Annals of Library and Information Studies
Vol. 64, March 2017, pp. 16-20

Developmental features of biomedical bibliographic databases

N. Subramanyam[a], M. Krishnamurthy[b] and A.Y. Asundi[c]

[a]Senior Librarian, M.S. Ramaiah Medical College, Bangalore - 560054,
Email: subs_saha@yahoo.co.in

[b]Associate Professor, DRTC, Indian Statistical Institute, Bangalore - 560059,
Email: mkrishna_murthy@hotmail.com

[c]Former Professor and Chairman, Department of Library and Information Science, Bangalore University, Bangalore - 560056,
Email: ashok_asundi@yahoo.com

Received: 07 April 2016; revised: 21 November 2016; accepted: 20 March 2017

The genesis of bibliographic databases can be traced to the abstracting and indexing (A&I) periodicals. The *Journal Des Scavans* published in 1665 did contain abstracts of articles, but the formal abstracting journal began in 1820 (English) and in 1830 (German). The growth of primary periodicals required libraries to subscribe to A&I services to search for papers scattered in different journals as it was not possible to acquire all the primary periodical titles. The electronic versions of A&I periodicals started appearing in 1960s with emergence of database concept. The *Chemical Titles* and the MEDLARS are considered to be the earliest bibliographic databases. The Silver Platter is believed publish first biomedical CD-ROM database the MEDLINE with search facility – SPIRS. The growth and development of bibliographic databases has continued since then and enters the digital era to serve the users. The present paper theoretically examines EMBASE, PubMed and IndMed databases.

Keywords: Abstracting and Indexing Services; Bibliographic Databases; Biomedical Literature; EMBASE; PubMed; IndMed

Introduction

The abstracting and indexing periodicals, also termed as secondary periodicals, have been serving the users and the librarians. The first Abstracting and Indexing (A&I) periodical came into being in 1820, with the *Pharmacopeia of the United States (https://en.wikipedia.org/wiki/United_States_Pharmacopeia).* Germany based *ChemischesZantralblatt* also was published around this period[1]. The famous *Chemical Abstracts Service*, by the American Chemical Society, started in 1907 and its print version continued till 2004 and is now available in the digital form on the internet. The A&I services have covered every field of universe of knowledge. However, it can be presumed that the number of A&I services in sciences, engineering and technology are more in number than in social sciences and humanities.

Williams[2] highlighted the emergence of bibliographic databases in 1974. Rice[3] profiled several bibliographic databases and gave an account of their various uses such as in collection development. The US National Institute of Health (NIH)[4] has presented the march of MEDLINE since its inception in 1971 and also outlined the emergence of PubMed in 1996 and narrates more "milestones".

Other biomedical databases to name are the BIOSIS and the EMBASE. The development features of these databases are highlighted in the context of optical storage technology and online information retrieval systems (DIALOG)[5]. Sridhar[6] highlighted the use of CD-ROM technology for bibliographic databases. The Silver Platter (Silver Platter Information and Silver Platter MEDLINE)[7] contributed to the first biomedical database on CD-ROM with search software known as; Silver Platter Information Retrieval Software (SPIRS). The development of full-text databases in medicine was reported in detail by Sievert et al[8]. The authors distinguish them by full-text retrievable only and full-

text retrievable and searchable. According them, 61 databases indexed as medicine and biomedicine is found on CD-ROM. This pace shows that the CD-ROM technology was used for both reference and full-text databases.

Nozoe[9] presents an evolution of medical bibliographic information services and has featured the developments from 1960 to 2007 and discusses the metamorphosis of health information services profiling the PubMed and MEDLINEplus.

In this paper an attempt is made to look at the features three biomedical bibliographic databases viz., PubMed, EMBASE and IndMed.

Bibliographic database concepts

The database concept was conceived and implemented to handle the data stored in separate files and integrated into one, and managed by software called database management system (DBMS). Tedd[10] states, "since the late 1960s many organizations involved in producing abstracting and indexing periodicals have used computer systems in the production of printed bibliographic information sources and the output was the bibliographic databases". So the database concept emerged for managing bibliographic data in 1960s and the first such example was the magnetic tape version of the *Chemical Titles* that appeared in 1961 as an alerting service, produced by the Chemical Abstracts Service, USA[11].

Biomedical bibliographic databases

The genesis of biomedical bibliographic databases can be traced to *Index Medicus* of National Library of Medicine, USA which began in 1879 as an index to biomedical literature and later was published as an abstracting service. The *Index Medicus* mechanization project realized in 1948 resulted into computerized information storage and retrieval system, known as MEDLARS, which became operational in 1964. The MEDLINE is the online version of MEDLARS.

The second important, comprehensive biomedical bibliographic database is the *Biological Abstracts* that started in 1926 and its computerized bibliographic version *BIOSIS* came into being in 1969. The use of BIOSIS by medical librarians was studied by Bishop[12]. The coverage of biomedical literature of *BIOSIS Previews* was compared with core biomedical sources like MEDLINE and EMBASE and it was found that the coverage of *BIOSIS Previews* was more on biological sciences than biomedical subjects[13,14].

The third is this series is *Excerpta Medica* started in 1947 and published by Elsevier. In the initial years, it covered over 3500 biomedical journals published throughout the world. In addition to all fields of medicine, *Excerpta Medica* also extensively covered drug and pharmaceutical literature (DIALOG)[15]. The *Excerpta Medica* Database became EMBASE from 1985 onwards (DIALOG)[16].

Development features

Storage and processing technologies

As early as 1982, it was reported that "combination of forces has conspired to transform database services "*from slow starter to high flyer*". It further adds that "among them the prime stimulants are 'advances in telecommunications coupled with rapid decrease in the cost of computer processors and the storage media'[17]. The trend continued and the cost of processing and storage has come down from US$ 193,000.00 per GB in January 1980 to US$ 0.07 per GB in 2009. It is further stated that "over the last 30 years, space per unit cost has doubled roughly every 14 months.

Networking and WWW

Another important turning point and supporting technology is networking and the Internet and World Wide Web providing remote and anytime, anywhere access to information, which added as a supporter to the diminishing costs of storage and processing technologies.

Optical storage technology

The optical storage technology also termed as "the New Papyrus" with multimedia features made its beginning in the 1980s, and with the high density disks becoming available, the bibliographic databases used this technology to deliver the content to the libraries around the world. Sridhar wrote, "CD-ROM databases gave a good break-through in promoting the use of secondary journals"[6]. They also introduced the CD Servers for accessing several CD-ROM Disks via a network and the CDNET Integrated Server by Meridian-Data was introduced in early 1990s, which was a cost effective, easy to use, high-performance solution for sharing many CD-ROMs on LAN[18].

BIOSIS was the one of the bibliographic databases used in this context. Biomedical databases were the first to use this technology, with Silver Platter pioneering the market entry with the first CD-ROM database coupled with search software SPIRS (Silver Platter Information Retrieval Software). The growth of optical disks namely, the CD-ROM technology was substantial in 1990s for the database storage. Rao[19] reported that the "CD-ROM Directory 1995 (4th Edition) published by TFPL, London, has listed 9551 CD-ROM Databases of 1900 publishers from 390 countries. These databases have been created by 2445 database providers".

Online information retrieval services

The online information retrieval services started with several experimental projects from 1954 with application of computers to bibliographic searches[20]. Bourne mentioned that "an investigation of online bibliographic search was first made by Bagley in 1957"[20]. Bourne and Hahn[21] in 2003 published a book entitled *"A history of Online Information Services, 1963-1976"*. The vital role of System Development Corporation (SDC), ORBIT and the Lockheed's DIALOG Services could also be recalled here, for their exclusive Online Information Retrieval Services. Finally, in 1971 the MEDLARS Online, as MEDLINE, went into operation as a nationwide database that included all journals indexed in *Index Medicus*.

Digital technology

The digital technology as is well known uses only the digits 0 and 1 and so can easily be represented in electronic circuits[22]. The digital technology applications in library saw the emergence of digital libraries, with library resources becoming digital information resources. Lahiri stated, "With rapid developments in performance and declining cost of devices required for digitisation, technology is no longer an important issue"[23]. The technology was used to build large bibliographic databases, which facilitated storage, search and retrieval functions. The digital technology also enhanced speed of data transmission, which enabled to networking of computers to access and share the resources from remote locations.

Technology convergence

Harmonious integration and convergence of multiple technologies have benefitted the bibliographic database products and services. The bibliographic data is a very complex unlike the other kinds of data. The bibliographic record standards like, ISBDs, MARC and the recent metadata standards have helped in the creation, search and access to database content and in particular the full text databases.

Implications on PubMed, EMBASE and IndMed

The genesis of PubMed can be traced to the *Index Medicus*. It includes MEDLINE and its contents date back to 1879. The database has over 25 million citations/records[24]. Even though it provides Web access, through virtual data transmission, the storage requirement is also enormously vast and thus the developments in the storage technologies have been one of the factors for the present growth scenario of these biomedical bibliographic databases.

The processing technology has also made its contribution in processing several variables, as far as PubMed/Medline is concerned. They cover over 5600 biomedical journals and some outside that subject too. The entire database is indexed to facilitate searching by keywords and maps by the MeSH terms. Taking into consideration, 25 million records, and each record has at least 5-10 metadata elements – for example; author(s), author Affiliation, Title, Keyword(s), Source, Language, Form, Abstract, and so on. Eliminating duplicates and repetitions, the search process would require searching 100+ million terms approximately. However the MeSH has only about 27,000 subject headings, but the user's query language is not restricted to only those 27,000 terms, but the query in natural language has to be transformed to the standard vocabulary. The coverage of PubMed dates back to 1879 virtually and has both Born Digital and Archival Materials. The citations of these databases use the NSTCD standards[25]. The technological developmental factors have a very clear bearing on these bibliographic databases, in searching and retrieval.

The EMBASE is a successor of ExcerptaMedica, published by Elsevier, The Netherlands, since 1947 (selectively from 1902)[24]. It has now 29 million records from 8500 journals. It includes all of MEDLINE citations, plus 6 million more than the MEDLINE and indexes 2000 journals which are not covered by MEDLINE and are European Union titles. It also includes 260,000 abstracts from over 1000

National and International Conferences held all over the world. The MEDLINE covers most USA based journals, whereas EMBASE's coverage is from 90 countries. This again profiles the EMBASE's storage requirements.

As far as processing is concerned it indexes more number of journals than PubMed and in addition to MeSH terms' synonyms, it has discovery search facilitated by the EMTREE, which goes for a deep search with advanced search facilities to match the natural language terms. The EMBASE, apart from all biomedical subjects has specialty content for Drugs and Pharmaceutical Subjects. It uses the Web 2.0 platform's RSS tool to enhance the search options and email alerts. These are some of the distinct features of EMBASE and it is possible with the developments in the processing technology.

IndMed and MedInd are the two national level web-based indexing and document delivery services, funded by the Indian Council of Medical Research (ICMR) in association with National Informatics Centre (NIC). IndMed is an index of select Indian medical journals. It covers about 110 Indian biomedical journals from 1985 onwards and was started in 2010[26]. The Indian journals in IndMed are also covered by the PubMed. The IndMed also makes use of the technological advances as in the case of PubMed and EMBASE. However, there are not many studies on the use of IndMed and analysis of its services.

Conclusion

From being print-based, A&I services have evolved being web-scale discoverable and in future, users shall be able to search through A&I services seamlessly.

References:

1. Wiegand WA and Davis DG. Jr. eds. *Encyclopedia of Library History,* (Routledge:Taylor and Francis Group; New York), 1994, p.3. Available at:https://books.google.co.in/books?isbn=1135787506. (Accessed on 7 April 2016)
2. Williams ME, Use of machine readable databases. *Annual Review of Information Science and Technology,* 9 (1974) 72-74.
3. Rice BA, Evaluation of online databases and their uses in collection evaluation, *Library Trends*, 33(3) (1985) 297-320.
4. National Institute of Health (NIH) U.S. National Library of Medicine.2013, Available at: https://www.nlm.nih.gov/-news/medline_35th_birthday.html (Accessed on 6 October 2016.)
5. DIALOG Information Retrieval Service. *Database catalog,* 1980, p.14.
6. Sridhar MS, *Managing modernisation of Library services using IT: Potentials and problems.* In DRTC Workshop (30). Advances in Information Technology: Impact on Library and Information Field. DRTC, Indian Statistical Institute, Bangalore, 1996,Paper-DD. p.7.
7. Silver Platter Information Inc. and Silver Platter MEDLINE(R) on CD: National Library of Medicine: (Available at http://www.company-histories.com/SilverPlatter-Information-Inc-Company-History.html) (Accessed on 6 October 2016.)
8. Sievert MEC, McKinin, EJ and Johnson ED, Full-text databases in medicine, *Journal of the American Society for Information Science.* 46 (10) (1997) 748.
9. Nozoe A, Evolution from medical bibliographic information services to health information services – MEDLARS, MEDLINE/PubMed and MEDLINEplus, *Journal of Information Processing and Management.* 50 (9) (2007) 580-593.
10. Tedd LA, *An Introduction to computer-based library systems*. 2nd edn. (John Wiley; New York) 1984, p.56, 216.
11. Sathyanarayana NV, *Online access: A beginners first lesson.* In Seminar on Library Automation and Information Retrieval. DRTC, Indian Statistical Institute; Bangalore,. 12-14 August, 1985, p.34.
12. Bishop D, Control and dissemination of information in medicine, *Advances in Librarianship,* 2 (1971) 102.
13. Van Camp G and Freeman AJ, BIOSIS Previews and MEDLINE – a biomedical perspective, *Online,* 1(1) (1977) 12-16.
14. Snow B, Alternative medicine information sources, *Database,* 2 (3) (1998) 18-20.
15. DIALOG Information Retrieval Service. *Database Catalog,* 1980, p.14.
16. DIALOG Information Services Inc. *Database Catalog*, 1992, p.36.
17. Far Eastern Economic Review. *Database services: From slow starter to high flyer. Industrial Japan '82.*1992, p.57.
18. Jayakanath F and Ravi S, A. *Access to CD-ROM Databases on IISc Campus Network: Experience of NCSI.* In DRTC Workshop (30). Advances in Information Technology: Impact on Library and Information Field. DRTC, Indian Statistical Institute, Bangalore,1996 Paper-DL. p.3.
19. Rao SS, CD-ROM Technology: World scenario and Indian experience, *Information Today and Tomorrow*, 15 (2) (1996) 3-4.
20. Bourne CP, Online systems- History, technology and economics, *Journal of the American Society of Information Science,* 31(3) (1980) 155-160.
21. Bourne CP and Hahn TB, A history of online information services, 1963-1976. (Cambridge, MA). 2003. A Review of the book published in the *Journal of the Medical Library Association*, 92 (2) (2004) 277-278.

22. Visual Encyclopaedia. *Technology and transport*. (Pentagon Press, New Delhi) 2002. p.7.

23. Lahiri A, Digitalisation–a new mandate: Editorial, *Information Today and Tomorrow,* 20 (2) (2001) 1.

24. Rickman K, Librarian, King Edward Memorial Hospital, Perth, Australia. *(Adapted for FIU by Barbara M. Sorondo)* Available at http://kemh.health.libguides.com/library/-search_tips/faqs/difference_between_pubmed_medline_embase. (Accessed on 4 April 2016)

25. National Standardization Technical Committee of Documentation in 1979, 25 national standards have been issued. A national committee and several sub-committees were set up corresponding to the format of ISO/TC 46. Sub-Committee No. 6 is in charge of the standardization of information retrieval 1979.

26. IndMed, Available at: http://indmed.nic.in/ (Accessed on 4 April 2016).