

电子信息领域 OA 资源综述^{*}

钱建立 吴广茂

收稿日期:2008-11-17

修回日期:2008-12-26

西北工业大学机电学院,710072 西安市高新路 25 号瑞欣大厦 10A, E-mail: QJL56789@126.com

摘 要 把 OA 资源分为了 Post-Print、Pre-Print 和金色 OA、绿色 OA,对目前支持论文检索的国内外电子信息领域的 OA 资源的主办单位、收录论文数、特点进行了比较,并给出了一些站点输入关键词“Filter”查到的论文数。结果表明,在计算机领域,Citeseer 最有影响,ScienceDirect 的论文整体水平较高,OpenJ-gate 收集的 OA 期刊和论文数更多;国内外中国科技论文在线的更新率较高,Socolar 收录的论文数较多,开网则目前集中于电子信息领域。

关键词 开放存取 预印本 OA 仓储 知识库 比较 电子信息领域

1 开放存取资源分类

在 Paperopen^① 的开发过程中,我们对 DOAJ、Open-J、Scolar 等收录的 OA 资源进行了梳理,并通过互联网对 OA 资源进行了搜索,对找到的 OA 资源进行了分析和归类^[1]。

从 OA 资源来源角度可以分为:

(1) Post-Print 论文。

这里又分为纯网络 OA 期刊(如 Plos Medicine)论文、传统期刊(如《西安电子科技大学学报》)的 OA 电子版和复合 OA 期刊(如 The AAPS Journal)。

(2) Pre-Print 论文。

如中国科技论文在线。

Pre-Print 和 Post-Print 的区别在于是否经过同行评议和结集出版。网络期刊大多结集出版,还经过同行评议(专家审稿),预印本则正好相反。因此,网络期刊论文的质量整体上比预印本论文的质量要高,但也不排除预印本论文中有很多高质量的论文,如 ScienceDirect 上的预印本论文整体质量就很好。

从 OA 论文的网络存在方式又可以分为:

(1) 金色 OA(Gold OA)。

包括上面提到的网络 OAJ 期刊网站、传统期刊 OA 网站。

(2) 绿色 OA(Green OA)。

包括个人自存档、机构仓储。

传统期刊的 OA 是杂志社或者出版商官方实施 OA,而自存档论文则是源于出版机构官方没有实施开放存取但是允许作者把自己发表的论文保存在作者自己的个人网站和所在机构网站上用于交流学习。

随着 OA 的发展,还出现了一些专门收集 OA 论文的仓储站点和搜索引擎,比如 Scirus、Google 学术搜索、Citeseer、Highwire、DOAJ、Open-J、Paperopen 等,这对 OA 的发展起到了积极的推动作用,也方便了读者的使用。下面介绍一些知名的支持论文检索的 OA 资源。

2 电子信息领域国外 OA 资源简介

以下介绍的 OA 资源中都存在电子信息领域的论文,但不一定所有的论文都属于电子信息领域。另外在下述资源中,不一定所有文献全部是 OA,但考虑到存在很多 OA 文献,就一并作为 OA 资源进行介绍。

2.1 Citeseer

网址: <http://citeseer.ist.psu.edu/> <http://citeseerx.ist.psu.edu/>

原名 ResearchIndex,是 NEC 研究院在自动引文索引(Autonomous Citation Indexing,ACI)机制基础上建设的一个

^{*} 陕西省科学技术研究发展计划项目:科技论文开放获取搜索引擎研究,编号 2007K04-11。

^① Paperopen 是空军工程大学开发的 OA 论文平台(www.paperopen.com),目前已经收录 OA 论文一百多万篇。

学术论文数字图书馆^[2], Citeseer 采用机器自动识别技术搜集互联网上 Postscript 和 PDF 文件格式的学术论文。1997 年开始提供互联网服务, 2008 年全面改版, 新版的网址为: <http://citeseerx.ist.psu.edu/>。截至 2008 年 6 月 9 日, 其数据库已超过 107 万篇论文, 主要涉及计算机科学领域, 主题包括: 互联网分析与检索、数字图书馆与引文索引、机器学习、神经网络、语音识别、人脸识别、元搜索引擎、音频/音乐等。

Citeseer 是最早的也是最有影响的计算机领域的 OA 资源, 目前国内外的论文平台都受到其影响, 例如, 其首先推出的研究热度、被引论文、同引论文等已经在国内知网等平台得到了应用。

Citeseer 收录的论文来源大多是一些国际会议文集, 还有一些知名期刊的部分论文, 包括 IEEE 的很多期刊。这些论文大多来源于作者和机构的自存档。Citeseer 自称具有全球最大的科学文献免费全文索引。广西大学图书馆黄日昆老师曾受理过一用户索取学术全文请求。用户在 INSPEC 和 Ei Compendex 文摘型数据库查到 14 篇有关计算机学科方面的文摘, 多为会议文献。经检索广西大学购买的几个国外全文数据库与馆藏书刊, 只获得其中的一篇期刊全文, 使用 Citeseer 检索, 检索出其中的 12 篇并成功下载全文^[3]。

Citeseer 核心技术是自动引文索引和机器自动识别。

2.2 Highwire

网址: <http://highwire.stanford.edu/>

由美国斯坦福大学图书馆建设^[4~6], 目前收录了 1154 种经过同行评议的高影响力期刊, 收录论文 476 万篇, 其中 188 万篇是免费的 (2001 年 5 月有论文 99 万篇, 免费论文 25 万) 篇^[7]。主要集中于生物学、人类学、医学、物理学和社会科学, 但是目前也收录了很多电子信息领域的论文。

其收录论文中 40% 是免费 OA 论文。免费期刊在期刊列表中有 Free ISSUES、Free TRIAL 和 Free SITE 三种标记之一。Free ISSUES 表示该刊在一定时限内 (过刊) 可免费提供全文, Free TRIAL 表示在一个特定时间前可免费提供全文, Free SITE 表示该刊全部论文均可免费提供全文。无标记的表示不免费提供全文, 对一般读者只提供题录或文摘。

免费全文多为发表后延迟 1 年 (其中不乏权威期刊), 少数为 2 年, 也有 6 个月的, 甚至还有 2 个月的, 实施 OA。延迟 6 个月的与国内可得到的印刷型非原版期刊时滞上已无明显时差, 延迟 2 个月则与国内用美元购买的印刷版原版期刊时滞相近。

此外, 其主页还提供了许多科技期刊的链接, 很有实用价值。目前, 已在全世界建立了一千多个影像站点, 中国的影像站点建在了清华大学图书馆。

<http://zgkjyqkj.periodicals.net.cn/>

2.3 DOAJ

网址: <http://www.doaj.org>

DOAJ 是一个重要的开放期刊平台, 由瑞典 Lund University 建立^[8,9], 截至 2008 年 6 月 8 日, 共收录了 3414 种期刊, 其中 1161 种可以提供全文检索和全文, 有全文 18 万篇^[10], 期刊的品种和论文数量仍在不断增加。其中有计算机科学、电子、工业工程等与电子信息相关学科。据消息称, 目前已和中国科技信息研究所进行国际合作, 探讨收录中国期刊的可能。

2.4 Scirus

网址: <http://www.scirust.com>

Scirus 由 Elsevier 出版社开发, 目前已将 3 亿个科技信息网页编入索引中。除此之外, 它还包括 MEDLINE 文摘、ScienceDirect 全文、USPTO 的专利、Beilstein 文摘、IDEAL 全文、NASA 技术报告、源于 E-Print ArXiv 的电子文献、源于 CogPrints 的电子文献、来自 Mathematics Preprint Server 的预印本、源于 BioMed Central 的全文、源于 Neuroscion 的新闻、来自 Chemistry Preprint Server 的预印本。

Scirus 为全学科科技文献平台, 其索引每月更新, 可检索 1973 年以来发表的文献。Scirus 提供的期刊资源可以免费查看题录和文摘, 其中大多数为收费论文, 但也有部分免费论文。最近几年, Scirus 荣获了多项国际大奖, 包括最佳专业搜索引擎的搜索引擎守望奖 (the Search Engine Watch award for Best Specialty Search Engine) 和最佳搜索引擎/目录指南的网络奖 (the Web Award for Best Search Engine or Directory)。同时, Elsevier 公司还提供预印本服务 Sciencedirect (<http://www.sciencedirect.com/preprintarchive>), 其预印本数据库只有 3 种, 数学、化学、计算机, 这是免费 OA 的。由于该公司出版的期刊大多被 SCI 收录, 因此其预印本论文的质量很高。该预印本数据库需要注册和登陆 (免费)。

2.5 Open J-Gate

网址: <http://www.openj-gate.com>

Open J-Gate^[11] 是印度 Informatics (India) Ltd 公司在 2006 年发布的一个 OA 平台, 截至 2008 年 6 月 8 日, 声称收录了 4386 种 OA 期刊, 其中 2367 种经过同行评议, 可提供免费的全文, 每日更新, 目前其官方网站声称收录了超过 100 万篇论文, 还宣称以每年 30 万篇论文的速度在增长, 但是没有看到其收录论文总数的确切数字。

它只提供了期刊列表, 没有提供期刊的分类列表, 旨在建立一个全科的 OA 平台, 期刊来源遍布世界各地。

提供高级查询, 笔者用 Filter 作为查询词, 从标题和关键词中, 查到了 2360 条记录。但其查询出来的全文链接, 多为原期刊的目录, 没有链接到指定的论文, 使用不便。

中国科技期刊研究, 2009, 20(3)

2.6 e-print arXiv

网址: <http://arxiv.org/>

arXiv 是美国国家科学基金会和美国能源部资助,于 1991 年 8 月由美国洛斯阿拉莫斯(Los Alamos)国家实验室建立的电子预印本文献库。主站点设在康奈尔大学,域名为: <http://arxiv.org/>,在世界各地设有 17 个镜像站点,中国的镜像站点设在中国科学院,域名为: <http://cn.arxiv.org/>。

arXiv 收录范围目前包含物理学、数学、非线性科学、计算机科学等四个学科共计 17 万篇预印本文献。收录的论文除作者提交的外,还包括 American Physical Society、Institute of Physics 等提供的 12 种电子期刊全文。数据库的全文文献有多种格式(例如 PS、PDF、DVI 等),需要安装对应的全文浏览器才能阅读,尤其是 PS 格式,要求在 UNIX 系统下安装 PS 格式的浏览软件。

要说明的是,提交到预印本库中的论文均未经过任何审核,实行文责自负,收入该数据库中的论文可以随时受到同行的评论,论文作者也可以对评论进行反驳。论文作者在将论文提交 e-print arXiv 的同时,也可以将论文提交学术期刊正式发表。论文一旦在某种期刊上发表,在 e-print arXiv 的该论文记录中将加入正式发表期刊的有关信息。

由于 arXiv 采取双向交流的方式,即用户不但可通过 WWW 界面或 E-mail 方式检索或获取文献,而且还能随时上传文献,因此 arXiv 的更新频率很高,几乎是每日更新。

2.7 Jstage

网址: <http://info.jstage.jst.go.jp>

Jstage 是“Japan Science and Technology Information Aggregator, Electronic”的缩写,由日本科学技术机构(Japan Science and Technology Agency)主办,提供硬件和软件,1996 年 10 月 1 日开始提供全天候服务。这是一个电子期刊的采编、发布、订阅综合平台,收录了 30 万篇论文。网站上没有看到收录的期刊总数,从期刊名称来看,大多是在日本出版的期刊,领域多是生物学、医学,电子信息的期刊不多,只看到了 *The Journal of the Institute of Electrical Engineers of Japan* 一种。没有查到关于“filter”的论文。

2.8 ScholarGoogle

网址: <http://scholar.google.com/>

Google 是 Google 公司的学术搜索引擎,可以提供来自学术著作出版商、专业性社团、预印本、各大学及其他学术组织的经同行评议的文章、论文、图书的摘要和全文。国外资料部分免费使用,中文资料提供了相关数据库链接,属有偿使用。

推荐一种使用 ScholarGoogle 的办法:从 EI、SCI 等只能查看摘要的二次文献搜索中去查找到有价值的论文,然后到

ScholarGoogle 上去查找免费全文。据笔者的经历,能查到全文的概率在 60% 以上(英文论文)。

3 电子信息领域国内 OA 资源

3.1 中国科技论文在线

网址: <http://www.paper.edu.cn/>

中国科技论文在线是经教育部批准,由教育部科技发展中心主办的科技论文网站。根据文责自负的原则,只要作者所投论文遵守国家相关法律,为学术范围内的讨论,有一定学术水平,且符合中国科技论文在线的基本投稿要求,即可在一周内发表。按自然科学国家标准学科分类与代码分为 39 类专业领域,和电子信息类相关的学科包括:信息科学与系统科学、动力与电气工程、电子通信与自动控制技术、计算机科学技术、航空航天科学技术。发表的论文允许同行评议。

目前一些大学要求博士生毕业必须向中国科技论文在线投稿一篇,因此其更新速度很快。截至 2008 年 6 月 8 日,共收录了首发论文 21990 篇、同行评议论文 9816 篇,优秀学者论文 43117 篇,自荐学者论文 2701 篇,科技期刊论文 64830 篇。

中国科技论文在线可为发表论文的作者提供该论文发表时间的证明,并允许作者同时向其他专业学术期刊投稿,以使科研人员新颖的学术观点、创新思想和技术成果能够尽快对外发布,并保护原创作者的知识产权。

3.2 中国预印本服务系统

网址: <http://prep.istic.ac.cn/eprint/index.jsp>

中国预印本服务系统是由中国科学技术信息研究所与国家科技图书文献中心联合建设的以提供预印本文献资源服务为主要目的的实时学术交流系统,是国家科学技术部科技条件基础平台面上项目的研究成果。该系统由国内预印本服务子系统和国外预印本门户(SINDAP)子系统构成。

国内预印本服务子系统主要收藏的是国内科技工作者自由提交的预印本文章,可以实现二次文献检索、浏览全文、发表评论等功能。从网站访问统计来看,此网站的访问量很小。

据其官方网站介绍,SINDAP 子系统是由中国科学技术信息研究所与丹麦技术知识中心合作开发完成的,实现了全球预印本文献资源的一站式检索。通过 SINDAP 子系统,用户只需输入检索式一次即可对全球知名的 16 个预印本系统进行检索,并可获得相应系统提供的预印本全文。目前,SINDAP 子系统含有预印本二次文献记录约 80 万条。但是笔者进行了测试,输入“Filter”,查询结果为 0。

中国预印本服务系统提供国内科研工作者自由提交的中国科技期刊研究, 2009, 20(3)

科技文章,一般只限于学术性文章。系统的收录范围按学科分为五大类:自然科学、农业科学、医药卫生科学、工程与技术科学、人文与社会科学。

3.3 奇迹文库预印本论文

网址: <http://www.qiji.cn/>

奇迹文库是由几个中国年轻的科学、教育与技术工作者(志愿者)创办的非营利性质的网络服务,主要收录中文科研文章、综述、学位论文、讲义及专著(或其章节)的预印本,同时也收录作者以英文或其他文字写作的资料。收录的学科范围主要包括自然科学(理学、数学、生命科学等)、工程科学与技术(计算机科学、信息处理、材料科学等)、人文与社会科学(艺术、法学、政治、经济、图书情报学等),其他分类(科学随想、毕业论文、热门资料等)。

3.4 开网

网址: <http://www.paperopen.com>

空军工程大学 OA 研究中心开发,2008 年 1 月正式发布。开网得到了陕西省科技规划项目和科技部 863 项目的支持,开通初期收集国内外科技论文 54 万篇,目前主要领域为电子信息,包括了《西安交通大学学报》、《西安电子科技大学学报》、《软件学报》等百余家中期刊。到 2008 年 12 月,其论文总数已达到 187 万篇。开网收集的论文全部可以给出全文链接。

另外,由于本文作者担任了开网的项目负责人,对其长远规划比较清楚,该项目预计到 2009 年底,OA 论文数量要达到 500 万篇。

3.5 Socolar

网址: <http://www.socolar.com/>

该平台由中国教育图书进出口公司建设,截至 2008 年 6 月 9 日该平台宣称收录了 OA 期刊 7066 种,包含文章 10972856 篇;OA 仓储 959 个,文章 3224892 篇。平台收录文章总计 14197748 篇。笔者查询了一些国内的 OA 期刊,如《软件学报》、《西安交通大学学报》等,均未发现。笔者输入关键词“Filter”,从摘要和标题中查得 4143 个记录,其查询速度也很快。

这是目前唯一宣称收录了 OA 期刊和 OA 仓储的平台,也是目前收录论文数最多的平台。

4 免费课件资源 MIT's OpenCourseWare

网址: <http://www.core.org.cn/OcwWeb/>

美国麻省理工学院的免费教学课件平台,无需注册,有些已经翻译为中文。教案包括了课程教学大纲、教学日程、课堂讲义、作业及其解答、实验指导书、试卷范例和相关资源

(参考文献)等,有的还提供音频和视频。其中电子工程和计算机科学的教案有 38 个。

教案和其他教学资源的 OA 是 2007 年新出现的一个发展新趋势^[12]。

5 结束语

结果表明,在计算机领域,CiteSeer 最有影响,Scirus 的论文整体水平较高,OpenJ-Gate 收集的 OA 期刊和论文数更多;国内中国科技论文在线的更新率较高,Socolar 收录的论文数较多,开网则目前集中于电子信息领域。

目前,互联网上散存着很多有价值的 OA 资源,如何进行分类整理,采用更人性化的搜索算法,让读者在最短的时间内,浏览最少的信息就能得到需要的知识,一直是科技传播的目标。OA 为科技进步的传播提供了一种新的思路,作为发展中国家如何利用这些 OA 资源来提高我国的科技创新能力是一个值得关注的课题。

参考文献

- 1 李若溪,黄颖,欧红叶,游中胜,Fytton Rowland. 国际学术出版开放式访问(OA): I-实践与前沿问题研究进展. 编辑学报, 2006, 18(3): 237-240
- 2 Cl Giles, K Bollacker S L. CiteSeer: An Automatic Citation Indexing System. New York: 1998
- 3 黄日昆. 网络引文搜索引擎 CiteSeer 评析. 情报杂志, 2004, (6): 31-33
- 4 John Sack. HighWire Press: ten years of publisher-driven innovation. *Learned Publishing*, 2005, 18: 131-142
- 5 魏一生,李双官,金静宜等. 免费提供论文全文的网站 HighWire Press. 医学信息, 2002, (3): 3-5
- 6 张晓燕,何玮等. 新版 Highwire Press 及其使用. 中华医学图书情报杂志, 2003, 12(2): 47-49
- 7 Stanford University. Highwire Press, [2008-06-08] [http://highwire.stanford.edu/\[EB/OL\]](http://highwire.stanford.edu/[EB/OL]).
- 8 Morrison H. Directory of Open Access Journals (DOAJ). The Charleston Advisor. January 2008 <http://eprints.rclis.org/archive/00012501/01/DOAJreview.pdf>
- 9 谭从蓉,阮春英. 关于 DOAJ 的概况、使用和影响分析. 图书馆建设, 2006, (1): 33-35
- 10 Lund University Libraries. DOAJ. [2008-06-08] [http://www.doaj.org/\[EB/OL\]](http://www.doaj.org/[EB/OL]).
- 11 Informatics (India) Ltd. Openj-gate. [2008-06-08] [http://www.openj-gate.com/\[EB/OL\]](http://www.openj-gate.com/[EB/OL]).
- 12 Peter Suber. 开放获取 2007 年进展 [EB/OL] [2008-06-08] <http://www.libspace.org/archives/open-access-in-2007.html> 2008-04-09