A comprehensive review of database resources in chemistry

Syed Sauban Ghani[1+]

1. Jubail Industrial College, General Studies Department, Jubail, Saudi Arabia

+Corresponding author: Syed Sauban Ghani, Phone: +96658392-0235, Email address: Syed_SG@jic.edu.sa

ABSTRACT: As scientific community worldwide is publishing a huge number of research articles in various fields; it is necessary to distinguish between databases that are efficient and objective for literature searches. This review offers information on the important points of the database. None of the databases are complete and perfect, but they complement each other. If a library can only afford one, choice must be based on the priorities of institutional needs. The benefits that databases can provide in the preparation of the literature review for developing future studies and dissemination of research are discussed. This paper provides an overview of the most frequently used free chemistry databases such as PubChem, Crystallography Open Database, PubMed, ZINC, ChemSpider, and Google Scholar. It also gives a brief description of three major commercial databases such as Scopus, Web of Science, and SciFinder. Thus, substance and citation databases that covers almost all areas of chemistry, has become an invaluable tool in bibliometric analysis.
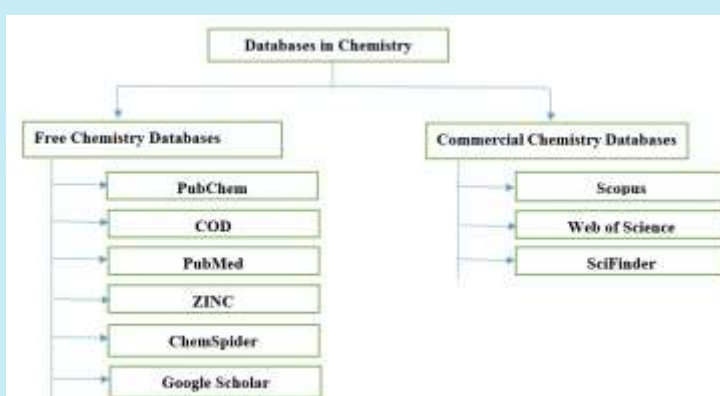
1. Introduction

The amount of information available today is growing at an exponential rate and the ability to search for the necessary information is one of the basic needs of knowledge. The abundance of technological and Internet resources can both simplify and complicate a researcher's world. Chemistry is an interdisciplinary subject upon which the other scientific disciplines are dependent to a certain extent. This vast information data must be systematically organized by the experts in the field. A database is an organized collection of data in any field[1]. In addition to the basic search techniques that are used today in almost everyday life, such as searching by keyword search engines, there are some areas of chemistry that are not so simple or meaningful. Chemical databases have now become a powerful tool in drug discovery. Database searches based on potential requirements for biological activity identifies compounds that are suitable for detailed analysis or indicate novel ways to achieve the desired activity[2]. Accessing chemical information that are stored in different kinds of databases by utilizing the means of computer are gradually becoming more significant. The dimension of all databases either in terms of the structures or that of the reactions are growing tremendously each year[3]. The proficiency of search algorithms executed within these databases are very crucial. Therefore, the computer supported databases are becoming useful tools for several research laboratories in industry as well as in academia.

It is must to have a better understanding of how the data is organized and interconnected for the effective search in any database. The databases are broadly distributed in two major group's viz. full-text and the structures, based on the category of data contained[4]. A full-text database is generally a set of documents in which indexes are created to facilitate their fast search. This type of database is commonly run by publishers of magazines and books, patent offices for patents or academic institutions. The largest database of this type on this globe is definitely operated by Google, in which documents and other accessible files are uploaded on the internet in the form of websites[5]. On the other hand, structured databases normally include a set of tables that contain records or rows, all of which have the same structure well defined by a set of fields or columns[6]. Each record is always assigned a unique "ID" or "Number" called as identifier or the primary keys, which are easily referenced. The Chemical Abstracts Service Registry Number (CAS RN) is an example of the primary key for the structure in the REGISTRY database[7]. Structured databases are generally classified into two large groups in terms of its contents as bibliographic and factographic. The bibliographic databases usually do not have the full text of a document, however it records information about a single publication, patents, and similar documents[8]. Typical fields in bibliographic records are - author, article title, journal name, volume, issue, year of publishing, pages, etc. The Digital Object Identifier (DOI) is a comparatively new parameter, which describes the distinctive placement of a document on the Web[9]. The bibliographic databases are secondary sources that interpret, analyze, and summarize, the primary source information to increase usability and speed of delivery, such as an online encyclopedia. Moreover, factographic databases consist of specific information extracted from primary documents, particularly in the area of chemistry that have details about chemical reactions and chemical compounds such as toxicological, spectral, physical, or chemical characterization[10]. It is must for the databases to allow the user to search for records by all field values as well as to create search queries for logical operators in order to be regarded as an ideal database system. Additionally, the thorough knowledge of the database structure, the syntax of the search language, and specific IT skills are the prerequisite for this search method. The method that is frequently adopted by the database creators is the inclusion of the "forms" which is already having names of the usual fields such as "bibliographic data" or "physical parameters" to give a user-friendly interface.

Search by chemical identifier is probably the simplest search by keyword, with the difference that chemical identifiers are a little more difficult to define. However, the free programs available to us can generate these identifiers if we are able to enter the structure of the compound (e.g. Chem-Sketch or Marvin- Sketch)[11]. The most common search for information on chemical compounds or their chemical products is the search for structures or substructures. Less common were also searched in structured databases under chemical, physical, or biological properties of the chemical compounds. There are many considerations that are involved in the construction and searching of chemical databases. Chemical structures that are commonly stored in databases, such as text, differ significantly from other entities therefore the different search modes too differ significantly, however some matches can be drawn. The reason for the existence of different databases is that each of them have its own function, however, none of them is perfectly a subset of any other. The subsequent process for any chemical databases using a specialized structural editor is to create a search query, give a chemical structure or substructure of a search compound. JME Editor was the most widely used structural editor of this kind but for the last couple of years or more have started to phase out this technology[12]. Consequently, the creators of chemical databases opt for JavaScript-based editors that is the recognizable technology of the future. So, the best among the structural editors nowadays is Marvin JS, which is widely used in the application of Reaxys. The most exciting and commercially available program for drawing chemical formulas is Chem Draw, which is marketed by Cambridge Soft. The most recent version of this editor permits the user to search for the diagrams directly in SciFinder[13].

The commercial chemical databases of the resource are the most popular and most widely used web applications of Scopus, Reaxys, SciFinder and Web of Science (WoS), in which the above search technologies are possible and are more closely related to this article. The chemical databases are nowadays searched to give novel

ideas for prime discovery. The comparison of the search possibilities of Reaxys, SciFinder and Web of Science sources were published in 2016 [14], and a comparison of two more chemically oriented sources - Reaxys and SciFinder was published by Jaroslav Silhanek[15]. In this paper, we will first describe the types of databases used in chemistry and the possibilities of the most important commercial tools. The authors chose some of the databases as the object of study on the assumption that these databases might provide the most informative and relevant results for a specific query. As the main source of the selected database for retrieving results from published journals, books, patents, conference abstracts, and other available relevant resources. After a quick description of several existing databases, we will also provide an overview of alternatives to freely available chemical resources, which in some cases, may replace the commercial resources. And finally, we will conclude by highlighting the weaknesses and shortcomings of the database as well as recommend the ways for their best possible utilization. The research criteria adopted were based on qualitative and quantitative characteristics of the database such as source, citations, searching and special features by analysing previous studies.

2. Experimental

The open Web compromises a rich pool of various chemical data sources if the user knows where to find out. It has been over many years since some emerging chemical databases were dominated by a handful of established players, the field has practically opened up to a variety of innovative newcomers. Although some of the original databases are no longer active, it is inspiring to see that several them continue to run and even flourish. It is of course more likely that still many more services will be created and some of them will become irrelevant in the coming years. The Internet now offers a varied range of free online chemistry databases, and this list is being continuously updated with new information and new entries. The following list summarizes some of the databases that are freely available for the users.

2.1. PubChem

PubChem is a public repository for information on chemical substances and their biological activities. It is regarded as the grandfather of all free chemistry databases, which search over 8 million compounds by a variety of criteria and is systematized as three linked databases viz. PubChem Substance, PubChem Compound, and PubChem Bioassay[16]. PubChem is a database of chemical molecules and their activities against biological assays. The National Center for Biotechnology Information (NCBI) maintains its system. PubChem can be freely accessed through a web user interface where millions of compound structures and descriptive datasets are freely downloaded via FTP. PubChem contains the descriptions of substance and small molecules having lesser than 1000 atoms and 1000 bonds. More than 350 database retailers add to the developing PubChem database. PubChem have a significant amount of literature-derived bioactivity data of chemical substances which are manually extracted from several thousands of scientific articles by data contributors such as ChEMBL and BindingDB and additionally, through integration of data from Drug Bank, the Hazardous Substances Data Bank and other databases[17]. The databases of these databases complement to the contents of PubChem.

2.2. Crystallography Open Database (COD)

Crystallography Open Database (COD) is an open-access collection of crystal structures of organic, inorganic, metal-organics compounds, and minerals, excluding biopolymers. This database is specifically designed to store information about the structure of molecules and crystals[18]. All data on this site have been placed in the public domain by the contributors. The COD can provide a link to CIF if there is a CIF available somewhere in the internet. The Crystallography Open Database has more than 360,000 entries and has various contributors, as well as contains CIFs as prescribed by the International Union of Crystallography[19].

COD has a website http://www.crystallography.net which provides proficiencies for all registered users to deposit published or unpublished crystallographic structures as personal communications or pre-publication depositions. Having such sort of a

setup that enables extension of the COD database by several users simultaneously. It also increases the chances for growth of the COD database and may be considered as one-step towards creating a worldwide Internet-based collaborative platform committed to the collection and curation of structural knowledge. Each structure deposited into the COD generate a unique seven-digit number, called COD number, which identifies a particular illustration of a structure determination. In general, COD does not accept duplicate structures.

2.3. PubMed

PubMed is a freely accessible web interface (since 1997) designed to search for records located primarily in the MEDLINE database of references and abstracts. It comprises more than 28 million citations for biomedical literature from MEDLINE, life science journals, and online books[20]. Citations may include links to full-text content from PubMed Central and publisher web sites. PubMed also provides access to older references even from the print form of Index Medicus dating back to 1951 or earlier in addition to MEDLINE. This bibliographic database is indexed by journal entries and other primary sources related to medicine. There is also information about publications in the field of medicinal chemistry or biochemistry. The PubMed identifier (PMID) is the primary key used in PubMed to identify the unknown in this database. The tool provided in PubMed facilitates saving searches, filtering search results saving sets of references retrieved as part of a PubMed search, configuring display formats of search terms and the extensive range of further options. PubMed records with recent increases in activity.

2.4. ZINC

ZINC is a commercially available free database of compounds for virtual screening and this database has brought virtual screening libraries to a comprehensive range of structural biologists and medicinal chemists. It contains more than 35 million available compounds in ready-to-dock, 3D formats[21]. Due to its structure-based virtual screening, it has numerous significant successes in recent years and is nowadays a common technique in initial stage of drug discovery in many of the pharmaceutical companies. ZINC can be easily used for download using the website http://zinc.docking.org. It is currently built from the catalogues of ten major compound vendors in several common file formats including SMILES, mol2, 3D SDF, and DOCK flexi base format and the number of molecules in ZINC is continuously growing. This database has been designed in such a way that it organizes data relationally so that it remains compatible to attain the objectives of efficient loading, incremental updates, querying, and data subsetting. These steps make them fast and efficient. Though exporting subsets of the database can make them slow, but this problem has been resolved by exporting the molecule subsets from the database into ready-to-download compressed files, and database-intensive work is scheduled in batch mode. This totally bypasses the relational database and subsets are downloaded speedily once it is ready. The web-based interface is fast as well as supports moderately complex queries and users may search ZINC based on several criteria. The ZINC server enables users to upload and process their own molecules, as we often come across molecules such as positive and negative controls that we need to dock that are not part of the existing database[22]. Henceforth, ZINC is much useful for virtual screening by experts and non-specialists equally and assist more researchers to attempt computational ligand discovery.

2.5. ChemSpider

ChemSpider is an open access chemical structure database, which provide rapid text and structure search access to over 67 million structures from hundreds of data sources. ChemSpider is one of the chemistry community's primary online public compound databases. ChemSpider serves data for tens of thousands of chemists every day and it lays the foundation for many important international projects to integrate chemistry and biological data, facilitate drug discovery efforts and help to identify new chemicals from under the ocean[23]. It is not just a search engine based on terabytes of chemistry data but also acts as a crowdsourcing community for chemists those have contributed their data, skills, and knowledge for the enhancement and curation of the database. Therefore, it can be said that ChemSpider seems like Wikipedia by promising participation and contributions from the

scientific community. ChemSpider can link open- and closed-access chemistry journals, environmental data, PubChem, Chemical Entities of Biological Interest (ChEBI), chemical vendors, Wikipedia, The Kyoto Encyclopedia of Genes and Genomes (KEGG), and few other patent databases[24]. These links allow a ChemSpider user to collect information of their interest, such as from where to buy a chemical, chemical toxicity, metabolism data, and so on. Amassing this level of related information through a usual search engine like Google or Bing is a time-consuming process. Additional features have been added to each of the chemical structures within the database, such as structure identifiers like SMILES, InChI, IUPAC, and Index Names, as well as many physico-chemical properties[25]. ChemSpider also offers access to a series of property prediction algorithms. The user can access this database by browsing http://www.chemspider.com/. The ChemSpider homepage as it appears on the desktop has been shown in Fig. 1. The provider of this service has been the Royal Society of Chemistry (RSC) since 2009, which gives more value to other positive and useful services.

For the known compound, it provides extensive information like all its possible names and identifiers (both standard and nonstandard), experimental and calculated physicochemical properties, toxicity and biological activity data, spectra (NMR, IC, MS, UV-vis), publications, patents, etc. The information that is available depends on what has been gained from the original sources and the links to it are available. The role of ChemSpider is to get information about all the compounds available on the web at one central location, make it easy to search and standardize their structures and names. It also improves the quality of chemical sources by using automated control of the structure and manual management of collaborating experts as well as provides a platform for data input and storage. Additionally, it tries to make it easy to access all data using a web interface optimized for mobile devices, mobile applications, and web services for data capture. It does integrate data into the RSC publication using the first links and use validated chemical names to search in Google Scholar, PubMed and RSC books, journals, and databases.
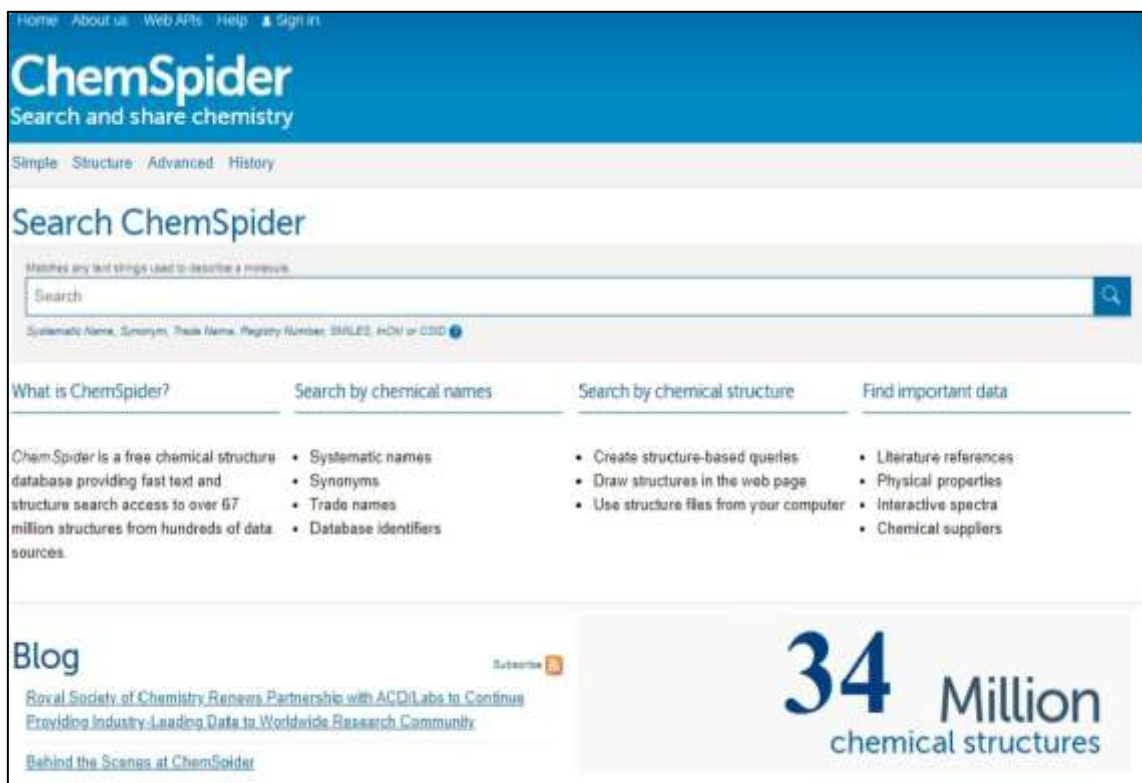

Figure 1. ChemSpider homepage as on desktop.

2.6. Google Scholar

Google Scholar is a freely available web-based search engine, available since 2004, that indexes the full text or metadata of scholarly literature through a range of publishing formats and disciplines. The indexed resources include online journals, conference papers, books, dissertations, thesis, patents, and any other significant literature. It is estimated that it contains approximately 389 million documents comprising articles, citations and patents which makes it the world's largest academic search engine in 2018 [26]. Google Scholar has now become indispensable for research and research dissemination that provides a systematized and instant process for users to build on through a sort of digital snowball for literature retrieval. The reason for the excel of Google can be attributed to its sophisticated natural language processing. In addition to the search, Google Scholar users are also able to create a personal profile with a list of their own publications and can generate census statistics and H-indexes like that of Web of Science. Nevertheless, if a user wishes to use a structured query in accordance to the field values in the bibliographic record or to find documents that have not been issued, it is preferable to resort to paid databases. The difficulties faced by the Google Scholar users are that they are not aware that when it is updated, includes old articles, as well as no suggestions are provided for limiting searches.

3. Results and Discussion

The three major commercial web database applications that are widely used in the field of chemistry are Scopus, Web of Science (WoS) and SciFinder. All the three databases contain extensive search options and are somehow remarkably similar in their chemical content as well as in their search mode, search effectiveness and interface. They have periodically undergone significant overhauling

and have intense competition between them. This competition has led to improvements in the services offered by them, which is, however, advantageous for users. As these databases are expensive, it is not feasible to have all these databases, therefore the scientific libraries must decide that which citation database will meet the requests of the consumers more effectively.

Despite of the similarities between them still there are differences between them that are worth of a detailed analysis.

3.1. Scopus

Scopus is the Elsevier's largest abstract and citation database of peer-reviewed literature, which was launched in 2004. The literature covers more than 49 million records including scientific journals, conference proceedings, and books[27]. Scopus offers a complete summary of the global research output in the area of science, engineering, medicine, humanities, and social sciences. Scopus database is the leading searchable citation and abstract source for searching literature that is continuously expanding and updating. Scopus offers smart tools that have the sorting and refining features to track, analyze and visualize research of more than 27 million citations and abstracts dating back to 1960s[28]. Researchers across the globe believe that use of Scopus had positive influence on the research finding as it is easy to use, saves time as well as provides quality outcome. The content on Scopus is derived from over 5,000 publishers, which is reviewed by an independent Content Selection and Advisory Board (CSAB) and then selected for indexing in Scopus. The metadata that is provided by publishers includes the following: authors name, affiliations, document title, volume, issue, pages, year, electronic identification (EID), source title, citation count, document type and digital object identifier (DOI). This metadata is integrated to different websites and platforms, which provides more precise search and enables retrieval of scientific information. Scopus provides International Standard Serial Number (ISSN) for journals, conference series or book series for series publication and International Standard Book Number (ISBN) for one-time conference or book publication[29]. The overall view of the working pattern of Scopus is given in Fig. 2.

However, the coverage of a journal by Scopus may be discontinued for a certain period i.e. it has breaks for some journals whereas for some journals Scopus makes a partial coverage. Several studies can be found in the literature making detailed descriptions of the main features of Scopus and comparing the databases with the aim of assessing the number of citations obtained by a particular set of documents in each of them.

Studies have analyzed the set of journals covered by each database as well as their interface accessibility and usability compared with Scopus, from the point of view of the number of items included, and of testing the breadth of coverage. The rankings from Scopus and WoS match at the top and the bottom but deviate considerably in the middle positions[30]. If the user is aware with search devices such as drop-down boxes and check boxes, even for the beginner Scopus is easy to

navigate. Scopus have some extraordinary features such as: it allows the user to go both forwards and backwards in time by linking to both citing and cited documents; it can link to the publisher's web site to view the document; citation accuracy is so accurate that 99% of citing references and citing articles matched exactly; can work in all the common web browsers like Chrome, Internet explorer or Mozilla.
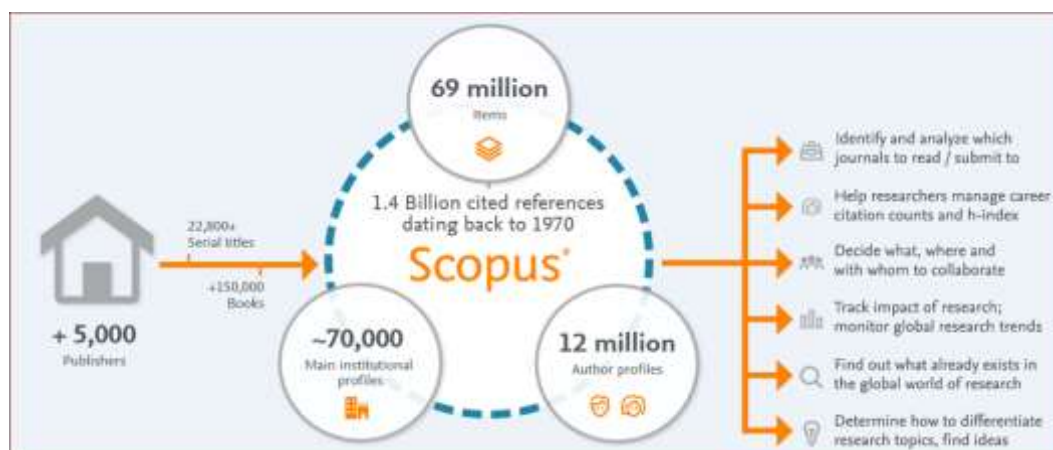

Figure 2. Working pattern of Scopus.

3.2. Web of Science

Web of Science is an ideal place to search the citation universe across subjects and across the world as it provides an access to the most reliable, integrated, multidisciplinary research that is connected through linked content citation metrics from multiple sources within a single interface[31]. As Web of Science adheres to a strict evaluation process, it assures that only the most influential, relevant, and credible information is included. The selection is made based on impact evaluations i.e. Impact factor (IF) that is a measure reflecting the yearly average number of citations to recent articles published in that journal and it includes open-access journals. Therefore, it allows the user to uncover the subsequent vast idea quicker. WoS connects the complete search as well as discover the process through Multidisciplinary Content; Subject Specific Content; Emerging Trends; Analysis Tools and Research Data. WoS precisely indexes the utmost significant literature in the world and has become the standard for research discovery and analytics. WoS links publications and researchers through citations and organized indexing in curated databases across every discipline. It uses cited reference search to track

past research and monitor current developments for around 100 years of indexed content including 59 million records as early as 1898 [32]. WoS were originally created by the Institute for Scientific Information (ISI) and now is maintained by Clarivate Analytics. WoS enables the user to acquire, analyze, and disseminate database information in a timely manner and is possible due to the creation of a common vocabulary, called ontology, for varied search terms and varied data. Furthermore, these search terms generated relate information across categories. WoS platform provides access where the user can search individually or through a combination of topic, title, author or author ID, editor, conference, language, journal title, digital object identifier (DOI), year published, organization, address, document type, funding agency, grant number, accession number, and PubMed ID[33]. The Web of Science Core Collection, as illustrated in Fig. 3, consists of the following six online databases: Science Citation Index Expanded; Social Sciences Citation Index; Arts & Humanities Citation Index; Emerging Sources Citation Index; Book Citation Index and Conference Proceedings Citation Index. Apart from the seven citation indices listed, additionally two chemistry databases, Index

Chemicus and Current Chemical Reactions permit the creation of structure drawings, consequently allowing users to locate chemical compounds and reactions.

Key features of Data Citation Index (DCI) on WoS is to search directly through millions of records from hundreds of evaluated data repositories in the Sciences, Social Sciences, and Humanities[34]. Each DCI record links directly to the repository so that users can quickly access the associated research data. Citations to data sets are indexed so that the user can measure their impact as well as track their influence. The Data Citation Index offers a single point of access to research data from repositories across disciplines throughout the world. In this index, descriptive records are generated for data objects and linked to literature articles in the Web of Science. As data citation practices increase, the resource aims

to provide a distinct picture of the full impact of research output, and act as an important tool for data attribution and detection. WoS has the most advanced features for citation analysis. It allows H-index to be honored that is now extensively used to assess the quality of the scholar and is able to search for multi-field quotes with respect to other databases. The newer WoS features provide search by the grant agency or the grant number. It is also possible to export the found export logs in various formats to a file or a web-based version of EndNote's personal bibliographic database. For subsequent import into another bibliographic database, it is appropriate to use the RIS structured export scheme, which is a recognized standard for these purposes. Controlling WoS is the easiest to learn with the help of a video tutorial, which are available to foreign operators.
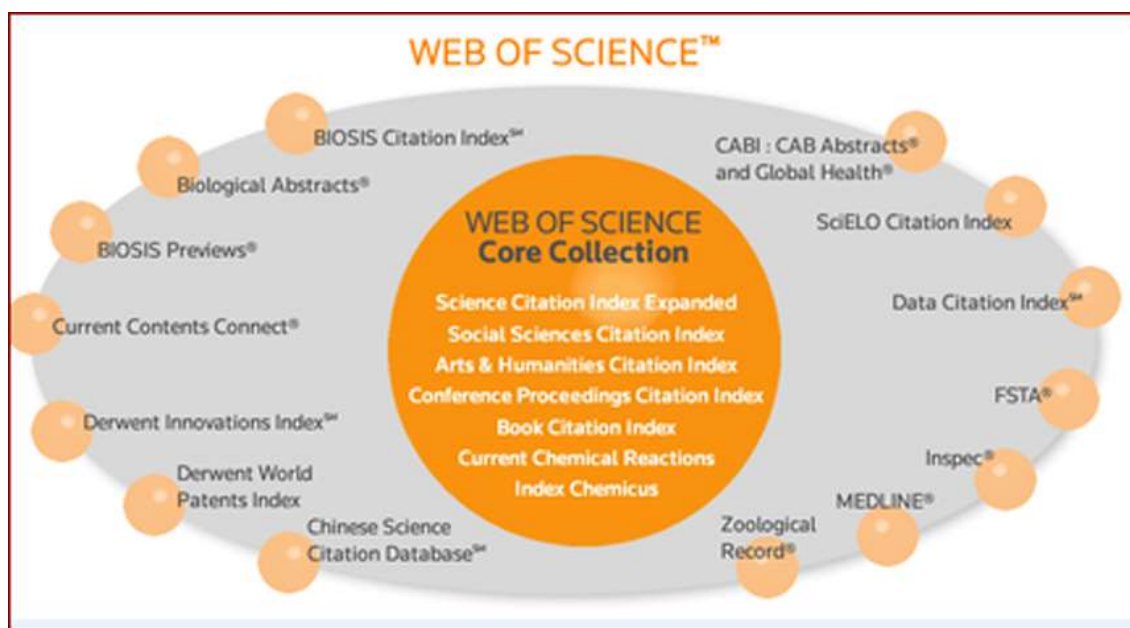

Figure 3. The Web of Science Core Collection.

3.3. SciFinder

SciFinder was launched by Chemical Abstract Service (CAS) in 1995 as a desktop application tool for Medals of Chemical Literature[35]. Today's application provides access to some databases produced by CAS, as well as to the freely available MEDLINE bibliographic database. SciFinder is a sophisticated search interface to six basic chemical related databases. CAS itself produces five of these databases. SciFinder Scholar is designed so infrequent searchers can

explore the chemical literature, thereby eliminating the need to learn the intricacies of searching CAS. SciFinder offers easy, convenient, and prompt access to CAS REGISTRY, the standard for substance information, proposing more substances than any other single-source tool including organic and inorganic molecules, DNA, RNA, proteins, polymers and Markush structures[36]. On daily basis CAS scientists gather and investigate published scientific literature across the globe, creating the best quality and most up-to-date collection of scientific

information in the world. Covering progresses in chemistry and linked sciences for the last 150 years, the CAS content pool empowers researchers, and information professionals with instant access to the trustworthy information required to catalyze innovation. SciFinder offer a direct search through the below mentioned database:

3.3.1. CAplus

The main Chemical Abstracts literature database of over 23 million references. It is a bibliographic database, which contains data from the most important chemistry journals for the CAS Source Index (CASSI), is available free of charge, where the user can search by CODEN, ISBN, ISSN, and naming or abbreviations for all sources used by CAS since 1907 [37]. The bibliographic records from this database are displayed in the SciFinder interface for relevant references to other CAS databases, and in most cases, references to the full texts of the document are also found.

3.3.2. CAS REGISTRY

It is a substance database containing information about all the compounds that CAS has ever been abstracted from the literature. It has a pool of more than 27 million organic and inorganic substances and 57 million bio sequences. Any previously unrecorded compound that is added to the database is assigned the new CAS Registry Number (CAS RN), which is a very wide-ranging identifier of chemical compounds, often used in chemical vendor catalogues[38]. For the most common chemicals, it is possible to search CAS RN by name of the compound or to find the name in a freely accessible web application operated by CAS - Common Chemistry[39]. If the user applies some forms and a structural editor from the SUBSTANCES section of the SciFinder interface, the results from the database will be displayed, and in each record, the user will be able to find the relevant links to the other CAS databases, including the CAS REGISTERS. Every day about 15,000 new compounds are added to the database. CAS REGISTRY serves as a universal standard for chemists worldwide.

3.3.3. CASREACT

It is a chemical information database that provides access to over 10 million reactions from the journal literature and patents. Most of the reactions are from the publications dating back to 1985, but the records published in 1840 can also be found[40]. If the SciFinder interface is used in the REACTIONS section, which allows search only according to data entered in the structural editor, and the records from this database is shown. At present, the CASREACT database has more than 83 million records, and each day adds about 30,000 new responses.

3.3.4. CHEMCATS

It is a Supplier Chemicals Database, which lists suppliers of commercially available chemicals worldwide. For each commercially available compound, a link to its vendor that leads to this database is available[41]. However, only those suppliers who joins CAS CHEMCATS program are able to find it. Nevertheless, it is often possible to find a trusted supplier with a distinctly more favorable price than the usual suppliers. Naturally, the current price of a chemical is often quite different from what is reported in SciFinder.

3.3.5. CHEMLIST

It is a regulatory chemicals database. It includes chemicals that appear on a list of regulated chemicals (toxic, hazardous, etc.). If a compound is found in the CAS REGISTRY database, it is also contained in the CHEMLIST database. The appropriate links can be found in the compound record in the REGULATORY INFORMATION section[42]. Currently, this database contains more than 348,000 entries, and about 50 new substances are added each week that are accumulated from the extensive group of national and international regulatory lists and inventories.

4. Conclusions

The dimension of almost all chemical databases has increased manifold in the last many years so the search engines must be equally more powerful. The outline of this research is the usefulness of the databases for teaching, learning,

and research as each chemical record retains the links to the original source of the material, thereby associating a micro attribution and these links let a database user source information of particular interest. Each of the commonly used chemical databases presented here has at least some overlap with each of the remaining ones, which means that each of these databases appears to have its own "niche". The user, looking for a variety would like to give attention to each of them. These databases thus seem to be a valuable resource to the chemical community as they offer a large collection of compounds, either with related sample availability or with a diverse and unique structure set. As all the investigated databases developed over the years, the detailed results of these databases essentially signify a snapshot in time. The description reported here may give a useful overview relative to some of the most important large chemical databases available. In PubChem, unique chemical structures are extracted from the Substance database and stored in the Compound database that provides an accumulated interpretation of information for a given chemical structure. COD database establishes a worldwide Internet-based collaborative platform committed to the collection and curation of structural knowledge. PubMed provided a general description of PubMed including its content and unique characteristics. ChemSpider provides the variety of information of a given compound including physical and chemical properties, molecular structure, synthetic methods, spectral data, and systematic nomenclature for millions of compounds in a single Web site. The ZINC database provides 3D molecules in several formats compatible with most docking programs. Google Scholar helps to identify the collection of publications for a specific research topic. There is a high association between WoS and Scopus databases that allows searching and sorting the queries by anticipated parameters such as first author, citation, and institution, etc. regarding impact factor and h-index. SciFinder meets its goal of effectively exploring the scientific literature and the search results are mostly truly relevant and often astonishingly inclusive regardless of the level of complexity or syntax of the query. The database that ought to be used depends on the user and desired information. Therefore, the user must investigate the up-to-date condition of the specific database before establishing a decision of acquisition and usage of any of the databases presented here, as there may be changes in scope, configuration, vendor, etc. Libraries willing to subscribe the database should make their choice based on the needs of the library.

5. Acknowledgment

6. References

[1] Masic, I., Review of most important biomedical databases for searching of biomedical scientific literature, Donald School Journal of Ultrasound in Obstetrics and Gynecology 6 (4) (2012) 343-361. https://doi.org/10.5005/jp-journals-10009-1258.

[2] Walters, W. P., Stahl, M. T., Murcko, M. A., Virtual screening—an overview, Drug Discovery Today 3 (4) (1998) 160-178. https://doi.org/10.1016/S1359-6446(97)01163-X.

[3] Hunter, L., Cohen, K. B., Biomedical language processing: what's beyond PubMed? Molecular Cell 21 (5) (2006) 589-594. https://doi.org/10.1016/j.molcel.2006.02.012.

[4] Tenopir, C., Ro, J. S., Full Text Databases, Greenwood Press, Westport, 1990.

[5] Bar-Ilan, J., Which h-index? - A comparison of WoS, Scopus and Google Scholar, Scientometrics 74 (2008) 257-271. https://doi.org/10.1007/s11192-008-0216-y.

[6] Chang, K. C-C., He, B., Li, C., Patel, M., Zhang, Z., Structured databases on the web: Observations and implications, ACM SIGMOD Record 33 (3) (2004) 61-70. https://doi.org/10.1145/1031570.1031584.

[7] Dittmar, P. G., Stobaugh, R. E., Watson, C. E., The Chemical Abstracts Service Chemical Registry System. I. General Design, Journal of Chemical Information and Computer Sciences 16 (2) (1976) 111-121. https://doi.org/10.1021/ci60006a016.

[8] Wright, K., McDaid, C., Reporting of article retractions in bibliographic databases and online journals, Journal of the Medical Library Association 99 (2) (2011) 164-167. https://doi.org/10.3163/1536-5050.99.2.010.

[9] Paskin, N., Digital Object Identifier (DOI®) System, In: Encyclopedia of Library and Information Sciences, Bates, M. J., Maack, M. N., ed., CRC Press: Boca Raton, 3rd ed., 2010, Ch. 7. https://doi.org/10.1081/E-ELIS3-120044418.

[10] Conklin, D., Fortier, S., Glasgow, J., Knowledge discovery in molecular databases, IEEE Transactions on Knowledge and Data Engineering 5 (6) (1993) 985-987. https://doi.org/10.1109/69.250082.

[11] Ertl, P., Molecular structure input on the web, Journal of Cheminformatics 2 (1) (2010) 1-9. https://doi.org/10.1186/1758-2946-2-1.

[12] Bienfait, B., Ertl, P., JSME: a free molecule editor in JavaScript, Journal of Cheminformatics 5 (24) (2013) 1-6. https://doi.org/10.1186/1758-2946-5-24.

[13] Mendelsohn, L. D., ChemDraw 8 Ultra, Windows and Macintosh Versions, Journal of Chemical Information and Computer Sciences 44 (6) (2004) 2225-2226. https://doi.org/10.1021/ci040123t.

[14] Bharti, N., Leonard, M., Singh, S., Review and Comparison of the Search Effectiveness and User Interface of Three Major Online Chemical Databases, Journal of Chemical Education 93 (5) (2016) 852-863. https://doi.org/10.1021/acs.jchemed.5b00601.

[15] Šilhánek, J., Comparisons of the most important chemistry databases - Scifinder program and reaxys database system, Chemicke Listy 108 (1) (2014) 81-106. https://projekty.upce.cz/sites/default/files/groups/admins/luva3059/2014_01_83-90.pdf.

[16] Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., Bryant, S. H., PubChem: a public information system for analyzing bioactivities of small molecules, Nucleic Acids Research 37 (2 Suppl) (2009) W623-W633. https://doi.org/10.1093/nar/gkp456.

[17] Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., Bryant, S. H., PubChem Substance and Compound databases, Nucleic Acids Research 44 (D1) (2016) 1202-1213. https://doi.org/10.1093/nar/gkv951.

[18] Gražulis, S., Daškevič, A., Merkys, A., Chateigner, D., Lutterotti, L., Quirós, M., Serebryanaya, N. R., Moeck, P., Downs, R. T., Le Bail, A., Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration, Nucleic Acids Research 40 (D1) (2012) D420-D427. https://doi.org/10.1093/nar/gkr900.

[19] Hall, S. R., Allen, F. H., Brown, I. D., The crystallographic information file (CIF): a new standard archive file for crystallography, Acta Crystallographica Section A A47 (1991) 655-685. https://doi.org/10.1107/S010876739101067X.

[20] Van Buskirk, N. E., The review article in MEDLINE: ambiguity of definition and implications for online searchers, Bulletin of the Medical Library Association 72 (4) (1984) 349-352. PMCID: PMC227511

[21] Irwin, J. J., Shoichet, B. K., ZINC-a free database of commercially available compounds for virtual screening, Journal of Chemical Information and Modeling 45 (1) (2005) 177-182. https://doi.org/10.1021/ci049714+.

[22] Sterling, T., Irwin, J. J., ZINC 15 – Ligand Discovery for Everyone, Journal of Chemical Information and Modeling 55 (11) (2015) 2324-2337. https://doi.org/10.1021/acs.jcim.5b00559.

[23] Pence, H. E., Williams, A., ChemSpider: An Online Chemical Information Resource, Journal of Chemical Education 87 (11) (2010) 1123-1124. https://doi.org/10.1021/ed100697w.

[24] Hettne, K. M., Williams, A. J., van Mulligen, E. M., Kleinjans, J., Tkachenko, V., Kors, J. A., Erratum to: Automatic vs. manual curation of a multi-source chemical dictionary: the impact on text mining, Journal of Cheminformatics 2 (4) (2010) 1-7. https://doi.org/10.1186/1758-2946-2-4.

[25] Williams, A., Tkachenko, V., The Royal Society of Chemistry and the delivery of chemistry data repositories for the community, Journal of Computer-Aided Molecular Design 28 (10) (2014) 1023-1030. https://doi.org/10.1007/s10822-014-9784-5.

[26] Zientek, L. R., Werner, J. M., Campuzano, M. V., Nimon, K., The Use of Google Scholar for Research and Research Dissemination, New Horizons in Adult Education & Human Resource Development 30 (1) (2018) 39-46. https://doi.org/10.1002/nha3.20209.

[27] Burnham, J. F., Scopus database: a review, Biomedical Digital Libraries 3 (1) (2006) 1-8. https://doi.org/10.1186/1742-5581-3-1.

[28] Bar-Ilan, J., Tale of Three Databases: The Implication of Coverage Demonstrated for a Sample Query, Frontiers in Research Metrics and Analytics 3 (6) (2018) 1-9. https://doi.org/10.3389/frma.2018.00006.

[29] Ball, R., Tunger, D., Science Indicators Revisited – Science Citation Index versus SCOPUS: A Bibliometric Comparison of Both Citation Databases, Information Services & Use 26 (4) (2006) 293-301. https://doi.org/10.3233/ISU-2006-26404.

[30] Gavel, Y., Iselid, L., Web of Science and Scopus: A journal title overlap study, Online Information Review 32 (1) (2008) 8-21. https://doi.org/10.1108/14684520810865958.

[31] Jacso, P., As we may search – Comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation-enhanced databases, Current Science 89 (9) (2005) 1537-1547. https://www.jstor.org/stable/24110924?seq=1#metadata_info_tab_contents.

[32] Bar-Ilan, J., Web of Science with the Conference Proceedings Citation Indexes: the case of computer science, Scientometrics 83 (3) (2010) 809-824. https://doi.org/10.1007/s11192-009-0145-4.

[33] Meho, L. I., Rogers, Y., Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science, Journal of the American Society for Information Science and Technology 59 (11) (2008) 1711-1726. https://doi.org/10.1002/asi.20874.

[34] Robinson-Garcia, N., Jiménez-Contreras, E., Torres-Salinas, D., Analyzing data citation practices using the data citation index, Journal of the Association for Information Science and Technology 67 (12) (2016) 2964-2975. https://doi.org/10.1002/asi.23529.

[35] Somerville, A. N., SciFinder Scholar (by Chemical Abstracts Service), Journal of Chemical Education 75 (8) (1998) 959-960. https://doi.org/10.1021/ed075p959.

[36] Cain, R., Schwall, K., Guiding your literature searching, Chemtech: the innovator's magazine 25 (8) (1995) 8-11. ISSN: 0009-2703.

[37] Baykoucheva, S., Comparison of the Contributions of CAPLUS and MEDLINE to the Performance of SciFinder in Retrieving the Drug Literature, Issues in Science and Technology Librarianship 66 (2011) 1-17. https://doi.org/10.5062/F42Z13FT.

[38] Fisanick, W., Cross, K. P., Rusinko III, A., Similarity searching on CAS Registry substances. 1. Global molecular property and generic atom triangle geometric searching, Journal of Chemical Information and Computer Sciences 32 (6) (1992) 664-674. https://doi.org/10.1021/ci00010a013.

[39] Fisanick, W., Mitchell, L. D., Scott, J. A., Stouw, G. G. V., Substructure Searching of Computer-Readable Chemical Abstracts Service Ninth Collective Index Chemical Nomenclature Files, Journal of Chemical Information and Computer Sciences 15 (2) (1975) 73-84. https://doi.org/10.1021/ci60002a003.

[40] Blake, J. E., Dana, R. C., CASREACT: more than a million reactions, Journal of Chemical Information and Computer Sciences 30 (4) (1990) 394-399. https://doi.org/10.1021/ci00068a008.

[41] Cavaller, V., Software review: SciFinder, International Journal of Competitive Intelligence, Strategic, Scientific and Technology Watch SciWatch Journal 1 (1) (2008) 15-17. https://hexalog.files.wordpress.com/2008/05/2-_-scifinder-english2.pdf.

[42] Murray-Rust, P., Chemistry for everyone, Nature 451 (2008) 648-651. https://doi.org/10.1038/451648a.