilitating complex tasks like a synthesis of patient cases and abstract research data. In the future, multimodal data used for retrieval because it is unattainable to depict the visual data.

Various techniques and methods to combine information from very different sources will be required in the future. The limited use of clinical data is another problem to be tackled. After retrieving the Information from EHR, the search results did not show the original values of sensitive attributes in EHR to the third parties. So we need a medical search engine with data privacy to avoid third parties stealing the sensitive private data of patients. Automated data extraction methods, including natural language processing, have to transform disorganised clinical notes into an organised, codified and hence determinable format. An efficient, flexible and scalable solution can be offered by the search engines or IR systems, which can further improve the value of disorganised clinical data. The challenges to

extracting Information locked in the medical text should not be very easy. The insignificant use of electronic health record data depends on recovering exact and full details about the specific patient populations.

D.CONCLUSION

Based on this literature review, various searching techniques have been recognised with their pros and cons in extracting Information from EHR. This paper provides an abrupt analysis of the syntactic and semantic search engines that uses discrete advances in multiple ways to fetch the relevant Data for a user query. Nowadays, a large volume of clinical data in the medical field is generated, and retrieving the relevant data for user queries is a challenging task. Various issues and future enhancements were discussed to meet the challenges efficiently and effectively by an EHR search engine technology. Still, the limitation is constructing Information recovering functionalities into electronic health records is very risky and there requires further work to better know the multiple requirements of the users to mine Information from narrative clinical documents.

REFERENCES

[1] R.C. Barrows Jr, M. Busuioc, C. Friedman. Limited parsing of notational text visit notes: ad-hoc vs NLP approaches. AMIA Annual Symposium Proceedings. 2000; 51–55.

[2] Judith L Bader, Mary Frances Theofanis. Searching for Cancer Information on the Internet: Analysing Natural Language Search Queries. Journal of Medical Internet Research. 2003;5:4.

[3] Ashish K Jha. The promise of electronic records: around the corner or down the road?. JAMA Journal of the American Medical Association. 2011; 306: 880-1.

[4] YanshanWanga, Stephen Wu, DingchengLi, SaeedMehrabi,Hongfang Liu. A Part-Of-Speech term weighting scheme for biomedical information retrieval. Journal of Biomedical Informatics. 2016; 63: 379-389.

[5] TielmanT.VanVleck, Lili Chan, Steven G. Coca, Catherine K. Craven, Ron Do, Stephen B. Ellis, et al. Augmented Intelligence with Natural Language Processing Applied to Electronic Health Records is Useful for Identifying Patients with Non-Alcoholic Fatty Liver Disease at Risk for Disease Progression. International Journal of Medical Informatics. 2019; 129.

[6] David A Hanauer, Gretchen Miela, Arul M Chinnaiyan, Alfred E Chang, Douglas W Blayney. The Registry Case Finding Engine: An Automated Tool to Identify Cancer Cases from Unstructured, Free-Text Pathology Reports and Clinical Notes. 2007; 205(5): 690-697.

[7] Gregg W, Jirjis J, Lorenzi NM, Giuse D. StarTracker: an integrated, web-based clinical search engine. In American Medical Informatics Association(AMIA). Annual Symposium Proceedings. 2003;P- 855.

[8] M. Gubanov and A. Pyayt. “Medreadfast: A structural information retrieval engine for the big clinical text". IEEE 13th International Conference on Information Reuse and Integration (IRI). 2012.

[9] Johnston D, Kaelber D, Pan EC, Bu D, Shah S, Hook JM, Middleton BA, A framework and approach for assessing the value of personal health records (PHRs). In AMIA, Annual Symposium Proceedings 2007, P. 374–378.

[10] Rachel L. Richesson. An informatics framework for the standardised collection and analysis of medication data in networked research. Journal of Biomedical Informatics. 2014;52 4-10.

[11] Luis Marco-Ruiz, David Moner, José A. Maldonado, Nils Kolstrup. Archetype-based data warehouse environment to enable the reuse of electronic health record data. International Journal of Medical Informatics. 2015; 702-714.

[12] N. A. Farooqui and R. Mehra. "Design of A Data Warehouse for Medical Information System Using Data Mining Techniques". 5th IEEE International Conference on Parallel Distributed and Grid Computing. 2018.

[13] D.C. Cron, D.M. Coleman, K.H. Sheetz, M.J. Englesbe, S.A. Waits. Aneurysms in abdominal organ transplant recipients. Journal of Vascular Surgery. 2014;59(3):594-598.

[14] Papagerakis S, Bellile E, Peterson L.A, Pliakas M, Balaskas K, Selman S, et al. Proton pump inhibitors and histamine 2 blockers are associated with improved overall survival in patients with head and neck squamous carcinoma. Cancer prevention research (Philadelphia, Pa.) journal. 2014;7:1258–1269.

[15] N. Talaat, S. Yapali, R. J. Fontana, H. S. Conjeevaram, A. S. Lok. Changes in characteristics of hepatitis C patients seen in a liver centre in the United States during the last

decade. Journal of Viral Hepatitis. 2014;22(5):481-488.
[16] David A. Hanauer, Qiaozhu Mei, James Law, RituKhanna, Kai Zheng. Supporting information retrieval from electronic health records: A report of University of Michigan's nine-year experience developing and using the Electronic Medical Record Search Engine (EMERSE). Journal of Biomedical Informatics. 2015;55:290–300.
[17] Brian L, Hazlehurst, Stephen E. Kurtza, AndrewMasica, Victor J. Stevens, Mary Ann McBurnie, et al. CER Hub: An informatics platform for conducting comparative effectiveness research using multi-institutional, heterogeneous, electronic clinical data. International Journal of Medical Informatics. 2015;84: 763-773.
[18] Samuel G, Finlayson, Mia Levy, Sunil Reddy, Daniel Rubin. Toward rapid learning in cancer treatment selection: An analytical engine for practice-based clinical data. Journal of Biomedical Informatics. 2016;104-113.
[19] Senathirajah Y, Kaufman D, Bakken S., User-composable electronic health record improves the efficiency of clinician data viewing for patient case appraisal: a mixed-methods study. EGEMS The Journal for Electronic Health Data and Methods. 2016;4(1):1176.
[20] Brecht Claerhout, DipakKalra, Christina Mueller, Gurparkash Singh, Nadir Ammour, Laura Melonia, et al. Federated electronic health records research technology to support clinical trial protocol optimisation: Evidence from EHR4CR and the InSite platform. Journal of Biomedical Informatics. 2019;90: 103090.
[21] David Perez-Rey, Ana Jimenez-Castellanos, Miguel Garcia-Remesal, Jose Crespo and Victor Maojo. CDAPubMed: a browser extension to retrieve EHR-based biomedical literature. BMC Medical Informatics and Decision Making. 2012.
[22] Biron. P, Metzger M.H, Pezet. C, Sebban. C, Barthuet. E, Durand. T. An information retrieval system for computerised patient records in the context of a daily hospital practice: the example of the Léon Bérard Cancer Center (France). Applied Clinical Informatics. 2014;5:191-205.
[23] KorayAtalag, Hong Yul Yang, Ewan Tempero, James R.Warren. Evaluation of software maintainability withopenEHR – a comparison of architectures. International Journal of Medical Informatics. 2014;83:849-859.
[24] Alberto G, Jácome, FlorentinoFdez-Riverola, AnáliaLourenço. BIOMedical search engine framework: lightweight and customised implementation of domain-specific biomedical search engines. Computer Methods and Programs in Biomedicine. 2016;131:63-77.
[25] Danilo Schmidt, KlemensBudde, Daniel Sonntag, Hans-Jürgen Profitlich, Matthias Ihle, Oliver Staeck. A novel tool for the identification of correlations in medical data by faceted search. Computers in Biology and Medicine. 2017; 85:98-105.
[26] Joshua D Stein, MoshiurRahman, Chris Andrews, Joshua R Ehrlich, ShivaniKamat, Manjool Shah, et al. Evaluation of an Algorithm for Identifying Ocular Conditions in Electronic Health Record Data. JAMA Ophthalmol. 2019; 37:491-497.
[27] David A Hanauer. EMERSE: The Electronic Medical Record Search Engine. AMIA 2006 Symposium Proceedings. 2006;Page - 941.
[28] Seyfried L, David A Hanauer, Nease D, Albeiruti R, Kavanagh J, Kales HC. Enhanced identification of eligibility for depression research using an electronic medical record search engine. International journal of medical informatics. 2009;78:13–18.
[29] Lawrence Chang, David Frame, Thomas Braun, Erin Gatza, David A Hanauer, Shuang Zhao, et al. Engraftment syndrome after allogeneic hematopoietic cell transplantation predicts poor outcomes. Biol Blood Marrow Transplant. 2014;20:1407–1417.
[30] Nong Chen, Dahanayake. Rethinking of Medical Information Retrieval and Access. IDEAS Workshop on Medical Information Systems: The Digital Hospital (IDEAS-DH'04). 2004.
[31] Tracy Edinger, Aaron M. Cohen, Kyle Ambert, Steven Bedrick, William Hersh. Barriers to retrieving patient information from electronic health record data: failure analysis from the TREC medical records track. AMIA Annual Symposium proceedings. 2012;180–188.
[32] Felix Köpcke, Stefan Kraus, Axel Scholler, Carla Nau, Jürgen Schüttler, Hans-Ulrich Prokosch, et al. Secondary use of routinely collected patient data in a clinical trial: An evaluation of the effects on patient recruitment and data acquisition. International Journal of Medical Informatics. 2013;185-192.

[33] Jiayue Zhang, WeiranXu, Jun Guo, Sheng Gao. A temporal model in Electronic Health Record search. Knowledge-Based Systems. 2017;126:56-67.
[34] Vijaya M Vemulakonda, Ruth A Bush, Michael G Kahn. Minimally invasive research?"Use of the electronic health record to facilitate research in pediatric urology. Journal of Pediatric Urology. 2018;374-381.
[35] Henning Müller, DevrimUnay. Retrieval From and Understanding of Large-Scale Multi-modal Medical Datasets: A Review. IEEE Transactions on Multimedia. 2017;19:2093–2104.
[36] S. Mukherjea, B. Bamba, P. Kankar. Information retrieval and knowledge discovery are utilising a biomedical patent semantic Web. IEEE Transactions on Knowledge and Data Engineering. 2005;17(8):1099-1110.
[37] EunjeongPark ,Hyo Suk Nam. A Service-Oriented Medical Framework for Fast and Adaptive Information Delivery in Mobile Environment. IEEE Transactions on Information Technology in Biomedicine. 2009;13(6):1049-1056.
[38] Sheau-Ling Hsieh, Wen-Yung Chang, Chi-Huang Chen, Yung-ChingWeng, Semantic similarity measures in the biomedical domain by leveraging a web search engine. IEEE Journal of Biomedical and Health Informatics. 2013;853-61.
[39] Carlos Roberto Valêncio, Rodrigo Dulizio Martins, Matheus Henrique Marioto, Pedro LuizPizzigatti Corrêa, Maurizio Babini. Automatic Knowledge Extraction Supported by Semantic Enrichment in Medical Records. International Conference on Parallel and Distributed Computing. Applications and Technologies. 2013.
[40] Haolin Wang, Qingpeng Zhang, Jiahu Yuan. Semantically Enhanced Medical Information Retrieval System: A Tensor Factorisation Based Approach. IEEE Access. 2017;5:7584-7593.
[41] Juan M. Silva , Jesus Favela. Context-Aware Retrieval of Health Information on the Web. Fourth Latin American Web Congress. 2006.
[42] Weider D. Yu, Seshadri K, Yilayavilli. A Semantic-Based Dynamic Search Engine Design and Implementation for Electronic Medical Records. 11th International Conference on e-Health Networking. Applications and Services (Healthcom). 2009.
[43] YaliniSenathirajah, David Kaufman, Suzanne Bakken. Cognitive Analysis of a Highly Configurable Web 2.0 EHR Interface. AMIA Annual Symposium proceedings. 2010;732–736.
[44] Andrew Tawfik, Karl M. Kochendorfer, DinaraSaparova, said alGhenaimi, Joi L. Moore, ''I Don't Have Time to Dig Back Through This'': the role of semantic search in supporting physician information seeking in an electronic health record. Performance Improvement Quarterly. 2014;26:75–91.
[45] Yunmei Shi, Xuhong Liu, YabinXu, ZhenyanJi. Semantic-based data integration model applied to heterogeneous medical information systems. The 2nd International Conference on Computer and Automation Engineering (ICCAE). 2010.
[46] Andrew A, Tawfik Karl M, Kochendorfer, DinaraSaparova, Said Al Ghenaimi, Joi L. Moore. Using semantic search to reduce cognitive load in an electronic health record. IEEE 13th International Conference on e-Health Networking, Applications and Services. 2011.
[47] William Hsu, Ricky K. Taira, Suzie El-Saden, Hooshang Kangarloo, Alex A. T. Bui. Context-Based Electronic Health Record: Toward Patient-Specific Healthcare. IEEE Transactions on Information Technology in Biomedicine. 2012;16:228 - 234.
[48] Sylvie Ranwez, Vincent Ranwez, Mohameth-François Sy, Jacky Montmain, Michel Crampes. user-centred and Ontology-Based Information Retrieval System for Life Sciences. Proceedings of the Workshop on Semantic Web Applications and Tools for Life Sciences, Germany. 2010.
[49] ZhihuiLuo , Riccardo Miotto, ChunhuaWeng. A Human-Computer Collaborative Approach to Identifying Common Data Elements in Clinical Trial Eligibility Criteria. Journal of Biomedical Informatics. 2013;46:33-9.
[50] Dongqing Zhu, Ben Carterette. Joint search in text and concept spaces for EMR-based cohort identification. IEEE International Conference on Bioinformatics and Biomedicine. 2013.
[51] Kai Zheng, Qiaozhu Mei, David A Hanauer. Collaborative search in electronic health records. Journal of the American Medical Informatics Association. 2011;282–291.

[52] Jiayue Zhang, Jimmy Xiangji Huang, Jun Guo, WeiranXu. Promoting electronic health record search through a time-aware approach.IEEE International Conference on Bioinformatics and Biomedicine. 2013.
[53] Julio Cesar Dos Reis, Rodrigo Bonacin, Edemar Mendes Perciani, Intention-Based Information Retrieval of Electronic Health Records. IEEE 25th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), 2016.
[54] Le Thi Hoang Diem, Jean-Pierre Chevallet, Dong Thi Bich Thuy. Thesaurus-based query and document expansion in conceptual Indexing with UMLS: Application in medical information retrieval. IEEE International Conference on Research, Innovation and Vision for the Future. 2007.
[55] KarthikNatarajan, Daniel Stein, Samat Jain, Noémie Elhadad. An analysis of clinical queries in an electronic health record search utility. International journal of medical informatics. 2010;79(07):515–522.
[56] Lei Yang, Qiaozhu Mei, David A. Hanauer. Query Log Analysis of an Electronic Health Record Search Engine. AMIA Annual Symposium proceedings. 2011;915-24.
[57] Gregory W Hruby, KonstantinaMatsoukas, James J Cimino, Chunhua Weng. Facilitating biomedical researchers' interrogation of electronic health record data: Ideas from outside of biomedical informatics. Journal
of Biomedical Informatics. 2016;60:376–384.
[58] Julia Hoxh, Chunhu Weng. Leveraging dialogue systems research to assist biomedical researchers' interrogation of Big Clinical Data. Journal of Biomedical Informatics. 2016;61:176-184.
[59] Spitz, A. L. Determination of the script and language content of document images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997; 19(3): 235-245.

[60] Wang, J. H., & Chang, H. C. Exploiting sentence-level features for near-duplicate document detection. In Asia Information Retrieval Symposium, Springer, Berlin, Heidelberg, 2009; 205-217.

[61] Hu, J., Kashi, R. S., Lopresti, D. P., & Wilfong, G. Medium-independent table detection. In Document Recognition and Retrieval VII. International Society for Optics and Photonics, 1999; 3967:291-302.

[62] Conrad, J. G., Guo, X. S., & Schriber, C. P. Online duplicate document detection: signature reliability in a dynamic retrieval environment. In Proceedings of the twelfth international conference on Information and knowledge management, 2003; 443-452.

[63] Chowdhury, A., Frieder, O., Grossman, D., & McCabe, M. C. Collection statistics for fast duplicate document detection. ACM Transactions on Information Systems (TOIS),2002; 20(2): 171-191.

[64] Lopresti, D. P. A comparison of text-based methods for detecting duplication in scanned document databases. Information Retrieval, 2001;4(2):153-173.

[65] Timothy, D. J., & Tosun, C. Arguments for community participation in the tourism development process. Journal of Tourism Studies, 2003;14(2): 2-15.

[66] Ho, P. T., & Kim, S. R. Fingerprint-based near-duplicate document detection with applications to SNS spam detection. International Journal of Distributed Sensor Networks, 2014;10(5):612970.

[67] Janani, M. R., & Vijayarani, S. An Efficient Algorithm for Document Clustering in Information Retrieval. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2016;4.

[68] YaqiangWang, Zhonghua Yu, Li Chen, Yunhui Chen, Yiguang Liu, Xiaoguang Hu, et al. Supervised methods for symptom name recognition in free-text clinical records of traditional Chinese medicine: An empirical study. Journal of Biomedical Informatics, 2014;47:91-104
[69] Ming Cheng, Liming Li, YafengRen, Yinxia Lou, JianboGao, Yafeng Ren, et al. A Hybrid Method to Extract Clinical Information From Chinese Electronic Medical Records. IEEE Access, 2019;7:70624 – 70633.
[70] Lijun Liu, Li Liu, Xiaodong Fu, Qingsong Huang, Xianwen Zhang, Yin Zhang. A cloud-based framework for large-scale traditional Chinese medical record retrieval. Journal of Biomedical Informatics. 2018;77:21-33.
[71] Jiao Zhang, Dan Chang. Semi-Supervised Patient Similarity Clustering Algorithm Based on Electronic Medical Records. IEEE Access, 2019;7:90705 - 90714.

[72] Al Omar A., Rahman M.S., Basu A., Kiyomoto S.. Medibchain: a blockchain-based privacy-preserving platform for healthcare data. In: International Conference on Security, Privacy and Anonymity in Computation, Communication and Storage, Springer.2017;534–543.
[73] Lee, S.H., Yang, C.S. Fingernail analysis management system using microscopy sensor and blockchain technology. International Journal of Distributed Sensor Networks. 2018;14 (3).
[74] Zheng, X., Mukkamala R.R., Vatrapu R., Ordieres-Mere J. Blockchain-based personal health data sharing system using cloud storage. In: 2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom). IEEE. 2018;1–6.
[75] Yue X., Wang H., Jin D., Li M., Jiang W. Healthcare data gateways: found healthcare intelligence on the blockchain with novel privacy risk control. J. Med. Syst. 2016;40(10):218.
[76] Guo R., Shi H., Zhao Q., Zheng, D. Secure attribute-based signature scheme with multiple authorities for blockchain in electronic health records systems. IEEE Access.2018;6:11676–11686.
[77] Liu, J., Li, X., Ye, L., Zhang, H., Du, X., Guizani, M. Beds: A blockchain-based privacy-preserving data sharing for electronic medical records. In: 2018 IEEE Global Communications Conference (GLOBECOM). IEEE, 2018;1–6.
[78] Zhao H., Zhang Y., Peng Y., Xu R. Lightweight backup and efficient recovery scheme for health blockchain keys. In: 2017 IEEE 13th International Symposium on Autonomous Decentralized System (ISADS). IEEE, 2017; 229–234.