Abstracting Database

Related terms:

Web Search Engine, Database Vendor, Electronic Content, Library Network, Research Material

Locating scholarly papers of interest online

Maureen Henninger, in Social Media for Academics, 2012

Overview of online scholarly search services

The rapid growth of the Internet began with three not quite concurrent events: the development of the World Wide Web in 1989; the lifting of the ban on commercial activity on the Internet (what was at that time NFSNet) in 1991; and the release of the first point and click browser, Mosaic, in 1993. In fact it was out of a need to collaborate and share scholarly information among the community of particle physicists at CERN (Conseil Européen pour la Recherche Nucléaire) in Switzerland that the team, led by Tim Berners-Lee, developed the web (Henninger, 2008).

In the 1980s there were many publicly available collections, archives and repositories of scholarly documents such as NetLib, as well as the large proprietary indexing and abstracting databases. While the main system available to search across much of the content of individual databases was ISI's Science Citation Index (at that time accessed via Lockheed's Dialog service),[1] in the public sphere, other than early systems such as Archie or Gopher, there were no tools for searching and retrieving scholarly articles.[2] One of the earliest of the public scholarly search engines was CiteSeer, which was developed at the NEC Research Institute to provide access to cited scientific scholarly literature and was available on the web in 1997. It was really the first decade of this century, powered by robust search engine technology for crawling the Internet and access by agreement or joint venture to some of the large indexes of proprietary databases[3] which saw the development of public, web-based scholarly search engines (see Figure 4.1).

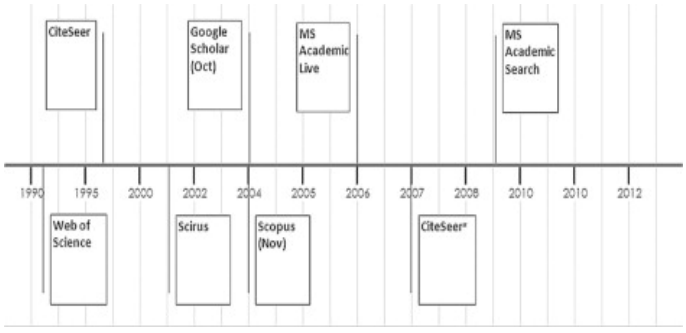

Figure 4.1. Timeline of the release of online academic search engines

Read full chapter

URL: https://www.sciencedirect.com/science/article/pii/B9781843346814500049

Higher education libraries

Shirley Oakley, Jennifer Vaughan, in Libraries in the Twenty-First Century, 2007

Research

Libraries maintain their traditional role in supporting research, but the changing research environment changes the way support is provided. Electronic access to search tools, indexing and abstracting databases and content have reduced reliance on large physical collections of research journals but increased the complexity of access pathways. Different aggregations of content have different search interfaces,

and multiple logins, adding complexity to the library's task in assisting researchers to use them. The demand for seamless discovery of and access to content over multiple platforms has led to the provision of search portals which provide a single search interface with direct links to content on many platforms. The Australian Academic and Research Library Network (AARLIN) is one such portal. AARLIN aims to provide a collaborative research information infrastructure that will:

- Streamline information access through targeted search options.
- Offer a choice of targeted services at the point of need on the research desktop.
- Share processes and workloads to maximise electronic resource discovery & delivery.
- Provide a portal for these services.
- Expedite access through use of single-sign-on and authentication technology (http://www.aarlin.edu.au/about.shtml).

Research is increasingly multidisciplinary, with research teams working across institutional and national boundaries. E-science, 'large scale science that will increasingly be carried out through distributed global collaborations enabled by the internet' (Research Councils UK 2006), requires new types of support infrastructure. Research data and research results produced electronically offer opportunities for improved access by researchers to information, but also require new methods of preservation. The Australian Research Information Infrastructure Committee (ARIIC) funds projects which will improve the access of Australian researchers to the information they need to carry out their research and make the results of Australian research widely available and easily accessible. Libraries are directly involved in funded infrastructure projects: Australian Research Repositories Online to the World (ARROW); Australian Digital Theses Program Expansion (ADT); Australian Partnership for Sustainable Repositories (APSR); and Regional Universities Building Research Infrastructure Collaboratively (RUBRIC). New initiatives in e-Science which facilitate researcher collaboration, data sharing and data preservation offer new challenges for libraries. The Dataset Acquisition, Accessibility and Annotation e-Research Technologies (DART) project, for example, builds on ARROW by storing datasets as well as the published research, making it possible to include annotations on datasets and publications.

Increased competition for research funding under the proposed Research Quality Framework (RQF) (DEST 2006c) places a significant value on the ability of university libraries to support researchers. There is increased pressure on libraries in the 'Dawkins' universities which do not have a long history of collecting for research support. The ability of the library to support researchers is considered a competitive advantage in the RQF funding regime.

Information sources

Alastair Smith, in Libraries in the Twenty-First Century, 2007

Indexing and abstracting databases

Accessing information in documents such as periodicals and conference proceedings has been a challenge for libraries. While general bibliographic databases such as Libraries Australia record, for example, the details of a periodical, say, *New Scientist*, they do not list the specific articles within the periodical. This is the role of indexing and abstracting databases. Indexing and abstracting databases have a long history, particularly in the sciences where much information appears in journals. Information in journal articles and conference proceedings is indexed and, in abstracting databases, a short summary or abstract is provided (see Chapter 9 for an account of indexing). An example of a long running indexing and abstracting service is *Chemical Abstracts*, started in 1907 by the American Chemical Society and indexing journal articles, conference papers and patents in pure and applied chemistry.

Originally indexes were produced manually, but with the advent of computer typesetting the printed indexes began to be generated from a computer database. The next stage was for the computer database itself to be searched, initially in batch mode – a program would be run overnight that would scan the database for particular combinations of keywords – but by the 1970s it became common for searches to be done in real time from remote computer terminals. Several 'database vendors' emerged, for example, Dialog (http://www.dialog.com/). A

database vendor acquires indexing and abstracting databases from a number of producers and mounts them on a server with a common search interface, charging users for access and paying a royalty to the producer.

A key to effective searching of databases is use of the subject descriptors attached to the computer records. These are generally drawn from a thesaurus or subject headings list. An example of this is the medical subject headings (MeSH) that provide the descriptors in the Medline database, which is the main database in the medical and health sciences, available on the open web as PubMed (http://www.pubmed.gov/). In order to search effectively for 'cancer', for instance, it is necessary to use the MeSH term 'neoplasm'. Another example of the use of specialised descriptors is Chemical Abstracts Registry Numbers (RNs) – if one wants to search for 'aspirin' in a database using Registry Numbers, one can search on 50-78-2 (the RN for aspirin) rather than all the alternative brand and chemical names for the material.

Citation databases are a specialised type of indexing database and work on the principle that scientific papers cite related, previous papers. To take an example, a searcher interested in the impact of rabbits on grazing and aware of Cooke's 1987 paper on the subject in the *Australian Journal of Ecology* can trace papers that cite Cooke's article, using a citation database, and have access to more recent information. Law case citation indexes had been compiled manually for a long time (from 1873 in the case of Shepards), but Eugene Garfield realised the advantages of computers in the creation of citation indexes for scientific information and in 1964 Garfield's Institute for Scientific Information (ISI) launched Science Citation Index. Initially a computer database was used to produce printed citation indexes but the index is now available as the web-based database, Web of Knowledge (http://isiwebofknowledge.com/), with coverage of social sciences, arts and humanities, as well as the conventional sciences. While ISI was for a long time the only general citation database vendor, competitors have appeared in the twenty-first century, such as Elsevier's *Scopus* (http://www.scopus.com/).

Until the 1990s, most database vendors were geared to information professionals. Databases charged by time spent online and had sophisticated interfaces that required people to be trained to use them effectively. Databases gave only bibliographic details, and library skills were required to locate the actual documents and obtain them by inter-library loan or using a commercial document delivery service. In the 1990s, however, it became increasingly common to link full text of articles with the indexing and abstracting databases, and with the growth of web-based user-friendly search interfaces, it has become common for library users, particularly in academic libraries, to access databases that allow seamless movement from an indexing function to a document supply function. Database vendors have been joined by journal publishers such as Elsevier and MCB in providing online access to the lull text of journals.

The 'open access movement' has facilitated the availability of research material on the open web through electronic journals, authors websites and institutional repositories. As a result it has become common to search for research material through internet directories and search engines. Google has taken this a step further with Google Scholar (http://scholar.google.com/) – an extension of the Google database, which indexes research writing on the open web. Despite scepticism that Google Scholar has the consistency and accuracy of more conventional indexing and abstracting databases (Jacsó 2005b), web search engines are likely to play an increased role in searching for research material.

The value of searching indexing and abstracting databases is that they access a much greater depth and authority of information than is available on the open web. Despite the open access movement, much research writing appears only in proprietary databases and is available only to members of subscribing institutions such as universities or research institutes. A valuable function of libraries in the twenty-first century is managing the availability of these proprietary databases to their clients both inside the library building and remotely from their home or workplace.

Qualitative analysis of journals' websites

Bhaskar Mukherjee, in Scholarly Communication in Library and Information Services, 2010

Miscellaneous issues

Other miscellaneous issues which are important for measuring quality of journals are provided in Table 4.6. Like the NOA journal JAM, 10 OA journals had the facility 'electronic mailing list' to alert the users to recent publications and upcoming events. This may be due to the fact that publishers of these OA journals wanted to circulate the published articles to a wider audience. Recently (during the writing of this chapter), *DLM* has introduced the RSS facility, which is yet to be adopted by other OA journals. Despite the fact that *LISTA* indexed 15 journals, *LISA* indexed 10 and the Web of Knowledge indexed only *INR*, most OA journals did not mention this on their websites. There were 8 indexing and abstracting databases which indexed *FIM*, 6 databases indexed *DLM* and *INR*, 5 indexed *LRS*, and 4 indexed *ARD* and *ITD*. The number of indexing and abstracting databases which indexed other OA journals is also mentioned in this table. This result is not consistent with the finding of Bjork (2004) in which he mentioned that 'one major drawback of OA journals so far has been that they have rarely been indexed in commercial indexing services'. Only two OA journals were enlisted in the Open Journal System, others are yet to be included. In most cases, the author owned the copyright. In 9 out of 16 OA journals, the copyright remained completely with the author. This may be due to the fact that all articles of these OA journals were available free of charge and if publishers ever changed their access policy, they will not be affected by access boundaries. It is also one of the major aims of Open Access publication. However, in two OA journals the copyright ownership was extended to both the author and the publisher, while in another two it was with the publishers only. *INR* used the 'Creative Commons License' for copyright and in *FIM* authors can choose either 'Creative Commons License' or 'Creative Commons Public Domain Dedication'.

Like all three NOA journals, the Google page rank was above 5 in 12 OA journals. ISI did not index any of these OA journals except *INR*. None of the journals carried advertising or product marketing, although *INR* carried a provision for advertising. This may be because most of these OA journals receive support from their sponsoring bodies (or may be because potential advertisers do not see the worth of advertising in these journals). The systematic arrangement of contents and hyperlinks made the overall appearance of these OA journals attractive and navigation of them is very simple. The subject coverage of 11 of the OA journals was not limited to the particular domain of librarianship, whereas the remaining 5 were slanted toward specific types of librarianship. The electronic ingredients of *DLM*, *ARD*, *EID*, *INR*, *IST*, *JDI* and *SJI* were very comprehensive and complete.

Introduction

Tibor Koltay, in Abstracts and Abstracting, 2010

The structure of the book

This book has eight substantive chapters of different lengths. The main content of the first two chapters is very simple, and this first introductory chapter is followed by a chapter on definitions.

Chapter 3 is considerably longer and discusses the characteristics of the abstract – the length, functions and types of abstract. The latter issue receives especial attention as we intend to present a holistic picture on informative, indicative and indicative-informative abstracts. It is in this chapter that we discuss the relationship between the different types of abstract and their function.

The subsection on the objectivity of the abstract addresses an issue that is also of great importance. One aspect of the discussion on objectivity revolves around the question of whether abstracts should be critical or not. To answer this question, we include the case of *Mathematical Reviews* (2008) which – despite the name – is an abstracting database that encourages the publication of critical abstracts. Notwithstanding this, other aspects of objectivity are also treated in this book.

The author abstract is also relevant to the discussion in Chapter 3. As the knowledge and skills related to this type of abstract are essential to a wide range of readers, we consider it in detail and include advice to writers of author abstracts and a discussion of the issues relating to structured abstracts.

Chapter 4 provides the answer to another question: 'What does an abstractor have to know?' This part of the work addresses the knowledge, skills, attitudes and competencies needed to become a good abstractor.

To get a deeper insight into the nature of abstracting, we outline some problems of everyday and semi-professional summarisation. It is particularly useful to give special attention to the questions of proficiency in summarisation because abstracts are created in the process of summarising.

The information literacy context of abstracting is also explained because it is the most modern and in many regards burning question for a wide circle, including all those who may be interested in abstracting. In light of this, a subsection discusses the relationship between information literacy and abstracting. This chapter concludes with a short subsection which deals with the teaching of abstracting.

Chapter 5 is about the practice of abstracting. In this chapter you will become acquainted with the terminological questions of abstracting. Possible structures of abstracts are also outlined. We give advice on what to leave out and what to include in a good abstract. We also discuss whether the abstract should be written in one or more paragraphs and what the sequence of information reflected in the abstract should be.

The subsection on types and techniques of reading is followed by thoughts on the language used in abstracts: the past tense and the third person, negation and negatives, the passive or active voice and the vocabulary.

Chapter 6 provides two examples of the creation of abstracts, showing the phases of analysis, editing and generating the final abstract. This chapter concludes with a short subsection about the evaluation of abstracts.

Chapter 7 bears the title 'Beyond language and style'. Were we to express the content of this chapter in the simplest way, we would say that it is about everything that we could not say in the previous chapters which is, nonetheless, of interest. The title of one of the subsections is telling: 'A genre in its own right'. It is in this chapter that we give insight into knowledge acquired about abstracting and comprehension, because comprehension is foundational to abstracting, provided that we take into consideration the differences that exist between 'normal' comprehension and comprehension directed towards abstracting.

Last but not least we address other related processes such as indexing and translation. The very last subsection throws light on some additional aspects of abstracting, and includes among others some models of the process. Finally, Chapter 8 recapitulates the main message of the book in a short form. The book also contains an index in which we indicate only those concepts that cannot be easily identified from the table of contents and do not figure in this introduction or the conclusion.

Library Applications

Johanna Olson Alexander, in Encyclopedia of Information Systems, 2003

VI. Electronic Content (E-Content)

Content creation, publishing, distribution, and transfer have changed dramatically because of text, image, and sound digitization. There are distinct formats of electronic content (e-content) but the linkages among these different formats are rapidly increasing. The resources discussed in the following section are available through licensed, pay-per-use, and/or free databases usually accessible via the Internet/Web.

VI.A. e-periodicals

Full text periodicals (including e-journals or e-zines, short for e-magazines) are usually text or HTML (HyperText Markup Language) files. These files are cheaper to produce, more easily downloaded, take less computer disc space than graphics and image formats, and are more easily manipulated and transferred. Full text periodicals usually lack graphics and are more affordable for the producer, but are less helpful to the user.

Full image periodical articles give the reader text and graphics such as charts, illustrations, etc. Generally, these image files, accessible from the Internet/ Web through a browser, are dependent upon external software programs (plug-ins) for creating, transferring, and reading files. Such programs as *Adobe Acrobat* are often used in conjunction with full image files but require the user to have access to the free reader program. These graphic files are in a transferable form called portable document format (PDF).

Increasingly, e-content providers are converting to XML formats (extensible Markup Language) to provide content *repurposing* or cross-media publishing—that is, the creation of content that can be distributed in multiple formats rather than creating content for one particular media. Content can then be made available through various means such as the Web, personal digital assistants, print, etc.

There are a variety of e-periodical "packages" or products. These packages may be provided by publishers (or alliances of publishers) such as *Project Muse* from Johns Hopkins University Press or Elsevier's *ScienceDirect*; library community retrospective and archival backfile initiatives such as *JSTOR* (journal storage project); publisher and scholarly society partnerships such as Stanford University's *HighWire Press* which focuses on value-added electronic functionality and graphics in science, technology, and medical journals; aggregator services such as EBSCOHost, which negotiate full text/image access to periodicals; commercial publishers that provide paper subscriptions and limited electronic access; and electronic only journals. These packages may be free but are more often fee-based licensed resources.

Other initiatives have sought to speed, expand, and archive scholarly communication within the research community and then provide access over the Internet. One example is *arXiv*, a free electronic preprint or e-print archive of scientific research in physics, mathematics, and other scientific areas. Authors can submit, replace, and/or remove papers from the *arXiv* collection and researchers can access the materials.

VI.B. Indexing and Abstracting Databases

Databases that provide indexing, bibliographic records, and abstracts (not full text/image) are another type of e-content form. These types of basic bibliographic/abstracting databases are enhanced by linkages to digital full text/image materials. Some of these databases provide links to selected full text/image materials but increasingly full e-content inclusion in the database or linkages to full e-content will be mandatory for a vendor to stay competitive.

VI.C. Linking Software and Services

E-journals/periodicals are most useful when the title has a stable URL and/or a persistent URL (PURL) allowing a direct and permanent link that is maintained in metadata search information. Some aggregators provide only e-journal access through their search interface and do not provide stable URLs. To improve electronic resource access, some aggregators negotiated stable URL connections for customers and developed methods to load 856 MARC field information into library databases and ILS. This allows users to access e-periodicals directly from library online catalogs. New aggregator *linking software and services* are cropping up. These services may have variant ways of processing or even slightly different products, but all are based on OpenURL specifications. Linking companies that have access to multiple vendors and database producers are now acting as link aggregators. Examples of these linking services include *jake: Jointly Administered Knowledge Environment* (http://jake-db.org/) and Openly Informatics Inc. (http://www.openly.com/link.openly/).

CrossRef is also based on OpenURL specifications and is a cooperative initiative of publishers. Implementation of *CrossRef* means a "researcher can click on a reference citation in a journal and immediately access the cited [reference, abstract], or article ... [using] Digital Object Identifiers (DOI), which are tagged to article metadata supplied by the participating publishers." (Publishers International Linking Association)

Linking technologies are now being integrated into ILS products. SFX is an example of a citation linking system based on OpenURL. SFX is being used by the ILS company Ex Libris, allowing customizable searching and access by the library "... to dynamically create [context-sensitive] links that fully integrate ... [a library's] information resources regardless of who hosts them—the library itself or external information providers." (http://www.sfxit.com/)

VI.D. Electronic Books (e-Books)

Electronic books, or e-books, are becoming more prevalent but have not had the popularity once forecast. Nevertheless, e-books are finding their way into academic and commercial markets. E-books come in forms including: (1) e-book apparatus that hold downloaded books, (2) entire book collections accessible via the Web, or (3) electronic reference sources. E-books may be freely available, as in the *Gutenberg Project* (http://promo.net/pg/), the *National Academy Press* (http://www.nap.edu), or *xrefer* (http://xrefer.com/), or for purchase or subscription, as is the case with *netLibrary* (http://www.netlibrary.com). These systems allow the user to search the book text, review and read, possibly highlight sections and/or take notes online,

print and download selections, and, depending on the pricing and licensing model, “check out” the e-book for in-depth reading and research. Other e-book reference sources may provide searching but not browsing. If there is a stable URL link, the e-book can be cataloged in a library's ILS providing a direct link to the electronic book. Some of these systems are self-contained systems that do not require any special software while others require plug-ins for specific access levels. Due to copyright protection issues and pricing model experimentation, there is a good deal of variation in the e-book industry. The National Institute of Standards and Technology, Open eBook Forum (OeBF), and the American Association of Publishers are some of the e-book standards setting organizations.

VI.E. Other Formats Including Data, Maps, and Gis

Statistical databases present data in HTML, PDF, or other easily readable formats. Many of these databases also allow data extraction for customizable reports and/or downloading in multiple formats such as comma-delimited, HTML, or spreadsheet formats. These variations are convenient, allowing the user to view and/or download data in a textual format or a spreadsheet format for calculation purposes. Appropriate plug-ins or data extraction programs are sometimes required. XML formats used with various tools will allow additional manipulation capabilities of research data. “XML Cover Pages” (http://xml.coverpages.org/) provides a useful list of XML applications in library, archival, and museum settings.

Cartographic and geospatial data, digital maps, and individual print maps are available in libraries. Map products range from easily accessed and updated digital maps and mapping systems available through the Web to highly specialized and technical geographical information systems (GIS). General issues of concern to libraries include:

- Indexing and accessing both print and digital map collections
- Ensuring quality relating to accuracy, print quality, and scale
- Providing appropriate hardware, software, and trained personnel to provide quality maps and services