ISSN:0975-9646

A Survey on Accessing, Retrieval Mechanisms and Challenges in Biomedical Literature

Thayyaba Khatoon Mohammed[1], Dr. A. Govardhan[2]

Research Scholar, JNTUH
Principal, JNTUCEH

Abstract: Information Retrieval methods occupy a key role in the present knowledge society. Access to Literature, analysis of data and integration of knowledge are key components of biomedical research. Although there are several methods and search tools exist, there is a huge need and demand for better approaches. Many of the researchers and scientists are losing valuable time and other resources because of the lack of access to right literature. This paper analyses the critical role of knowledge access mechanisms and challenges that exist in biomedical research with a focus on omics disciplines. Further, it surveys the current state of art to address the needs of biomedical researchers for better and efficient access to information and better integration of data and knowledge that is extracted from biomedical literature. The different data sources existing on biomedical literature, different existing search tools used to access the biomedical citations, bibliographic data sources, full text literature, citation data bases existing and challenges in accessing the literature are presented. This paper also gives a critical analysis on various methods used in search and query optimization for reducing the response time of query processing with challenges in accessing biomedical literature.

Key terms: Text mining, biomedical data bases, medical literature search tools, Access mechanisms, Query optimization, similarity search, performance improvement.

1. INTRODUCTION

There are several publishers available on biomedical data. Literature on biomedical data includes several concepts of medicine. data base search includes finding the relevant biomedical concepts based on user query, or finding high impact factor biomedical articles, or retrieving a query answer on biomedical literature. In the process of innovations in biomedical field, the basic work of any researcher is to find the past works related to the specific concept in biomedical field from the beginning of the work to the implementation of study. Biomedical literature access is essential for biomedical researchers not only scientists but also data base curators, clinicians for taking crucial clinical decisions. There are several medical journals exists whose electronic form is available online. User can use the basic search mechanisms to search for the literature. Literature search may be presented in three ways i) citation data base search ii)full text achieves iii) bibliographic data base search.

Figure 1 [38] illustrates the scope of biomedical literature and classification of biomedical informatics. Biomedical literature access is essential for several types of users including biomedical researchers, clinicians, database curators, and bibliometricians. In the past few decades, several online search tools and literature archives, generic as well as biomedicine-specific, have been developed. We present this chapter in the light of three consecutive steps of literature access: searching for citations, retrieving full-text, and viewing the article. The first section presents the current state of practice of biomedical literature access, including an analysis of the search tools most frequently used by the users, including NCBI's PubMed, Google Scholar, Web of Science, Scopus, and Embase, and a study on biomedical literature archives such as PubMed Central.

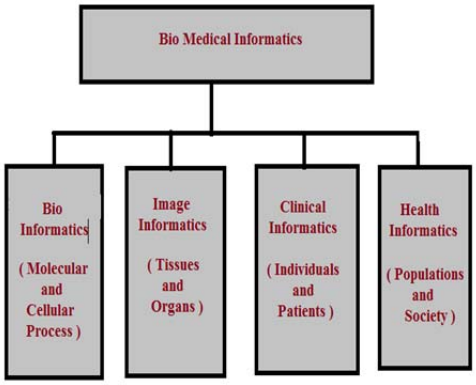


Fig.1. Classification of biomedical informatics

The biomedical community is attached to peer review and publication in high-quality journals has a central role in a researcher's career. Set against this traditional journal subscriptions are stretching the finances of universities and research institutions. Furthermore, it is important that availability and accessibility are not adversely affected by copyright issues. This adds to the pressure to seek new methods of publishing research findings in an openly accessible way.

From many years the biomedical scientists and researchers has benefited from PubMed [1], a free bibliographic database provided by the National Library of Medicine (NLM) [1] in the US. Since 2000, a direct link to full-text articles in PubMed Central (PMC) [19] [3] has been provided. The concepts of PMC is derived from both self-archiving by researchers and publishers [3] making their journal content available in the repository (a list of participating journals is available in PMC[3]). The entire content of full open access journals is directly deposited in PMC by publishers immediately upon publication. There also exist a number of subscription-based journals which make their full content available in PMC after an embargo

period of 6-12 months (see the PMC journal list). The existence of this structured international disciplinary archive makes the biomedical field more advanced than most others; it has revolutionized medical publishing by radically changing the way research results are circulated and has undoubtedly facilitated biomedical research, innovation and discoveries.

2.BIOMEDICAL LITERATURE DATA SOURCES

2.1 UNESCO: UNESCO provides open access and has a dedicated Global Open Access Portal (GOAP)[4] in operation since April 2011 that is expanding rapidly with the addition of new open access initiatives and projects at a global level. The organization places particular emphasis on scientific information emanating from publicly funded research, and works with partners to improve awareness of the benefits of open access among policy-makers, researchers and knowledge managers. UNESCO also recently released its 'Policy Guidelines for the Development and Promotion of Open Access'[5].

Figure 1. A typical storage room for medical records

2.2 The US NIH Public Access Policy: In 2007 US Congress passed a law requiring that all NIH-funded manuscripts be made freely available in PMC no later than 12 months after publication94. The NIH policy honors, and is consistent with, US copyright law. It is based on the principle that an author, as the creator of the work, holds the copyright in the original paper. Traditionally, authors transfer all their rights to the publisher when an article is accepted for publication in a journal[6].

Example of a UK PubMed Central (UKPMC) European funder: the Austrian Science Fund (FWF):

The Austrian Science Fund (FWF) has an open access policy which covers all disciplines130. he FWF joined UK PubMed Central (UKPMC) in March 2010. UKPMC guidelines were developed for Principal Investigators (PIs) funded by the FWF. Now, around 3,000 papers with FWF acknowledgement are deposited in PMC. The compliance rate of the FWF is one of the highest of all UKPMC funders. In addition to a reminder system developed with UKPMC and advice offered to PIs, this high rate is attributable to 3 factors [9].

First, publication costs for FWF projects are not budgeted within a project but are covered by extra funds provided by the FWF for up to 3 years after conclusion of the project. Second, the FWF covers costs for publishing in both gold open access journals and the so-called hybrid journals offered by some publishers. Third, for Wiley-Blackwell and Elsevier133 publication costs are transferred directly from the FWF to the publishers, with no charge payable by authors. Although this procedure incurs high costs, it makes it much easier for PIs to use FWF funding and deposit papers in UKPMC[10].

2.3 European Commission: The European Commission first showed interest in 2005 with a study on the economic and technical evolution of scientific publication markets in Europe[7]. Open access to research results rapidly became an important aspect of the broader area of knowledge circulation, which the Commission undertook to explore in a coherent way within the Seventh Framework Programme (FP7) and in line with specific statements made in the EU communications *Innovation Union*[7] and *A Digital Agenda for Europe*[7].

2.3.1European Research Council:

The open access policy of the European Research Council (ERC) is spelled out in the ERC Scientific Council guidelines dated 20 June 2012, where the Council reaffirmed its open access policy and strongly encouraged ERC-funded researchers to make their publications publicly available in open access as soon as possible, and no later than 6 months after the official publication date of the original article, using discipline specific repositories [6].

2.3.2 Science Europe (formerly EUROHORCs):

In 2008, the former European Heads of Research Councils (EUROHORCs, now regrouped under Science Europe [8]) issued a statement including a set of recommendations for the adoption of a minimal standard on open access by its 50 Member Organizations (research performing and funding organizations).

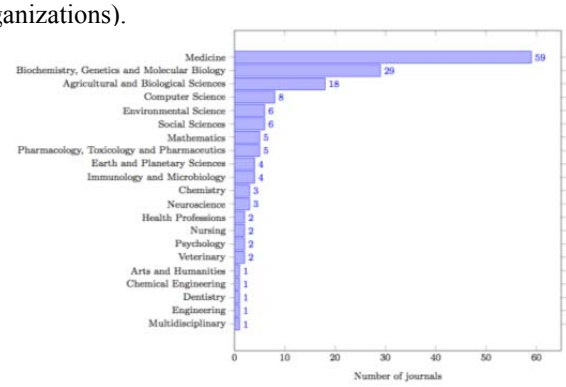


Figure 2. Number of % published journals by topic.

We are also having Thomson Reuters, ISI Web of Knowledge, NLM–National library of Medicine databases, MEDLINE, PubMed Central, ELSEVIER as the host of Databases, EMBASE, Scopus. In biological research, there are thousands of specialized data repositories, focusing on particular molecules, organisms or diseases, which offer sets of richly annotated records [11].

To ensure data of the highest quality, manual data entry and curation (annotation) processes are generally performed on these databases. Database curators are domain experts who search biomedical research literature for facts of interest, and manually transfer knowledge from published papers to the database[12]. This helps experts to consolidate data about a single organism or a single class of entity, often in conjunction with sequence information. Most importantly, this process makes the information searchable through a variety of automated techniques given that the curators use standardized terminologies or ontologies. However, as the volume of biomedical literature increases, so does the burden of curation, making annotation databases incomplete and inconsistent with the literature [12].

2.4 Thomson Reuters Integrity: provides researchers with reliable, detailed information, from the perspective of a scientist, across multiple disciplines to support successful drug research and development. By integrating biological, chemical and pharmacological data on more than 457,000 compounds with demonstrated biological activity and over 297,000 patent family records[13], Integrity provides a unique knowledge solution designed to empower your discovery and development activities.

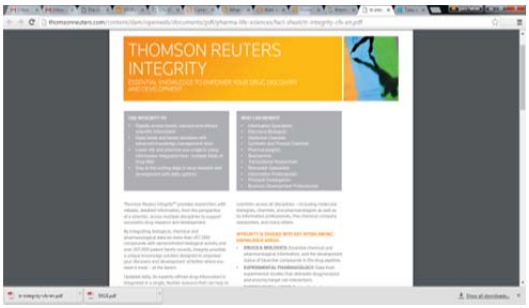


Figure 3: Thomson Reuter's integrity [13]

Using the resources of Thomson Reuters integrity the user group get benefited are biologists, medical scientists, chemists, pharmacologists, biochemist researcher [13].

2.5 Elsevier: Elsevier was founded in 1880 and took the name from the Dutch publishing house Elsevier which has no connection with the present company. The Elsevier family operated as booksellers and publishers in the Netherlands. The expansion of Elsevier in the scientific field after 1945 was funded with the profits of the newsweekly *Elsevier*, which first issue appeared on 27 October 1945[14].

Elsevier has two distinct operating divisions: Science & Technology and Health Sciences. Products and services of both include electronic and print versions of journals, textbooks and reference works and cover the health, life, physical and social sciences.

Elsevier products and services include VirtualE, SciencDirect, Scopus, Scirus, EMBASE, Engineering Village, Compendex, Cell, SciVal, Pure, and Analytical Services[14].

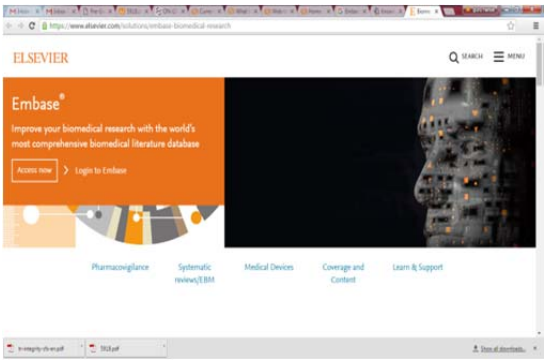


Figure 4: Elsevier solutions EMBASE biomedical literature [15].

2.5.1 EMBASE:

EMBASE for Excerpta Medica data BASE)[15] is a pharmacological and biomedical database of published literature developed to support information managers and pharmacovigilance in comply with the regulatory requirements of a licensed medicine.

Apart from this 28 million reports, Embase's database constantly rises every year at a rate of over 900,000 records. This Huge information is used in both educational and professional environments to retrieve any published biomedical or medicine related information. At present, Embase provides further customization for a personal experience like implementing an RSS feed and email alert system that notify a user of an update he or she may embrace interest in. With new medicine and disease related information all the time released, Embase is restructured every day to provide a complete and reliable basis of information[15].

2.5.2 SCOPUS:

Scopus is the leading abstract and citation database of peer reviewed literature, Books, Scientific journals and conference proceedings. Delivering a broad overview of the world's study and research output in the fields of science, technology, medicine, social sciences, arts and humanities, Scopus features stylish tools to follow, analyze and think about research [17].

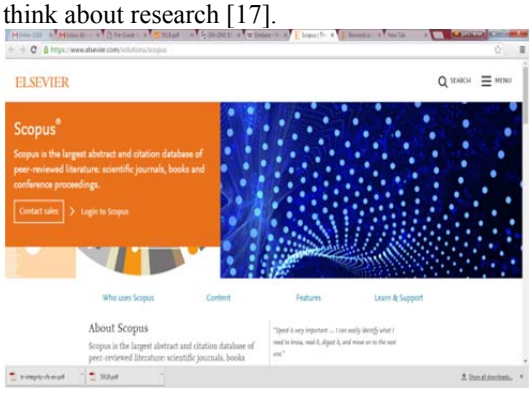


Figure 5: Elsevier solutions SCOPUS biomedical citation data base

SCOPUS is a bibliographic information base that contain citations and abstracts for academic journals and articles. It cover nearly 22,000 titles from about 5,000 publishers , of which around 20,000 are peer reviewed journals in the technical, scientific, medical, and social sciences [17]

3.LITRATURE SEARCH TOOLS

A search tool provides contact point to biomedical literature records. The search tool contains a medicinal citation database developed by indexing literature. The search interface provides mainly services as:(i) functionality to support query processing for the keyword exploration, similarity search (e.g. search by abstract, journal, title, author, etc.), (ii) Ranking mechanism for catalog the keyword search citations appropriate to the query, to filter and resorting the keyword search results; in addition to indexed information, each reference contains a link to get back the full literature piece of writing on a full text library.

PubMed [1] is the extensively used search tool devoted to life sciences and biomedical literature. Launched in 1996, PubMed is a openly available reference database developed and maintained by the U.S. National Library of Medicine. To date, PubMed contain more than 22.8 million documents for biomedical literature belongs to MEDLINE indexed journals and manuscripts deposited in PubMed Central, and also the NCBI Bookshelf. PubMed articles are indexed by the restricted vocabulary thesaurus, the search algorithm is based on automatic term mapping algorithm. The PubMed citation database is updated every day. PubMed citations date back to the early 1950s and around half a million date back to 1809. The PubMed edge offers the token search, and allows the most advanced queries by varied fields such as publication date and author name, PubMed entering date, grant number, editor, status of MeSH indexing for MEDLINE citations, etc. A remarkable feature of PubMed is the associated citations algorithm based on document similarity [18].

3.1 PMC(PubMed Central):[17][18] PMC is a free library of life sciences and biomedical journal literature at the U.S. National Institutes of Health's National Library of Medicine (NIH/NLM). It is a storage area for journal literature deposited by participating publishers, as well as for writer manuscripts that are submitted in compliance with the NIH Public Access Policy[17] and similar policies of other research funding agencies. PMC does not publish journal or articles itself.

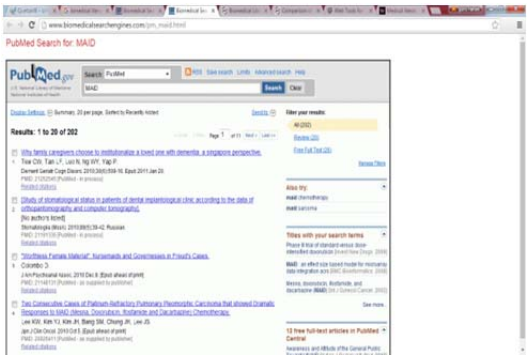


Figure 6: Pubmed central example for MAID (a cyclin D binding protein)[22]

PubMed found [22] 202 items from this search. It is not clear without further assessment why some of the items were found, but it is clear that the results consist of maids as support persons. PMC offer publishers many ways to participate and put their content in the library. Complete text XML is at the heart of PMC's design philosophy hence there is a need for all current content. Free right to use is also a core principle of PMC [19].

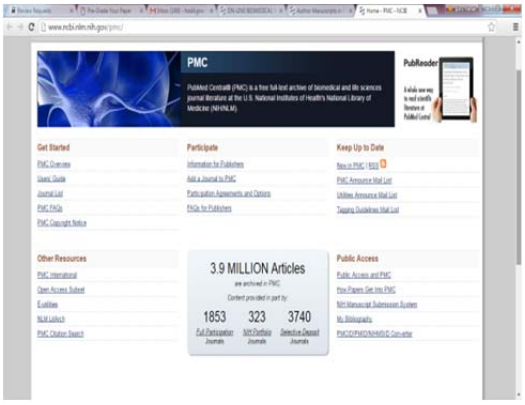


Figure 7 : NCBIs NIH/NLMs PubMed Central[19]

3.2 SciGlobe Search: SciGlobe[20] is superior search steam engine for biomedical literature which uses power of Conceptual Relationship, Searching, and a unique linguistic technology to find rapidly and carry the most important applicable search results. The benefit of the SciGlobe linguistic search compared to traditional key word[20] searching is its capability to give the intention of the author to link the search terms jointly. SciGlobe search provides superior quality content and not just very long list of publications.

The SciGlobe Advantage

[20]Intuitive search term method just asks SciGlobe a query (Does cancer cause diabetes?) Conceptual relationship searching decrease search end result from millions of hits to just hundreds of the .relevant papers stored search capability that can be configured for sending automatic weekly update of new documents. This enables you to link your search consequences to your collection subscriptions for direct access to full text documents or items. This Provides access to a huge collection of abstracts from PubMed PLUS NIH grant applications, TOXLINE, and more than 2.5 million full-text articles Unlimited and free of cost.

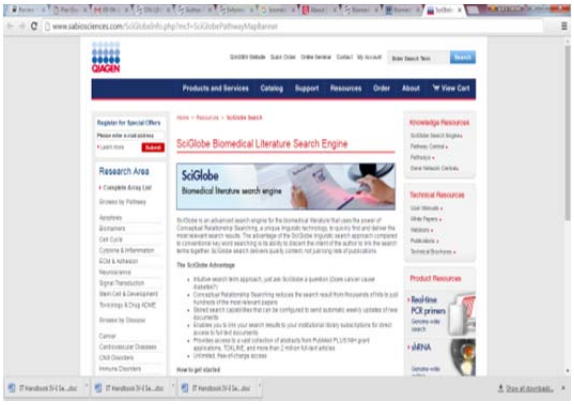


Figure 8: SciGlobe search engine for biomedical literature[20]

Search tools	Google scholar	Quertle	PubMed central	ISI Web of knowledge	Scirus	springer	EmBase	Scopus
Founder	Google Inc.	Quetzal	U.S National Institute of Health	Thomson Reuters	Elsevier	Springer	Elsevier	Elsevier
Year	2004	2008	1996	2004	2001	1842	2000	2004
Covering content of journals	Peer reviewed journals and Non-peer reviewed journlas		Medline Indexed journals	8,500 science journals	Peer reviewed journals	Peer reviewed journals	8,400biomedical and phrenological journals	Peer reviewed journals
Temporal coverage	unknown	Unknown	1809 to present	1900 to present	1995 to present	unknown	1880 to present	1995 to present
Biomedical citations	unknown	Unknown	26 million	90 million	55million	unknown	25 million	55million

Table 1. Summary various search tools

3.3 Elsevier Search Tools: EMBASE[18] is a subscription-based biomedical database developed by Elsevier in 2000. This search service focus was developed mainly for biomedical and clinical practice with specific focus on drug safety, drug and medicine discovery, development, and pharmaco vigilance research. EMBASE contains 25 million indexed records and indexes complete text documents from 8,306 journals, out of which 7,203 publish english language articles. EMBASE is often compared with MEDLINE, and contains 500 thousand records, covers 2,000 journals not included by MEDLINE. The EMBASE database is updated regularly and nearly 100 thousand records are added every year. EMBASE has digitally scanned the articles from 1947 to 1973. A temporal coverage of EMBASE dates back to 1947, some articles may also date back to 1880s. The records are indexed by Emtree thesaurus for medicine and chemical database. This allows for profound indexing of documents and flexible token searching using term mapping. The search capability is improved using auto complete and synonym hint features.

Web of Science [18], initiated and developed by Thompson Reuters in 2004, is a citation database which cover 12,000 top tier international and regional journals. in every area of humanities, natural sciences, social sciences, and arts[18].

Scopus [21], launched in 2004 by Elsevier, is a citation database for peer-reviewed literature from life sciences, physical sciences, health sciences, humanities and social sciences, Scopus, By November 2012, included citations from 400 trade publications , 19,500 peer-reviewed journals, 360 book series, and is updated one-to-two times weekly also temporally, citations date back to 1823. Scopus contain more than 18,300 citations from the health, life, and physical science areas.

3.4 Google Scholar [2] is a Web search engine owned by Google Inc and launched in 2004. Google Scholar indexes items from many disciplines from peer reviewed online journals of American and European publishers, scholarly books, and non peer reviewed journals.

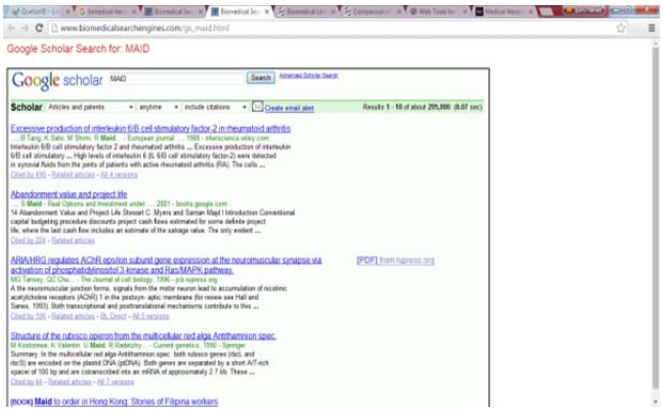


Figure 9: google scholar search example for MAID (a cyclin D binding protein)[22]

Google Scholar[22] found "about 295,000" items from this search. Many results are from the author name Maid. Result #3 has nothing to do with MAID nor is the term MAID (or any of its synonyms) found anywhere in the article. Result #5 is about household help. This indicates more hits are not better

3.5 QUERTLE:

Quertle[22] is a high-speed growing option in the biomedical search engine region Quertle is the only major biomedical search engine that focus on semantic searching to improve the relevance of results. Quertle is freely accessible.

Search and ranking algorithms: Quertle uses natural language processing to match your search against actual assertion made by the writer, instead of keyword matching or proximity. Quertle ranks the results based on relevance,

where relevance is mainly determined by how many times the article supports the assertion.

Content: Quertle's content covers all of PubMed plus a large number of full text documents, the NIH Reporter database of NIH grants, biomedical news reports, the TOXLINE database of toxic effects of chemicals, and whitepapers like research information from biomedical companies.

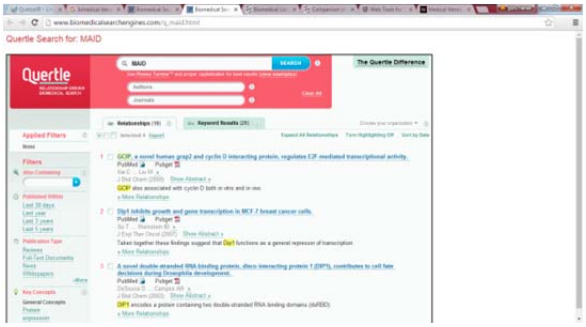


Figure 10: Quertle search example search example for MAID (a cyclin D binding protein)[22]

Quertle [22]found 18 items that refers to the MAID protein. Because Quertle automatically searches for all aliases of search terms, the items were found because GCIP, Dip1, etc. are equivalent to MAID. The 18 results are extremely relevant and hold assertions by the authors about this protein. An additional 6 results are found by traditional keyword searches (the Keyword Results tab)

4.NCBI'S SEARCH TOOLS FOR BIOMEDICAL LITERATURE SEARCH

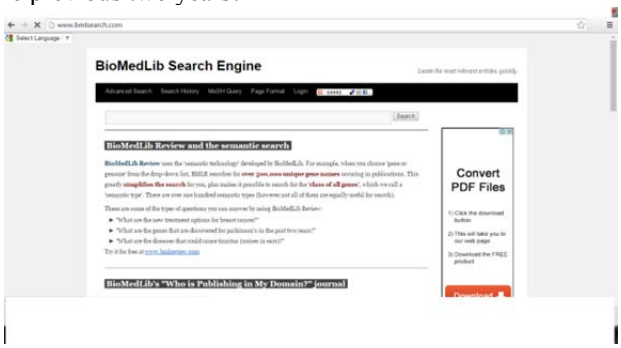
National center for biomedical information in that BioMedlib uses the 'semantic technology' (meaningful technology) developed by BioMedLib.

For ex: when you select 'gene or genome' from the drop down list of items, then BMLR searches for over 500,000 unique gene identification names. It reduces the complexity of the search and makes it promising to search for the 'class of all genes';

Some of questions can be answered by using BioMedLib Review:

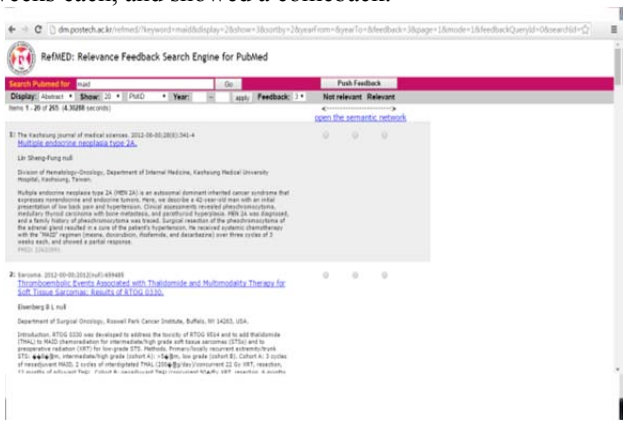
► "What are the possible new treatment options for breast cancer?"

► "What are the genes that are exposed for parkinson's in the previous two years?"

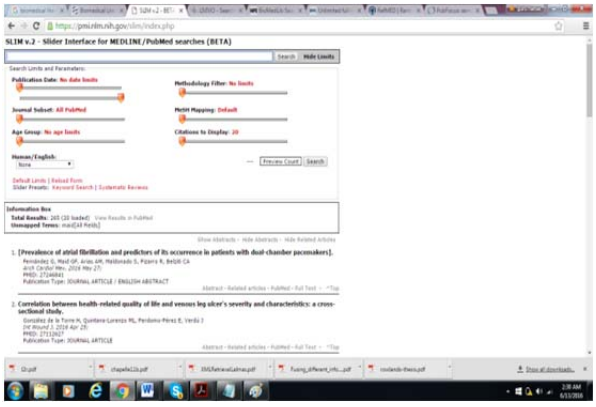


4.1 ipubmed : PubMed is a Windows based software system for automatically receiving e mail alerts about new publications listed on PubMed. The e mail alerts have links to newly accessible abstracts listed on PubMed telling publications those were selectively returned from a particular list of queries. **PubMed** AlertMe is freely availble software.

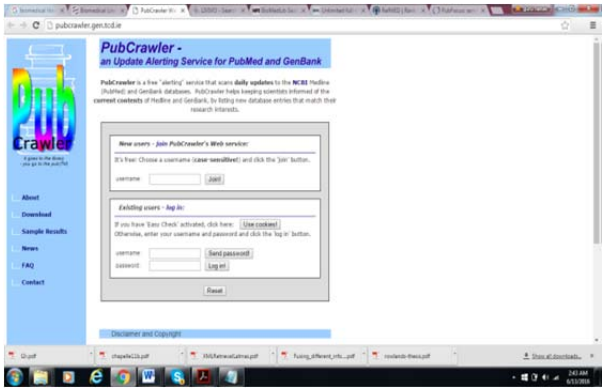
4.2 Refmed : Relevance Feedback Search Engine RFSE for PubMed, applied for example on multiple endocrine neoplasia type 2A (MEN 2A) is an autosomal dominant hereditary cancer syndrome that expresses non endocrine and also endocrine tumors. Here, we illustrate a 42 year old man with a first presentation of low back pain and hypertension. Clinical assessments and evaluations revealed medullary , heochromocytoma, thyroid carcinoma with bone metastasis, and parathyroid hyperplasia. Surgical resection of the heochromocytoma of the adrenal gland resulted in a restore to health of the patient's hypertension. He received the systemic chemotherapy with the "MAID" regiment about 3 cycles of 3 weeks each, and showed a comeback.



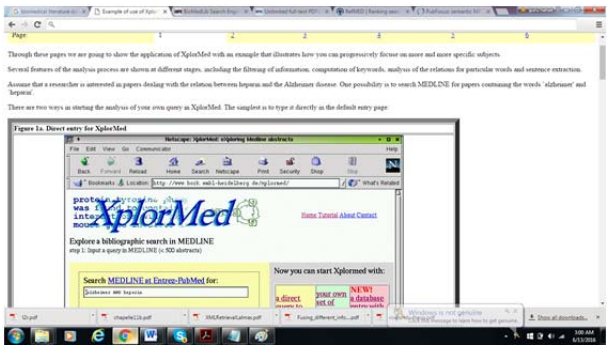
4.3 SLIM : Slider Interface for MEDLINE/PubMed searches (BETA). Features interactive slider bars in the search form.



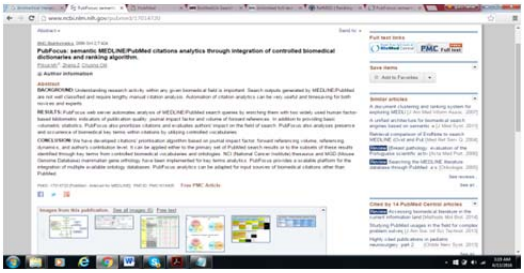
4.4 **PubCrawler** : It a cost free "alert" service that scan frequently updates to the NCBI Medline and GenBank databases. It also helps in keeping scientists informed about the the recent contents of Medline and GenBank, by listing new database entries that match with their research interests. This service is cost free.
It is given to the public in the hope that it will be helpful and useful.



4.5 **XplorMed:** This shows the way we can gradually focus on more and many more specific subjects disciplines. Many features of the analysis processes are shown at diverse stages, including the filtering of information, calculation of tokens, and analysis of the relations for sentence extraction and particular terms.
There are two ways to start the study of your own query in XplorMed. The easy type is straight in the default entry page and, the second way starts to perform the query at NCBI and to stock up it in your own hard disk. The name of that file will be provided next to XplorMed.

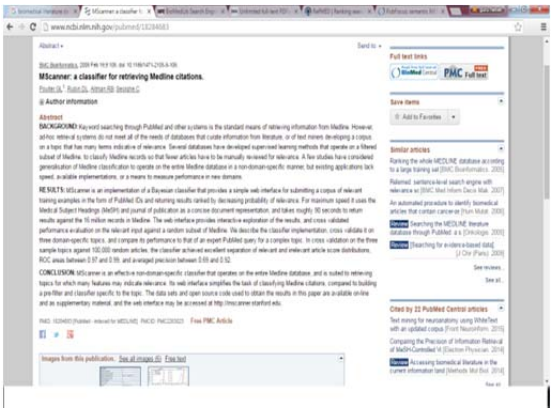


4.6 PubFocus: PubFocus web server automate examination of MEDLINE/PubMed search queries by elevating them with two widely used human factor based bibliometric indicators of publication excellence: volume of forward references and journal impact factor. In addition to given that basic volumetric statistics, Pub Focus also prioritizes citations and evaluates authors' impact on the field of search.it also analyses occurrence and presence of biomedical tokens within documents by utilizing controlled vocabularies .it has also developed citations' prioritization algorithm based on forward referencing volume, journal impact factor, referencing dynamics, and author's contribution level. It can be applied either to the main set of PubMed search consequences or to the subsets of these results identified through tokens from controlled biomedical vocabulary and ontology.

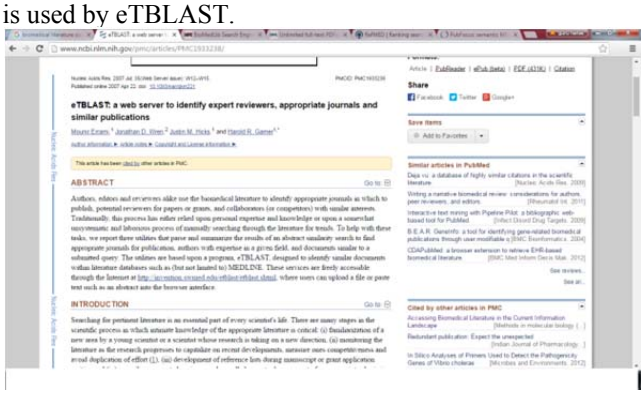


4.7 **MedEvi**: Apart from retrieving textual evidence and proof of relationships between bio medical concepts from Medline, the on hand search engines such as PubMed, have limits when applied for the reason of seeking textual evidence of relationships between agreed concepts. The limitations are largely due to the difficulty that the search engines do not productively deal with multi term queries which mean semantic relations stuck between the terms. To talk to this problem, we present MedEvi, a new search engine that imposes positional constraint on occurrences of similar multi-term queries, based on the study that terms with semantic relations which are clearly stated in text are not originate too far from each other.

4.8 MScanner: MScanner is an helpful non-domain-specific classifier that operate on the whole Medline database. It is suitable to retrieve topics in which most features point to relevance. Its web interface simplifies the task of classifying Medline documents. The data sets and open source code used to get the results in this paper are available on line and as supplementary and additional material.



4.9 ETBLAST: a web server to recognize authority reviewers, suitable journals and alike publications. We have developed ETBLAST, That uses a mixture scheme to take out and weight tokens enclosed within the submitted query to make out a subset of literature in Medline, and performs a sentence arrangement to compute a final quantitative gain as a measure of similarity, relevancy and, presumably. This tool then output a list, similar to PubMed, but ranked instead by this similarity score. It shall be renowned that ETBLAST and PubMed both of them find similar abstracts, but by dissimilar methods and PubMed's Related Links is restricted to only finding similarity among the records are at present in Medline, not arbitrary text, as is used by eTBLAST.



5. BIOMEDICAL LITERATURE ARCHIEVES

A text search tool is integrated with multiple text library where articles can be retrieved for additional use. As of June 18 2013, out of the 22.8 million documents in PubMed, 4 million documents are associated to their free full text archives. Out of the documents linked to free full text archives, 2.30 million are archived from the PubMed Central [3] literature library, and the remaining either contain direct links to journal's website (e.g., BMJ Journals Journal of Cell Anticancer Research, Biology, Oncotarget) or to wide-ranging text archives developed by most important publishing companies.

Springer's SpringerLink [29] launched in 1996. It's archives full text content accessible from 1996. SpringerLink cover just about 7.7 million full text articles from electronic books and journals from all disciplines and fields. Out of which 6.4 million are classified under the category of chemical, life , biomedical, public health, and medical sciences. Supplementary additional material also archived with each article.

Every literature archive provides one or more media or formats where the retrieved literature can be read by the user. at present, the aforesaid literature archives offer at least four types of viewing media. The first and foremost view is the classic view wherein the object can be viewed on the archive website itself. This view may not have any page breaks and needs to be read by scrolling up and down through a single long page.

This is the default HTML format view by most literature sources for reference. The second viewing media is the PDF format (.pdf extension) wherein the article can be downloaded onto a device.

Literature Archive (Provider, Year)	Temporal Coverage	Full-Text Biomedical Articles and Archive Coverage (Approx)	Viewing Media
PubMed Central (U.S National Library of Medicine)	1950-Present	2.7 million from 3,700 journals including full participation, NIH portfolio, selective deposit	Classic, PDF, EPUB, PubReader
CINAHL Plus with full-text(EBSCO,2010)	1937-Present	768 journal and magazines, 275 books and monographs from nursing and allied health disciplines.	PDF
Springer Link(Springer,1996)	1860-Present	6.4 million from biomedical, chemical,life,public health and medical sciences.	Classic, PDF, EPUB
ScienceDirect (Elvier,2000)	1823-Present	8,077 life and health sciences journals and book chapters.	PDF
Wiley Online Library(Wiley-Blackwell,2010)	Unknown	Journals, online books and reference works (biomedical coverage unknown)	Classic, PDF

Table 2. Comparison of biomedical full-text literature archives

6.FUTURE TRENDS TO IMPROVE BIOMEDICAL LITERATURE ACCESS

In terms of the search tools, we can expect to see more smart applications and reader friendly based on advanced IR and NLP techniques in order to help readers find and digest article more efficiently and proficiently. Furthermore, By the use of social media such as blogging and tweeting, innovative ways of sharing and also recommending papers will gain more importance in the future, in addition to the traditional search-based mechanism.

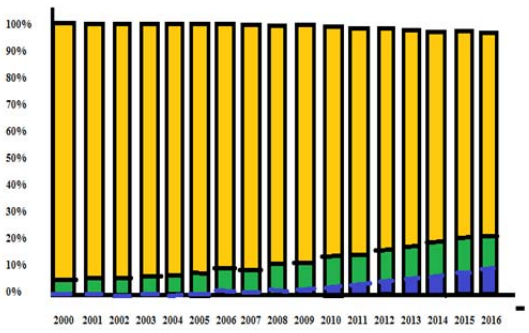


Figure 3. Year-on-year growth in availability of full-text articles in UKPMC and proportion of those that are open access in comparison to the number of abstracts available in PubMed. Of the 14,25,000 articles indexed by PubMed in 2016.

In the future, biomedical literature search could also be personalized. That is, search results are personalized towards the interests of individual researchers based on their own work and/or past searches. To be more precisely, in other terms, same query by two different users might return different search results.

7. TEXT MINING [35]

Our situation is that we are data rich, but information poor. The abundance of data, coupled with the need for powerful data analysis tools, has been describe as a *data rich but information poor* situation. Simply stated, data mining refers to extracting or "mining" knowledge from large amounts of data. The mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining.

Navigation Cost [35]

The following themes emerged:

A. Costs include access, transaction, entry, staff and infrastructure costs

B. Benefits include: efficiency exploring new horizons; improved research and evidence base; and improving the research process and quality. Unlocking hidden information and developing new knowledge;

C. Broader societal and economic benefits are also highlighted. For example cost savings and productivity gains, innovative service development, business models, and medical treatments

8. COSTS ASSOCIATED WITH BIOMEDICAL TEXT MINING [34, 35]

Access costs

Where text mining explores copyrighted materials, the copyright holders may need additional payment to allow their matter to be used in text mining. As several Consulters highlighted, this means that most text mining is partial to exploring Open Access documents where no additional charges are incurred.

Transactions costs

Transaction costs in this context relate to the effort necessary to enable text mining to take place. This is mainly associated with obtain authorization to mine particular corpora of documents. For example, establishing authorization to digitize alone takes about the equivalent of 1 FTE 10 as part of the national SHERPA/RoMEO service which offer information about publishers‟ policies with respect to self-archiving pre-print and post-print research papers.

Entry costs

Entry overheads refer to the resources required to develop and/or configure text mining tools to be used within a explicit context. There are some common tools available that need little configuration; however, higher end tools generally necessitate adaption and significant training before they can be used in a diverse area[9].

Staff costs

Text mining is presently a very specialized activity, requiring noteworthy technological and analytical skills as well as domain expertise. For example, the Big Data report [3] forecasts a shortage of 140,000–190,000 people in the USA with the necessary profound analytical skills to develop and support data and text mining.

Infrastructure costs

Text mining over large collections requires noteworthy computational and storage resources. For example, as discussed in 2.1, copies all the papers need to be made and annotated, and huge data repositories build.

9. SIMILARITY SEARCH

Bioinformatics is a growing young field that applies computer technology in molecular biology and develops algorithms, to manage and analyze biological data. DNA and protein sequences are essential biological data and exist in large volumes as well. It is key to develop effective methods to compare and line up biological sequences and determine bio sequence patterns.

Measuring the similarity[35][36] among words, sentences, paragraphs and documents is an important component in various jobs such as information retrieval, word-sense disambiguation, automatic essay scoring, document clustering, short answer grading, machine translation and text summarization. This survey discusses the existing works on text similarity. They are partitioned into three approaches; Corpus-based, String-based, and Knowledge based similarities.

In addition to samples of combination between similarities are presented.

- String-Based Similarity
- Corpus-Based Similarity
- Term-based Similarity Measures
- Hybrid Similarity Measures
- Knowledge-Based Similarity

String similarity measures operate on string sequences and character composition. . This survey represents the well known string similarity measures which were used in SimMetrics package [1]. As shown in figure 1, 14 algorithms will be introduced briefly; seven of them are character based and the other are term-based distance measures.

10. BIOMEDICAL CHALLEGES AND ISSUES [3,12,19,24,37,40]

As biomedical research and healthcare continue to develop in the genomic/post genomic era a number of challenges and opportunities exist in the wide area of biomedical informatics. Here . we define bioinformatics as the field that focuses on information, data, and knowledge in the context of biological and biomedical research. We then provide informatics frameworks for organizing and thinking about challenges and opportunities in the field of bioinformatics.

Biomedical Information Access and Retrieval: As the volume of data and knowledge increases, it is becoming crucial to biologists that they must be able to access the relevant pieces when they need it. Related to information

across varied sources is the fact that interdisciplinary and inter professional research is becoming important, Thus research findings are published in a huge range of journals.

Biomedical Decision Making: Apart from the field of bioinformatics, small explicit research into the area of decision making is done. Within clinical or medical informatics there is a rich history of research into systems designed to help care providers and patients (healthcare consumers) make better decisions surrounding diagnosis (what disease or illness is it that a patient has) and management (which of the options for treatment are best factoring in details of the situations and the values of the patient).

Evaluation and Socio Technical Dimensions of Biomedical Systems: The bioinformatics literature has a huge number of papers published on bioinformatics systems and a large number of papers on specific applications.

Biological Perspective: The rapid growth in biological data and the introduction of that raw information into extremely integrated databases on the Internet.

Data Storage, Interoperability, Standardization, and Retrieval: The huge growth in biological information acquired at every level of the Bio organization, from the simple DNA sequences to the global ecosystem, created serious challenges in data storage, retrieval and display. These challenges are met by new developments in nanotechnology, search algorithms, and virtual reality tools.

Data Publication and Knowledge Sharing: NIH requires all data generated by research it funds to be published easily. It should be accessible and sharable in electronic format. Creating challenges for current approaches such as journals and websites. The meaning of "publication" started to evolve, and libraries in particular are becoming involved in providing for the distribution of raw data from scientific experiments (see DSpace: http://www.dspace.org/). Additionally, increased use of "tele presence" tools such as the Access Grid (http://www.accessgrid.org/) and online collaboration/knowledge sharing tools such as Ask Me (http://www.askmecorp.com/) presents new and innovative infrastructure in support of the research effort.

Analysis/annotation Tool Development and Distribution/access: The development of Open Source bioinformatics tools among many different departments/groups at Universities and other institutions presented the need to develop the means of making these "home brew" tools made available to the general bioresearch community.

Networking and Communications Tools: The highly varied nature of the modern research in biology enterprise has from its beginning needed a very high degree of networking and communications among many individual researchers and organizations—the Human Genome Project itself would have been impossible without the use of Internet to develop and facilitate the distributed method to sequencing and annotating the human genome.

Publication/comprehension of Biological Information: Novel means of publication of data—wikis with their strength for fast and constant review, data posting on websites such as the Gene Expression Omnibus (GEO: http://www.ncbi.nlm.nih.gov/geo/), efforts such as the e-cell Project (http://www.e-cell.org/) and virtual disease models such as the Entelos Diabetes virtual patients (http://www.entelos.com/) to help understand biological systems—are becoming essential to use digital biological information for clinicians and basic biology researchers.

The Bio Research Program at the University of Washington: The University of Washington Health Sciences Library Bio Researcher Liaison

11. CONCLUSION

Research on biomedical technology fields is expanding day by day. In this scenario the literature on biomedical and life sciences is becoming huge data. As this number of citations existing on biomedical literature, it is very important to access desired information by the scientists, clinicians and database curators from data bases. Demand for new inventions in biomedical field leading to publish so many articles by biomedical researchers and bibiliometricians. This paper presents different data sources existing on biomedical literature. Different search tools used to access the biomedical citations. Full text archives and citation data bases are existed. Also challenges in accessing the literature, Methods used in search and query performance improvement for reducing the response time of query processing. This survey is done on various existing literature survey on biomedical data accessing. And Presented comparison of study through this literature survey which can be used for our work.

REFERENCES

1. PubMed. US National Library of Medicine, National Institutes of Health;http://www.ncbi.nlm.nih.gov/pubmed.
2. Google Scholar. Google; http://scholar.google.com/
3. PubMed Central. US National Library of Medicine, National Institutes of Health; http:// www.ncbi.nlm.nih.gov/pmc/
4. http://www.unesco.org/ci/goap.
5. http://unesdoc.unesco.org/images/0021/002158/215863e.pdf.
6. http://orbit.dtu.dk/ws/files/52941673/Open_Access.pdf.
7. http://ec.europa.eu/research/science-society/pdf/scientificpublication-study_en.pdf
6. www.scienceeurope.org.
7. www.fwf.ac.at/de/public_relations/oai/pubmed.pdf.
8. Open access' means 'free to read and free to re-use', at least for noncommercial purposes, although all the content of UKPMC is free to read. http://ukpmc.blogspot.co.uk/2012/05/increasing-proportion-of-ukpmcarticles.html.
9. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3544328/.
10. http://cdn.intechopen.com/pdfs-wm/5918.pdf.
11. http://thomsonreuters.com/content/dam/openweb/documents/pdf/pharma-life-sciences/fact-sheet/tr-integrity-cfs-en.pdf.
12. https://en.wikipedia.org/wiki/Elsevier.
13. https://www.elsevier.com/solutions/embase-biomedical-research.
14. https://en.wikipedia.org/wiki/Embase.
15. https://www.elsevier.com/solutions/scopus.
16. On-Line Biomedical Databases–The Best Source For Quick Search Of The Scientific Information In The BiomedicineActa Inform Med. 2012 Jun; 20(2): 72–84. doi: 10.5455/aim.2012.20.72-84.

17. Accessing Biomedical Literature in the Current Information Landscape *Methods Mol Biol*. 2014 ; 1159: 11–31. doi:10.1007/978-1-4939-0709-0_2.
18. http://www.ncbi.nlm.nih.gov/pmc/pub/pubinfo/.
19. http://www.sabiosciences.com/SciGlobeInfo.php.
20. Scopus - Document Search. Elsevier;http://www.scopus.com/home.url.
21. http://www.biomedicalsearchengines.com/.
22. http://www.ncbi.nlm.nih.gov/pubmed/24788259.
23. http://link.springer.com/protocol/10.1007/978-1-4939-0709-0_2.
24. http://www.researchgate.net/publication/262024298_Accessing_Biomedical_Literature_in_the_Current_Information_Landscape.
25. http://www.bmlsearch.com/
26. http://ipubmed.ics.uci.edu/#/?q=ipubmed.
27. http://dm.postech.ac.kr/refmed/?keyword=maid&display=2&show=3&sortby=2&yearFrom=&yearTo=&feedback=3&page=1&mode=1&feedbackQueryId=0&searchId=#
28. https://pmi.nlm.nih.gov/slim/index.php.
29. http://pubcrawler.gen.tcd.ie/.
30. http://xplormed.ogic.ca/example/index.html.
31. http://www.ncbi.nlm.nih.gov/pubmed/17014720.
32. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2387223/.
33. http://www.ncbi.nlm.nih.gov/pubmed/18284683.
34. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1933238/.
35. Y. Yamamoto, T. Takagi / Journal of Biomedical Informatics 40 (2007) 114–130