

Digital Libraries and Search Engines for Software Engineering Research: An Overview

Mark Turner, Keele University

The aim of this document is to provide informal information relating to the major search engines and digital libraries that are frequently used by the software engineering research community. The details provided are particularly focussed towards the needs of evidence-based researchers performing systematic literature reviews but the details may be useful to researchers performing any form of literature survey. Where the information is available, information is provided relating to the resources indexed by each digital library or search engine, and other relevant information concerning the searching facilities provided. It is intended that this document will be updated periodically with information relating to different electronic databases or search engines, or as new information is discovered about the databases already covered.

1. Digital Libraries

The following provides details on the main digital libraries utilised by software engineering researchers, specifically IEEE Xplore, ACM Portal, and ScienceDirect.

1.1. IEEE Xplore

Firstly, it should be noted that IEEE Xplore lists search results ranked by “relevance”, where relevance is calculated by the following algorithm:

- Frequency of the search term in the full abstract/citation record
- Text length of the abstract/citation record
- For AND queries, proximity of the search terms throughout the abstract/citation record

1.1.1. Access restrictions

A list of what can be accessed by subscription level (IEEE Member, Institutional Subscriber, or Guest) can be seen at the following URL:

http://ieeexplore.ieee.org/guide/g_oview_access.jsp

1.1.2. Searching

Use of the basic search facility without using field codes to direct where to search, results in the search defaulting to the field code ‘metadata’. The field code ‘metadata’ searches the following:

Document Title, Author, Publication Title, Abstract, Index Terms, Affiliation

Therefore, the basic search does not search the full text of the indexed resources/PDF files. Full text can be searched when using the basic search facility by including '<in>pdfdata' at the end of the search string. However, a login is required to search the full text of indexed articles.

Logging in allows the user to search specifying certain field codes, as outlined in Table 1:

Abstract	ab
Affiliation	cs
All fields	metadata
Author(s)	au
Catalog number	ca
CODEN	cn
Conference date	cy
Document title	ti
Editor(s)	au
Full text	pdfdata
Index term	de
ISBN	in
ISSN	in
Issue number	is
Issue part number	pt
Meeting date	cy
Part number	pt
Publication name	jn
Publication year	pyr
Subject term	de
Title	ti
Volume	vo

Table 1 – Field Codes available in IEEE Xplore
(http://ieeexplore.ieee.org/xplorehelp/Help_fieldcodes.html)

Logging in also allows the user to access the advanced search facility (see Figure 1):

OPTION 1
Enter keywords or phrases, select fields, and select operators Help

in
 in
 in

» Note: If you use all three search boxes, the entries in the first two boxes take precedence over the entry in the third box.

OPTION 2
Enter keywords, phrases, or a Boolean expression Help

» Note: You may use the search operators <and> or <or> without the start and end brackets <>.
» Learn more about [Field Codes](#), [Search Examples](#), and [Search Operators](#)

Publications

Select publications

- IEEE Periodicals
- IET Periodicals
- AIP/AVS Periodicals
- IEEE Conference Proceedings
- IET Conference Proceedings
- IEEE Standards

Other Resources (Available for Purchase)

- IEEE Books
- Educational Courses

Standard Status
(Applies to IEEE Standards only)

Status:

Select date range

Search latest content update (12 Jan 2010)

From year: to:

Display Format

Citation Citation & Abstract

Figure 1 – Advanced Search Screen in IEEE Xplore

A complete user guide to IEEE Xplore can be accessed at the following URL:
http://ieeexplore.ieee.org/guide/g_oview_guidepdf.jsp

1.1.3. IEEE Computer Society Digital Library (<http://www.computer.org/portal/web/csdl/home>)

The IEEE Computer Society Digital Library (CSDL) indexes journals and conferences proceedings published by the IEEE Computer Society. The list of the resources indexed by the CSDL is available at the following link:
<http://www.computer.org/portal/web/csdl/home>

The advanced search facility offers the ability to search particular journals by name, and also to restrict the search to only resources to which the user is subscribed (see Figure 2). It also offers similar search facilities to IEEE Xplore, such as the ability to search via a restricted set of field codes (Full Text, Title, Author, Date, ISSN/ISBN, or Exact Phrase) as well the ability to combine search phrases using Boolean operators.

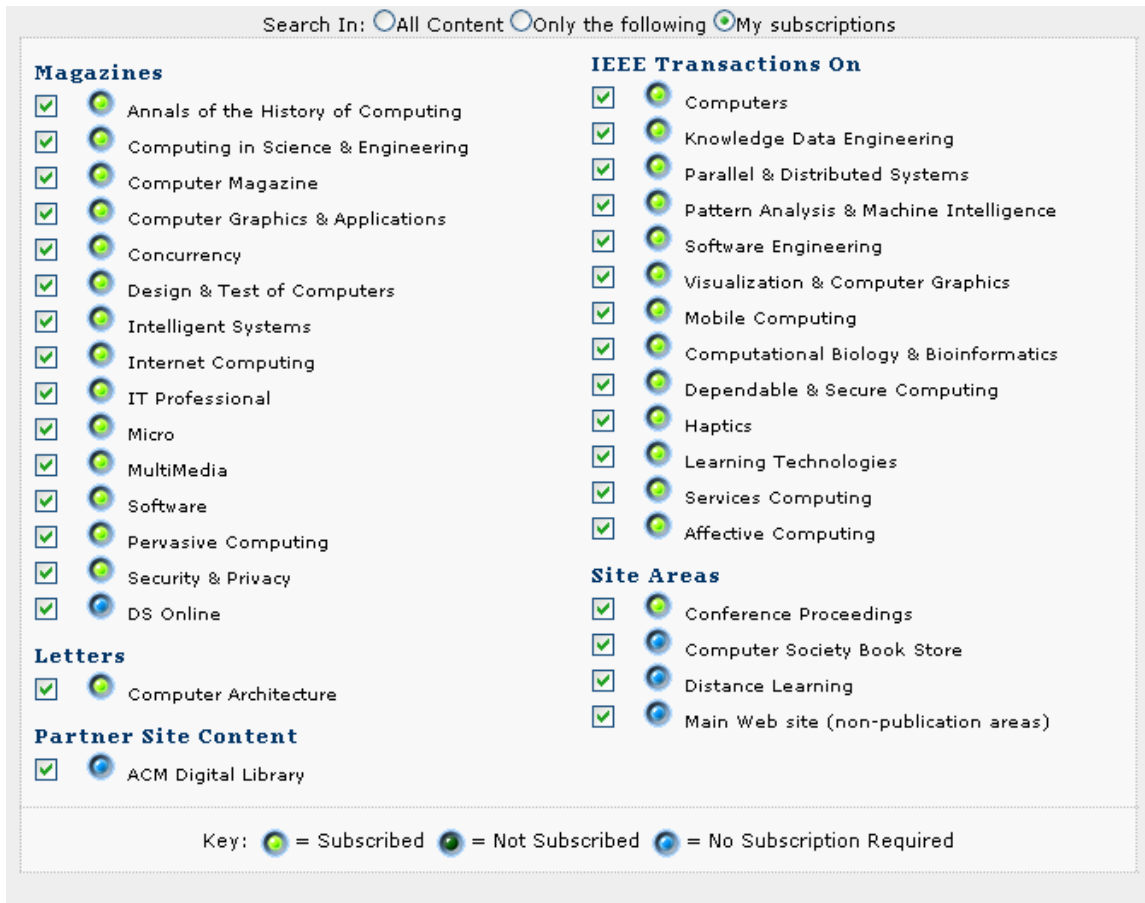


Figure 2 – CSDL Advanced Search Facility

A known paper, published by the IEEE Computer Society, was searched for in both the CSDL and IEEE Xplore and the paper was found in both digital libraries (Turner *et al.*, 2003). This result indicates that the CSDL content is a sub-set of the content indexed by IEEE Xplore. Further evidence is provided by the fact that the IEEE Computer Society journals are listed under the content indexed by IEEE Xplore, for example IEEE Transactions on Software Engineering (<http://ieeexplore.ieee.org/xpl/periodicals.jsp?letter=S&type=0&list=all>).

The IEEE Conference Publishing Services Web site also describes how all conference publications are published in both IEEE Xplore and IEEE CSDL.

“...all CPS published conference publications are automatically included in *IEEE Xplore* and the *IEEE Computer Society (CSDL) Digital Library*.” (IEEE Computer Society CPS, 2009).

1.2. ACM Portal (<http://portal.acm.org/dl.cfm>)

The ACM portal indexes journals and conference proceedings published by the ACM, along with selected publications from affiliated organizations. A list of publications by affiliated organizations is available:

http://portal.acm.org/browse_dl.cfm?linked=1&part=affil&coll=portal&dl=ACM&CFID=65759424&CFTOKEN=68988647

The above link provides the following details:

- [ALGOL Bulletin](#) *Computer History Museum*
- [Computational Linguistics](#) *MIT Press*
- [International Journal of Network Management](#) *John Wiley & Sons, Inc.*
- [Journal of Computing Sciences in Colleges](#) *Consortium for Computing Sciences in Colleges*
- [Linux Journal](#) *Specialized Systems Consultants, Inc.*
- [Mobile Networks and Applications](#) *Kluwer Academic Publishers*
- [Personal and Ubiquitous Computing](#) *Springer-Verlag*
- [Proceedings of the VLDB Endowment](#) *VLDB Endowment*
- [The Journal of Machine Learning Research](#) *JMLR.org*
- [The VLDB Journal — The International Journal on Very Large Data Bases](#) *Springer-Verlag New York, Inc.*
- [Wireless Networks](#) *Kluwer Academic Publishers*
- **ACL Workshops:** [ACL Workshops](#)
- **ACL:** [Annual Meeting of the ACL](#)
- **ANLC:** [Applied Natural Language Conferences](#)
- **CASCON:** [IBM Centre for Advanced Studies Conference](#)
- **COLING:** [International Conference On Computational Linguistics](#)
- **CSCL:** [Computer Support for Collaborative Learning](#)
- **EACL:** [European Chapter Meeting of the ACL](#)
- **HLT:** [Human Language Technology Conference](#)
- **ICLS:** [International Conference on Learning Sciences](#)
- **MUC:** [Message Understanding Conference](#)
- **NAACL-ANLP:** [ANLP/NAACL Workshops](#)
- **NAACL:** [North American Chapter Of The Association For Computational Linguistics](#)
- **TINLAP:** [Theoretical Issues In Natural Language Processing](#)
- **VLDB:** [Very Large Data Bases](#)

In terms of searching, ACM by default searches the *full text and all metadata* of all indexed articles. The advanced search (Figure 3) is also available without the need to have an account.

Enter words, phrases or names below. Surround phrases or full names with double quotation marks.

SEARCH

Words or Phrases
Find [any field] with
all of this text (and)
any of this text (or)
none of this text (not)

Names
Find [any field] with names
using all any none of the names

Keywords
Find author's keywords
using all any none of the keywords

Affiliations
Find company or school
using all any none of the affiliations

Publication
Find publication
using all any none of the names
Published since [year] Published before [year]
In publication types Journal Proceeding Transaction Magazine Newsletter

Conference
Find sponsor names
using all any none of the names
Find year (yyyy)
using any none of the years

Find location
using any none of the locations

Identification codes
Find ISBN/ISSN
Find DOI

Computing Classification System (CCS)
Find node
Find subject/noun

Required components
Results must have Full Text Abstract Review

Figure 3 – ACM Advanced Search Screen

1.3. ScienceDirect (<http://www.sciencedirect.com/>)

ScienceDirect is a digital library that primarily indexes material published by Elsevier.

1.3.1. Indexing

ScienceDirect indexes the following:

- Over 2000 journals published by *Elsevier* (including *Information and Software Technology*).
- Plus dynamic linking to journals from around 350 Science/Technology publishers through CrossRef (<http://www.crossref.org/>).

A full list of titles available through ScienceDirect is available:

http://pts.sciencedirect.com/journalholdings/sd_holdings_guest.asp

The above list can be searched to determine if ScienceDirect indexes a particular journal. Note that both Web of Science and Scopus are also able to search Elsevier journals.

1.3.2. Searching

The basic search allows for searching of the following:

- All Fields (which searches all metadata and the full text of all articles indexed)
- Author
- Journal/Book title
- Volume, Issue, Page

The advanced search facility can be accessed only when logged in, which also includes Athens or institutional access. The advanced search, as shown in Figure 4, allows the user

to search for specific terms within particular fields ('Abstract, Title and Keywords', Authors, Specific Author, Source Title, Title, Keywords, Abstract, References, ISBN, ISSN, Affiliation, or Full Text).

Figure 4 – ScienceDirect Advanced Search Screen

2. Search Engines (index multiple sources)

The following provides details about a selection of the other search facilities available which, instead of specialising in a particular publisher, instead index content from multiple digital libraries.

2.1. Web of Science (<http://isiknowledge.com>)

2.1.1. Searching

The basic search facility in Web of Science allows the user to search using various search terms and Boolean operators. However, the search forces the user to search in a pre-defined list of fields and, as it only indexes content from other sources, does not allow for the searching of the full text. The field tags that can be used are as follows:

TS=Topic
 TI=Title
 AU=Author
 GP=Group Author
 ED=Editor
 SO=Publication Name
 PY=Year Published
 CF=Conference
 AD=Address
 OG=Organization
 SG=Suborganization

SA=Street Address
CI=City
PS=Province/State
CU=Country
ZP=Zip/Postal Code
FO=Funding Agency
FG=Grant Number
FT=Funding Text

The Boolean operators that can be used are as follows:

AND
OR
NOT
SAME

The advanced search facility allows for searching using the above field tags as well as to combine them, e.g.

TS=(nanotub* SAME carbon) NOT AU=Smalley RE

It also allows for restricting the search to particular years, languages, and different document types. However, like the basic search facility, the advanced search does not appear to allow the user to search the full text of the indexed articles.

2.1.2. Indexing

Web of Science searches five databases:

- Science Citation Index Expanded
- Social Sciences Citation Index
- Arts and Humanities Citation Index
- Index Chemicus
- Current Chemical Reactions

The Science Citation Index Expanded database is of most relevance as it includes Computer Science as a discipline. The SCI-expanded database covers over 8,700 journals, and is maintained by Thomson Scientific. The computer science journals that are covered include IEEE, ACM and IET published journals. A search of the master journal list included:

IEEE Transactions on Software Engineering,
IEEE Software,
IEEE Computer
Empirical Software Engineering,
Information and Software Technology,
IET Software,
Software Practice and Experience.

The master journal list for Web of Science is available, and can be searched to determine if a particular journal is indexed:

<http://www.thomsonscientific.com/cgi-bin/jrnlst/jlsearch.cgi?PC=MASTER>

Web of Science authorises journals for inclusion and journals can be submitted to be indexed and, if chosen, are added to the above list and included in the search.

2.2. Scopus (www.scopus.com)

Scopus is similar to Web of Science in that journals are authorised for inclusion and the engine has its own index list. Some universities have institutional access, but for those without institutional access Scopus also offers an Athens login. A full list of the journals that it indexes is available:

<http://www.info.scopus.com/detail/what/titles.asp>

The list includes:

- IEEE Transactions on Software Engineering,
- IEEE Software
- Empirical Software Engineering,
- Information and Software Technology,
- IET Software,
- Communications of the ACM.

Similar to many other search engines, Scopus has the ability to select where to direct a search (including “All fields”, “Author”, and “References”) but as it only indexes material published elsewhere, it does not appear to have a full text search facility.

2.3. Google Scholar (<http://scholar.google.co.uk/>)

It is very difficult to find out exactly which journals are indexed by Google Scholar. However, it does index the ACM and IEEE databases and does not appear to include Elsevier journals (Meho & Yang, 2007). Past experience with performing SLRs has however found that there is some crossover between Google Scholar and ScienceDirect, indicating that Google Scholar does search at least some Elsevier journals.

As it is based on a standard ‘web crawler’ method of indexing (as used by Google), it is arguably more suited than the other databases for finding ‘grey literature’ (such as TRs and Masters and PhD theses).

Note that Google Scholar will return several results that are listed as [citation]. These links cannot be clicked on, as according to the Google Scholar help page: “These are articles which other scholarly articles have referred to, but which we haven't found online. A great deal of scholarly literature is still offline, and until these papers are available online, citation-only results can help researchers find as much relevant information as possible.” [Google Scholar Help, 2009].

It is very difficult to determine if Google Scholar searches full text or exactly where Google scholar gets the results that are returned. As it uses the standard web crawler method, it can be assumed that it will search the full text of PDF files when available, or will search what metadata is freely available from the different indexed sources.

2.4. CiteSeer (<http://citeseer.ist.psu.edu/>)

CiteSeer is similar to Google Scholar or other search engines in that the indexing is done automatically via a web crawler. The crawler searches for academic publications in PDF, PS, or Microsoft Word format, with filters to remove any documents that do not meet academic standards. Documents can also be submitted for inclusion via a Web form. Links are made via citations and the references included in the indexed documents.

Therefore, it is very difficult to determine exactly what sources are included but, like Google Scholar, it will also include some ‘grey literature’ that would be not be indexed by the digital library search engines.

In terms of the search facilities, according to the help page “CiteSeer indexes the full-text of the entire articles and citations. Full Boolean, phrase and proximity search is supported.” [Citeseer Help, 2009].

Note that Citeseer appears to be in the process of being replaced by CiteseerX (<http://citeseerx.ist.psu.edu/>).

3. Summary

Table 2 provides a summary of the details of each search engine or digital library covered.

	Digital Library (DL) or Search Engine (SE)	Login Required (Y or N)	Full text search
IEEE Xplore	DL	Y	Y (when logged in)
ACM Portal	DL	Y	Y (without logging in)
ScienceDirect	DL	Y	Y (without logging in)
Web of Science	SE	Y (Institutional access or Athens)	N
Scopus	SE	Y (Institutional access or Athens)	N

Google Scholar	SE	N	Y (although it is assumed this is only possible when the full text is publicly available)
Citseer	SE	N	Y (although it is assumed that this is only possible when the full text is publicly available)

Table 2 – Summary of Digital Libraries/Search Engines

References

Citeseer help (2009): Available: <http://citeseer.ist.psu.edu/citeseer.html> [01/12/2009]

Google Scholar help (2009): Available
<http://scholar.google.co.uk/intl/en/scholar/help.html> [01/12/2009]

IEEE Computer Society Conference Publishing Services (2009), Frequently Asked Questions regarding publishing with CPS, Available:
<http://www.computer.org/portal/web/cscps/faq> [18/01/2010]

Meho, L.I. and Yang, K. (2007) A New Era in Citation and Bibliometric Analyses: Web of Science, Scopus, and Google Scholar, *Journal of the American Society for Information Science and Technology*, 58(13), pp. 2105–2125.

Turner, M., Budgen, D., and Brereton, P. (2003). Turning Software into a Service, *IEEE Computer*, 36(10), pp. 38-44.