Searching and Reviewing the Computer Science Literature : A Guide for Research Students

A Brooks

Abstract

It is impossible to account systematically for all relevant previously published material. This article describes ways of searching and reviewing the computer science literature which should help optimize detection of relevant material and so put research on a surer foundation.

Keywords : bibliography, literature, review, search

1 Introduction

It is impossible for any individual to account systematically for all relevant previously published material. The problem arises as to what subset of the literature (books, journals, conference/workshop proceedings, technical/internal/research reports, theses, newsletters/bulletins, and electronic bulletin boards) anyone ought to read. At least two factors compound this problem. Firstly, there is an enormous range to the depth and quality of reporting in the computer science literature. The journal publications of the BCS, ACM, and IEEE, are good starting places; but it could be the unrefereed workshop paper which is the most important to read. Secondly, new journals seem to be advertised almost monthly. Both the BCS and the ACM are expanding their range of journals. Whilst advice from supervisors about suitable reading material will be useful, remember that supervisors have a lot less time to spend trying to keep up-to-date with the literature. This article, primarily intended for beginning research

students, describes ways of searching and reviewing the computer science literature. Many large organisations employ professional information scientists to carry out literature searches. By contrast, academic researchers generally act as their own information officers. Hopefully this guide communicates the most important things any beginning research student ought to know and do. Most obviously, you should attend any induction course offered by the main library, keep a brief of searches undertaken, and consult earlier guides and bibliographies e.g. [7], [6], and [8]. A recently published guide to information sources in artificial intelligence/expert systems [2] is good. This guide also has some coverage of the general computer science literature and should be consulted after reading this paper.

2 A First Approach

2.1 General Articles

As the new research student is often not in a position to assimilate full length journal papers, without a great deal of effort, you can make a better start by reading articles in the more popular computing magazines and journals e.g. *Byte, Communications of the ACM, IEEE Software* and *Computer*. Scanning through volumes for the last few years should locate easily read articles of relevance. The December issue of the *Communications of the ACM* has a useful annual index. If any difficulty is encountered then even more general articles can be found in encyclopedias such as the *Encyclopedia of computer science and engineering* [9], the *Encyclopedia of Artificial Intelligence* [10], or even the *McGraw-Hill Encyclopedia of Science and Technology* [5]. If jargon presents a problem there exist several useful dictionaries such as the *Dictionary of Computing* [3]. Know the jargon of your topic, as this can influence the quality of key word searching, whether manually through abstracts or online to a database such as *INSPEC* which includes the online equivalent of *Science Abstracts C : Computer and Control Abstracts*. The reference section of your main library

will have a selection of such dictionaries and encyclopedias. Knowledge of these should be obtainable without moving from your own desk as most main libraries have networked online access to their catalogues. Get details from your Departmental library representative or systems support. Wide area networks also make it possible to search the catalogues of the main libraries of other institutions. Substantial review articles should then be found and read.

2.2 Substantial Reviews

Review journals include *Computing Surveys*, *Advances in Computers*, and *Annual Review of Computer Science*. Scanning the *Annual Reviews of Information Science and Technology* might also provide relevant articles. Check if your library has the *Index to Scientific Reviews* and the *Index to Scientific and Technical Proceedings*, both published by the Institute for Scientific Information. There may have been an entire conference devoted to your research topic, some of the articles of which should contain review material. Check the *Lecture Notes in Computer Science* series. Pay a visit to local bookshops that serve your institution. A sales assistant may be able to direct you to a recently published book on your chosen specialism. Look out for the *Tutorial* series from the *IEEE Computer Society Press*. Do also check recent higher degree theses awarded by your Department. A previous student may have written a review covering part or all of the literature that interests you. If your main library does not have a review paper your interested in, this does not mean that you can't obtain a copy. Get out on Inter-Library Loan any review paper (and indeed any paper that you need to read) that your main library does not have.

There are two main types of review. Some reviews merely present summaries of the literature, whilst other reviews contain a substantial critical element. The former have their uses, but make sure you locate some of the latter. Incidentally, one of the hallmarks of most good research papers is a section on related work with a substantial critical element. *Review* might be a good key word to use in a search of the literature (in conjunction with other key words describing your

research topic).

2.3 ACM Guides

At an early stage consult the *ACM Guides to the Computing Literature*, which are undoubtedly some of the best literature aids. Your Departmental library should have a set of these as well as your main library. These guides are prepared annually and the usual key word indexing is supplemented by other indexing methods. Scan these guides for say the last ten years. You should easily find dozens of relevant papers.
The ACM Press have recently brought out three specialist bibliographic guides - Resources in Human-Computer Interaction, Resources in Ada, and Resources in Parallel and Concurrent Programming. Make sure your main library gets these guides and others following in the series. By the way, if you insist on writing an algorithm, do check the volumes of *COLLECTED ALGORITHMS FROM ACM*.

2.4 Abstracts

Manual abstract scanning is perhaps the least glamorous way of locating relevant literature. All the same you should do it. Determine which sections of the abstracts are relevant and scan for say the last ten years. Your main library should have at least one relevant abstracting journal e.g. *Science Abstracts C : Computer and Control Abstracts* and *Computer and Information Systems Abstracts*. Also scan the relevant sections of *Computing Reviews*, which present critical summaries of selected articles. Be aware that some very specialised abstracting journals have been launched in recent years such as the *Turing Institute Abstracts in Artificial Intelligence*. With the experience of manual searching behind you, you will be more effective at online searching (see later). You will also appreciate the importance of being able to write good abstracts for your own publications.

2.5 Reading Papers

What do you do with all the copies of papers that you have begun to accumulate or are about to accumulate ? You read them, of course. But how ? In assessing relevance, some experienced researchers will read the abstract, introduction, and conclusion sections first. They might also glance at any diagrams or tables in the main body of the paper and quickly scan the references to see what literature is cited. Sometimes the title is enough for experienced researchers who, with full mastery of the research context, go straight to the main body of the paper. If the paper is particularly relevant, then you must try to read and **understand** the main body of the paper. You might have to contact the author(s) to clarify certain points. Don't be hesitant about doing this. You might also have to read and understand some of the referenced papers. If you can criticise the contents, you will have understood a paper. Read to criticise as there's a lot of sloppy work about. A copy of a well read paper by an experienced researcher will usually be annotated with underlines and question marks in an attempt to clarify the strengths and weaknesses of the paper and to distinguish between fact and opinion. There are really no shortcuts to becoming an effective reader of scientific papers - read, read, and read some more.

Whilst making notes from papers, and building up your own collection of copies of papers, it is **extremely important** that you record **full** bibliographic details. This will be useful when you want to refer to something again (inevitably), and when you want to cite for your thesis. If you don't take the necessary care, valuable time can be easily wasted. Fortunately, most journal papers have relevant bibliographic details inscribed on each page. Books and conference/workshop proceedings, however, have essential bibliographic details (e.g. editor(s), publisher, place of publication, edition, date) spread over one or two pages at the beginning. Don't forget.

2.6 Foreign Language Material

If foreign language material is encountered, several things can be done. It is unlikely that funds will be available for professional translations, so first find out if the author has published an earlier or later article on the same theme in English, or in a journal which receives automatic translation (e.g. *Soviet Journal of Computer and System Sciences*). Locate the original as sometimes an English abstract is given or the tables and diagrams may give clues to relevance. The librarian, someone in the modern languages department, or even a fellow researcher, may be willing and able to translate some paragraphs to help assess relevance. You might also get lucky and find yourself working next to a research group which specialises in machine translation. Details about searching for a possible translation are given in [2], or ask the librarian. Research students in the UK and North America are truly fortunate in that English, according to Large [4], is the most frequently used medium 'accounting for between 50% and 90% of publications in the major disciplines'.

2.7 Browsing

Do not be afraid to spend time browsing around either your Departmental library or the main library whenever you cannot be bothered doing much else. Some of the most important articles are found by chance, just this way. Some researchers even advocate setting aside some time each week for browsing. Your Departmental library may have ad hoc lists of internal/technical/research reports from other institutions and even copies of some of these articles if a reciprocal distribution agreement exists. Such reports provide the most up-to-date information, but are not usually refereed, and are of varying depth and quality. When you're bored, attack this so-called *grey* literature.

Browsing the serials section of your main library may turn up journals that you had otherwise not considered as core journals for your research topic. You may also turn up special issues of journals which are devoted entirely to your research topic. Browse the books section as well for conference proceedings.

A quick browse in the computer science reference section at nearby University of Glasgow main library revealed bibliographies on database design, abstract data types, parallel programming, expert systems, robotics, logic programming, simulation, and privacy and computers. *Bibliography* might be a good key word to use in a search of the literature. Incidentally, get to know the main library of any neighbouring institution. The journal coverage might be better, and you might not need to make quite so many Inter-Library Loans.
If you're working in an applied area, do also browse the trade/professional magazines. Some knowledge of users and markets can be quite enlightening.

2.8 A First Review

At the end of the first approach, after say three months, write a review. Only by tackling some research of your own will you eventually develop a proper critical appreciation for others' work; when writing a first review, however, there are some obvious things that can be done. Try to distinguish fact from opinion. Point out areas of controversy where conclusions differ. If the fundamental issue at stake can be spotted then so much the better. Distinguish between conclusions which are tentative and those that are secure. Report if error estimates are missing in empirical work. If your own statistics are shaky, see [11]. Easy criticisms can usually be made regarding the failure to properly execute and analyse empirical work (e.g. see the criticisms by Brooks [1] [1] of some empirical work in human-computer interaction) and the failure to provide a framework for comparison (e.g. vision researchers often use their own favourite pictures in image processing work and schedulers often only work with their own particular timetabling problem). Sometimes you have to move outside of the mainstream computer science literature to come across criticisms. For example, you can easily gain the impression from the AI literature that expert systems are marvellous and that their use can produce huge cost savings. But in the area of computer-aided medical diagnosis, according to Sutton [13], the benefits accrue

[1] Well, there's nothing like a self-citation.

from better record keeping and not through the use of computer diagnosis per se. Sutton's paper was found by scanning abstracts in *Index Medicus* after a helpful librarian had pointed out the relevant sections.
After the first approach the literature must be attacked again. By this time, of course, your supervisor will have started you on a short research exercise to get your feet wet.

3 And Beyond

3.1 Citation Searching

All the papers you have located thus far will have reference lists tracing the development of the subject backwards in time. But you can trace forwards in time from these papers, and so get even more up-to-date, by doing citation searches using *CompuMath*. A few explanatory words are in order here. *CompuMath* comes in the three main parts : the Source Index, the Permuterm Index, and the Citation Index. The Source Index is arranged by first-named author and gives titles and references of papers. The Permuterm Index is arranged by key word, allowing a reader to locate a paper containing certain key words in its title. The Citation Index indicates who has cited a given paper - this is a remarkably useful facility for tracing developments of ideas forward in time, criticisms to papers, and errata in papers. *CompuMath* is published throughout the year and so is more up-to-date than annual accumulations such as the *ACM Guides to the Computing Literature* which are often not published until the latter half of the following year. Doing citation searches at this stage usually shows up deficiencies in earlier scanning. Use the Permuterm Index for key word searches. A copy of *CompuMath* will probably not be kept in your Department library, but seek it out at your main library. Do use it. You should find citation searches of review papers particularly valuable.
The *Science Citation Index*, which has a wider source coverage than *CompuMath*, should be used in addition to *CompuMath* for citation searches on

key papers. This can be particularly important if the research topic is multi-disciplinary.

3.2 Specialist Interest Groups and User Groups

A very good way of keeping up-to-date not only with the literature but also forthcoming workshops, conferences, courses, products and services, etc. is to join one of the many dozens of *Specialist Interest Groups* (*SIGs*). Many of these groups produce their own bulletins or newsletters. Some publish the electronic mail addresses of members, thus making it very easy to contact other researchers. Specialist bibliographies are occasionally published in the bulletins or newsletters and you may see advertisements for specialist online databases (e.g. the *HILITES* database for human-computer interaction) and electronic bulletin boards. If a published paper poses you several questions, you could consider quizzing the author(s) via electronic mail - but please make sure your questions are to the point. The BCS has *SIGs* ranging from *Advanced Programming* through to *Technology for Software Protection*. Likewise the ACM has *SIGs* ranging from *SIGACT* (Automata and Computability) through to *SIGUCCS* (University and College Computing Devices). Details of the BCS *SIGs* can be found in the BCS *SPECIALIST GROUPS HANDBOOK* [2] and most main libraries actually stock the output of the ACM *SIGs*. Government-led research initiatives also spawn *SIGs* to facilitate cooperation between academia and industry.

User Groups spring up around all sorts of activities, both hardware and software related. If you are using a fairly sophisticated piece of hardware or software in your research, the chances are that there is a *User Group*. Such groups help disseminate information about errors in the product, new product releases etc. By joining *SIGs* and *User Groups*, you will find out about the research community active in the area of your research topic. If there are active research groups in nearby institutions, then arrange a visit and make personal contact.

[2] Your Department should have a copy somewhere .....

Such groups may have already established a large collection of reports and other materials that are difficult to access.

3.3 Use a Computer. Why Not ?

The last decade saw a huge expansion in online bibliographic databases. *INSPEC* provides some of the best coverage of computer science and you should contact your main library about access and costing procedures. Also consider using *COMPUSCIENCE* which is the online version of *Computing Reviews* and the *ACM Guides to the Computing Literature*. It is usual that some member of the library staff will do the search of these online databases for you, but in your presence. The search is normally preceded by a discussion of key words to use, as an inappropriate choice can be costly in terms of time and money. With so many hosts and online databases available, it is perhaps well to have knowledgeable library staff perform the searches. Library staff should be able to identify any specialist online databases which give coverage of reports in the area of your research topic. They are also good at remembering to use synonyms and wild card characters when dealing with different key word endings. Some researchers do, however, prefer to interact with the systems themselves. Depending on the host(s) used, academic discounts offered, and depth of search, a single session can typically cost between £20-£60. Some databases are distributed as CD-ROMs (e.g. the *Applied Science and Technology Index*) and you should check the CD-ROM collection of your main library. CD-ROM databases are usually updated quarterly.

You should be a more effective searcher of the online databases if you follow the suggested approach and first get experienced at manual searching. Be aware that even the most specialised of online databases have their weaknesses. In a recent article on online AI databases, Stern [12] recommends a search of a variety of databases to ensure fairly comprehensive coverage of the AI literature. It is likely, however, that searching even just one specialist online database will retrieve as many papers as you can reasonably manage.

The University of London's Computer Literature Database (CLDB) is worth a mention here. Although it is limited in periodical coverage, it has a citation facility, and the service is currently free to the academic community. E-mail A.BROWN%BIRKBECK@qmw.ac.uk for details.
In 1991, through a joint purchase agreement, many of the Institute of Scientific Information's files became accessible over the UK academic network *JANET*. Look out for the information sheet from your main library.

3.4 Network News

Established communication networks link computers together worldwide. *Network News*, also called *Usenet*, is the most popular. Try typing *rn* at your terminal or workstation. Immediately subscribe to *newsgroups* with the description *newuser* in their titles and read the introductory articles. If need be, get advice about *Newtork News* from your systems support or a more experienced research student. Spend some time searching for the *newsgroups* that cover your area of interest and subscribe to them. It has to be said, however, that a lot of *Network News* has a poor signal/noise ratio. The better *newsgroups* usually have their content edited by a moderator. In all likelihood your Department will have a *local newsgroup* called something like *cs.research*. Do subscribe to this as well. You might find that some *Network News* contains information about specialist bibliographies that can be downloaded using *anonymous ftp*. *Anonymous ftp* indicates that a user can log into a remote site as user *anonymous* with an arbitrary password (a convention is to use some sort of identification such as your electronic mail address) and so download databases and software of interest. Other *Network News* may contain lists of available internal/technical/research reports which, as mentioned earlier, contain the most up-to-date information. It may even be possible to read an electronic version of a conference paper before actual going to the conference. Again, if need be, get advice from systems support or a more experienced research student. Incidentally, reading and contributing to *newsgroups* can be a way of getting hold of answers to all sorts of

technical questions, both hardware and software related. A word of warning, however. Discipline your use of *Network News* and don't get hooked on it.

3.5 Conference/Workshop Proceedings

Conference/workshop proceedings contain much more recent material than journals and so are a very important source of information. Browsing through recent proceedings will reveal the current focus of research in a whole area. It can be difficult, however, to get hold of proceedings of conferences or workshops that you personally don't attend. If you see a poster advertising an important conference or workshop (on *Network News* or on a Departmental notice board, for example) that you can't attend, take down the necessary details so that you know who to approach about getting a copy of the proceedings or so that you can try to obtain the proceedings through an Inter-Library Loan.

4 Final Remarks

By the end of your first year as a research student you should have completed a substantial, critical literature review. You should then have a well ordered collection of papers (with full bibliographic details) occupying one or two drawers of a filing cabinet. As your project progresses, keep up-to-date by scanning the core journals which cover your area of interest. Also scan *Current Contents : Engineering, Technology and Applied Sciences*, which is produced weekly by the Institute for Scientific Information and which lists the titles of papers in the main journals as they are published. Keep in mind a handful of important papers and maintain a citation search on them using *CompuMath* and the *Science Citation Index*. Don't forget to browse and keep track of relevant *Network News*.

Try to get along to at least one workshop or conference per year. Some of these provide opportunities for research students to present their work. Participating will sharpen your presentation skills and knowledgeable members of the audience

may help by criticising your work and directing your attention to some other relevant literature.

Once well into your research project, you will benefit from reading theses in your area. See [2] or ask the librarian for details on how best to locate relevant theses. One source of information is *Dissertation Abstracts International*, which is available on CD-ROM. If you have trouble trying to locate any material, do speak to the librarians.

Every time you have an idea, ask yourself : 'has somebody already done this ?'. Hopefully this article goes some way towards ensuring that you will be able to answer this question quickly and effectively.

5 Acknowledgements

I would like to thank Marc Roper, James Miller, Andrew D McGettrick, Tony O'Grady, Ian Watson, George McCaskill, and an anonymous critic for their suggestions.

References

[1] A Brooks. Evaluation : Where formality demands informality ? Research Report HCI-1-90, Department of Computer Science, University of Strathclyde, 1990.

[2] P J Hancox, W J Mills, and B J Reid. *KEYGUIDE TO INFORMATION SOURCES IN Artificial Intelligence/Expert Systems*. Mansell Publishing Limited : London, and Ergosyst Associates, Inc. : Lawrence, KS, USA, 1990.

[3] V Illingworth, editor. *Dictionary of Computing*. Oxford University Press : Oxford, third edition, 1990.

[4] J A Large. The foreign language barrier and electronic information. *ONLINE REVIEW*, 14(4):251–266, 1990.

[5] McGraw-Hill Encyclopedia of Science and Technology, 1987.

[6] Darlene Myers, editor. *Computer Science Resources : A Guide to Professional Literature*. Knowledge Industry Publications, Inc. : New York, 1981. Published for the American Society for Information Science.

[7] Alan Pritchard. *A Guide to Computer Literature, An Introductory Survey of the Sources of information*. Clive Bingley Ltd. : London, second edition, 1972.

[8] LITERATURE IN COMPUTING. Pages 867–872 in [9].

[9] A Ralston, editor. *Encyclopedia of computer science and engineering*. Van Nostrand Reinhold : New York, second edition, 1983.

[10] S C Shapiro and D Eckroth, editors. *Encyclopedia of Artificial Intelligence*. Wiley : New York, 1987. 2 vols.

[11] G W Snedecor and W G Cochran. *Statistical Methods*. Ames, Iowa : Iowa State University Press, seventh edition, 1980.

[12] D Stern. Artificial Intelligence Databases : A Survey and Comparison. *DATABASE*, 13(4):19–24, August 1990.

[13] G C Sutton. Computer-aided diagnosis : a review. *British Journal of Surgery*, 76:82–85, January 1989.