

SAVVY SEARCHING

FORUM

Péter Jacsó

*The University of Hawaii at
Manoa
jacso@hawaii.edu*

In this era of rapidly increasing options for information retrieval, information professionals and librarians need to be more and more aware of the need for savvy searching. Savvy searching for me means knowing (a) what information sources to use, (b) on what medium, (c) offered by which publisher, (d) at what price and (e) at what real cost. In other words, using an actual example, you have to know when to use the *Social SciSearch* database of the Institute for Information Science, or the *Social Science Abstracts* of H.W. Wilson, or *PAIS* of the eponymous company or a Web-borne alternative. (I use this adjective to distinguish those databases that were born on the Web from the ones that had existed before the Internet and were merely ported to the Web from their proprietary systems, or the telnet-based Internet version, and are now borne by the Web. Amazon is a Web-born database, but *Books in Print* - in any of its Web incarnations on DIALOG, Ovid, Infotrac - is merely Web-borne.

Once you decide that for a particular topic PAIS is the most appropriate because, say, it has the most international journal base and the topic has a strong interdisciplinary and public affairs slant on an international scale (like environment protection), you have to decide which of its versions to use - the CD-ROM version or the online version? If you consider the latter alternative the question arises: should you use the version published by Dialog, Ovid, SilverPlatter or OCLC? You may need to know which one has the most current version - if currency is of high importance to you or to your client. You need to know what is the actual price per connect hour and per item displayed/printed.

Price is never a direct predictor of real costs, however. You don't always get what you pay for. Which search service you are most familiar with may have more impact on the bottom line. Beyond that it matters how intuitive the user interface is when you have to brush up your knowledge of the search procedures on the potential hosts without incurring much charge. Your ability to do a quick and efficient search may also depend on whether the PAIS thesaurus is part of the database to look up the terms of the controlled vocabulary and see the broader, narrower and related terms. Would the software allow you flexibility in sorting the output by language, journal name and date? Price itself would not guide you in such issues, because it might well happen that the version that has the thesaurus implemented, offers the most flexible output definition and sorting options and has the most intuitive interface also has the lowest price - as is the case, for example, with Ovid's implementation of ABI/INFORM. Typically, it also means the lowest cost because you are likely to spend the least time preparing and conducting the search and postprocessing the output.

If you lean toward the CD-ROM version as a librarian and consider subscription for your college (rather than for a single search) then which one would be more appropriate: the one published by PAIS or the one published by SilverPlatter? The former has only a DOS version, the latter comes in DOS, Windows and Mac formats. You may shun the DOS version until you realise that its windowing technique is quite attractive, and its language options (English, French, German, Spanish,

Portuguese) pop up automatically in the publisher's version but are not even browsable in the other DOS version. But it is browsable in the WinSPIRS version, which offers user-defined formats and reasonable sort options, and the prices are comparable. Then again, browsing in the field-specific indexes of the publisher's version is better than in the basic index of the SilverPlatter version, which lumps together words and terms from the title, the descriptor and the annotation (abstract) fields. Beyond all these technical issues, the decisive criterion may be which search software the students and faculty (and your staff) are more familiar and comfortable with.

This was a relatively easy scenario for several reasons. One is that all the file producers (Institute for Scientific Information, H.W. Wilson and PAIS) considered for the online search are known to have high quality products. All of them have a clear policy of journal coverage, and the policy is known to be applied consistently. The list of journals covered is readily available to check, and the journal names are entered in consistent and correct format (a rarity in databases). Their updating pattern is predictable and reliable.

So are their classification and indexing, but with their own weaknesses and strengths, which the savvy searcher is supposed to know or find out about from his or her own experience or from published reviews or articles like this one. Records in the ISI database have abstracts only if one was published with the original. The Wilson databases usually have the most detailed and most informative abstracts. PAIS has the shortest annotation but its subject indexing is perhaps the best. ISI uses automatically generated index terms (Keyword Plus). H.W. Wilson uses modified Library of Congress Subject Headings, which may have been appropriate for the printed abstracting/indexing services with their preference for strong pre-coordination, but for the interactive online and CD-ROM searching ad-hoc post-coordination by users is much better. In addition, only PAIS has a publicly available thesaurus. In terms of currency, ISI is unbeatable, but for a specific search may have fewer of the most important titles covered than the other two. All of the online database services have reliable and powerful search software (although for full-text databases and multiple databases there is still no real competition for Dialog - yet).

The decision may be far more difficult when there are many more options (as is the case with the versions of most databases produced by the US government, such as ERIC, GPO, MEDLINE and LC-MARC),

or with less well-known databases and/or publishers on the Web (as is the case with the great variety of quotations databases and bilingual dictionaries, and with some of the book directories and catalogues).

On the subject of the Web, there is such a level of concern about low-quality Web resources that sooner or later the American Psychological Association will classify it as a disease. Every John and Jane who gets publication space or a public-speaking opportunity parrots these concerns. The truth is that many of the dubious-quality Web sites, and especially the Web databases, can be recognised a mile off by an experienced information user. (This may not be true for the novice - and therefore innocent - Web user, but this journal is meant for the former category, leaving plenty of opportunity for general interest magazines to address the latter group.) The other truth is that low-quality databases have been with us information professionals for decades before the Web phenomenon, and many years before the widespread use of online/CD-ROM searching. Classic studies of printed abstracting/indexing services from the 1960s by Alan Gilchrist in *Aslib Proceedings* and by John Martyn and Margaret Slater in the *Journal of Documentation* pointed out serious deficiencies. Wilfrid Lancaster in the *Bulletin of Medical Library Association*, Mary Turtle and William Robinson in *RQ*, Charles Bourne in *Information Processing and Management*, and Martha Williams and Lawrence Lannom in the *Journal of the American Society for Information Science* in the 1970s warned about the unprofessional collection, organisation and processing of data in professional databases. This was continued throughout the 1980s and 1990s by Carol Tenopir, Reva Basch, Anne Mintz, Ruth Pagell and many others, who have been fighting against mediocre database quality for years and have been given due credit in a recent review of database quality issues in a top-ranking serial publication (Jacsó, 1997).

The 'Péter's Picks and Pans column' in *Database* magazine has never had a shortage in the past five years of databases that deserved panning for the brutal neglect of purportedly core journals, the utterly inconsistent treatment of author, journal, country and language names, the irresponsible handling of supposedly controlled vocabulary terms, or the dereliction of duty with regard to database updating. Perceived clout does not always matter. You can even find signs of database-wide recklessness in an expensive directory that debuted in 1998, although it is compiled by a well-known company, and published by a traditional online publisher.

Most of those who hopped on the bandwagon of generally dissing the quality of Web-based information sources do not seem to be aware that traditional online and CD-ROM file producers and publishers often have committed worse offences against high-paying customers than those who make resources available on the Web for free. Many of the free Web resources turn out to be more complete, consistent, accurate and timely than their fee-based competitors. There should be no blind trust in 'professional' databases produced and published by stalwart companies in the information industry, nor global distrust of or prejudice against unknown databases that appear on the Web out of the clear blue yonder. That is what a savvy searcher has to know, or has to be able to recognise. The secret to survival is the same as in

playing cards: 'to know when to hold 'em, know when to fold 'em, know when to walk away, and know when to run', whether those databases are Web-based or traditional ones.

And that is what Péter Jacsó is going to help readers of *Online and CD-ROM Review* to do in his Savvy Searching column discussing the principles and practical methods of evaluating and comparing databases.

Reference

Jacsó, Péter. Content evaluation of databases. In Williams, Martha E., ed. *Annual Review of Information Science and Technology* (ARIST), volume 32. Medford, NJ: Information Today, Inc. for the American Society for Information Science, 1997.

The author

Péter Jacsó

Péter Jacsó is an associate professor of the Information and Computer Sciences Department at the University of Hawaii. He is the recipient of the 1998 Pratt-Severn/ALISE National Faculty Innovation Award for his innovative use of information technology in curriculum design and coursework. He is a frequent speaker at national and international conferences. His columns are published in *Database*, *Information Today*, *Computers in Libraries* and *Link-Up*. He has received various awards for his writings, including the 1998 Louis Shores/Oryx Press Award of the ALA Reference and User Services Association for his discerning database reviews.