

引文搜索引擎CiteSeerX调查评析

□ 刘莎 / 武汉大学信息管理学院 武汉 430072

摘要: CiteSeerX的兴起与开放获取、电子科研和引文标引系统有着直接的关系。文章通过分析CiteSeerX的搜索原理和重要功能,与同为学术搜索引擎的Google Scholar进行对比,明确其优势所在,指出它的不足,给出相关改进建议。

关键词: 引文搜索引擎, 开放获取, 文献检索, 数据库, 知识库

DOI: 10.3772/j.issn.1673—2286.2011.12.009

CiteSeer是NEC研究院(现为NEC实验室)在自动引文索引(Autonomous Citation Indexing, ACI)机制的基础上,由三位研究人员Steve Lawrence、Lee Giles和Kurt Bollacker研制开发建设的一个学术论文数字图书馆。这个引文索引系统提供了一种通过引文链接来检索相关文献的方法,它的目标是从多个方面促进学术文献的传播和反馈^[1]。

1 CiteSeerX的兴起与发展

CiteSeerX作为CiteSeer的换代产品,它的兴起与发展深受时代因素的影响,包括以下三个方面:

(1) 开放获取

开放获取作为一项交流学术信息和彰显学术价值的国际运动,自20世纪90年代兴起以来一直是业界研究的热点之一。它通过把同行评议过的科学论文或学术文献放到互联网上供用户免费使用,而不需考虑版权或注册的限制,旨在打破学术研究的人为壁垒,促进学术信息的广泛交流,提升科学研究的公共利用程度,保障科学信息的长期保存,提高科学研究的效率^[2]。CiteSeerX秉承开放获取的理念,提供免费的全文下载服务,支持科研人员上传自己发表与未发表的研究论文,成为一个学术科研的公共平台。

(2) 电子科研

传统的科学社区都是围绕智力资源如图书、特殊装置而形成的。过去,形成和维持一个科学社区最大的障碍就是成员们分散居住在各地,彼此共享信息不便。21世纪,面对面的科研合作已经逐步发展为网络交流

的方式。这股在线科研协作的浪潮被命名为E-science(电子科研)或者E-research(电子研究),指的是通过互联网技术进行的分布式和大规模科研协作,各个学科的科学家和学者利用这些服务可以进行高效率的网络协同工作^[3]。CiteSeerX作为一个电子科研的平台,协助网络实验室的发展,允许科学家们分享他们的关键智力资源,而不受时间地点的限制。

(3) 引文标引系统

科学文献的开放极大地提高了科学信息交流的效率和使用频率。但是,散落在网络上的研究论文缺乏有效的检索入口,研究人员查找论文费时费力。自动引文标引系统ACI及在此机制上开发的CiteSeerX正是致力于解决这一问题的新一代引文搜索引擎。

CiteSeerX坚定地践行开放获取的理念,以技术之翼辅助科研之体,以求打破这种学术壁垒。CiteSeer创建于1997-1998年,是一个搜索引擎和计算机信息科学领域的数字图书馆,向公众免费提供约700,000篇学术论文全文和约10,000,000的引文信息。

在CiteSeer创建之后的10年间,研发人员不断对CiteSeer运行中暴露的问题和用户的反馈建议进行分析,并由美国国家科学基金会(National Science Foundation)和微软研究院(Microsoft Research)资助,为该搜索引擎重新设计了系统结构和数据模型,CiteSeerX于2007年投入运行。

2 CiteSeerX的搜索原理与功能简介

2.1 CiteSeerX搜索原理^[4]

自动引文标引系统ACI可以自动标引电子格式的文献生成引文索引,CiteSeerX利用ACI技术可以很好地实现这一过程,其步骤如下:

(1) 采集文献

ACI通过搜索Web,监测电子邮件列表、新闻组或者直接与出版商联系获取文献。部分熟悉ACI系统的科研人员将自己的研究论文直接提交给系统,从而被标引入库。CiteSeerX又通过与部分上网期刊签订互惠协议,采集其文献,从而使用户直接通过CiteSeerX查阅期刊主页,最终提高期刊的访问流量和订购率。

(2) 文献处理

CiteSeerX将下载的Postscript和PDF文件转换成文本格式,判断其是否为研究论文。ACI根据参考文献列表的格式,包括引文标识符、缩排等特征将引文分割,再抽取每一条引文款目的可检字段,进而制作成引文列表。

(3) 查询和浏览

CiteSeerX的查询遵循以下操作规则:①支持布尔逻辑运算;②支持近义词运算;③不支持词组的精确匹配;④词干法有效,通配符语法无效。

(4) 相关文献算法

CiteSeerX采用三种算法计算文献相关度:TFIDF算法、LikeIt算法、CCIDF算法。

2.2 CiteSeerX功能简介

(1) 关键词检索

CiteSeerX首页的检索框默认是对篇名、作者、摘要、文本内容进行检索。对于未加双引号的检索词默认为模糊匹配检索,只要有一个词匹配成功则返回文献标引信息,按相关度进行排列;也可在首页的单一检索框自行构造组合检索式,检索完整字句用双引号。

(2) 作者检索

CiteSeerX支持布尔运算和近义词运算。如检索Jon Jablonski的文章,切换检索专题到authors,可同时运用下面的Disambiguated Search复选框,帮助消除歧义。如果检索词为全名Jon Jablonski,则不能得到所有该作者的论文,只有那些名字是“Jon”未经缩写的作者的文章能够被检索出来。在作者检索中,只用姓或者名字缩写加姓进行检索,会得到比较好的效果。

(3) 引文检索

CiteSeerX的引文和全文检索可同时进行,只需在

普通检索或高级检索界面勾选“Include Citations”即可。在高级检索界面中可对检索范围和检索结果排序规则作出规定,如对文献引文的最小数目作出限定,则引文数少于规定数目的文章将不被检出。

(4) 引文分析

CiteSeerX可对其数据库中的文后引文进行分析处理,它将“Related Documents”标签分为两个列表:活跃的参考文献和同被引文献。这两项功能可以帮助用户分析引用率高的文献的分布情况和被引现状,提高资源获取的针对性。“Version History”显示该文献的元数据。由于CiteSeerX支持通过验证授权的用户对元数据进行修改,而不删除或覆盖以前的记录,用户可看到文献元数据的所有历史版本。用户的群体参与,使得CiteSeerX数据库中元数据不断地修正与完善。

对于引用与被引用的分析,CiteSeerX采用了ACI技术,可自动识别参考文献标识,因而能准确地分析出文后参考文献总数。但对于被引情况,CiteSeerX只能在自己收集的文献数据库中进行统计,结果造成与实际被引数量的不一致。图1中,CiteSeerX给出22条文后参考文献数目,经查证,该数目准确无误。但对于引用这篇文献的数量统计则与从Google Scholar检索出的1489存在较大差距,再次说明Google Scholar的数据收录范围远大于CiteSeerX,前者将非正式发表的文献也统计在内。为进一步查证,笔者又进入发表这篇文章的正式出版物ACM进行检索,查证结果如图3。ACM显示该文献的被引数据为470条,与CiteSeerX较为接近,说明CiteSeerX的收录范围应与ACM相差较小。多出的数据一方面是因为ACM作为正式出版物从收稿到出版存在时滞,另一方面是因为CiteSeerX的实时收录功能和科研人员参与贡献的未发表文献。

CiteSeerX还在首页为用户提供众多引用分析排行链接,用户可看到通过引用率分析得出的最高引用率文章排行、最高引用率引文排行、最高引用率作者排行、综合影响因子排行。这项功能对于跟踪学科前沿十分有益。

3 CiteSeerX与Google Scholar综合对比

Google Scholar 是Google公司宏大的网上图书馆计划的一部分,其服务对象主要是科学家和各类学术研究人员。一方面它过滤了普通网络搜索引擎中大量对学术人士无用的信息,另一方面,Google 与众多学术出版商合作,加入许多普通搜索无法获取的加密内



图1 CiteSeerX检索实例引用与被引情况

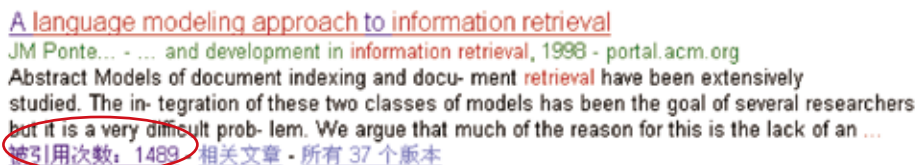


图2 Google Scholar检索实例被引情况



图3 ACM检索实例被引情况

容,并要求合作商免费提供文献的文摘。可见,Google Scholar和CiteSeerX在功能覆盖、用户群定位、检索原理等方面存在诸多相似之处,对比这两种检索系统,揭示学术搜索的一般特点和规律,可以帮助广大用户在海量网络信息资源中更加高效地获取所需信息。综合二者的特点,主要从以下几个方面对二者进行比较:

(1) 检索界面

CiteSeerX界面紧凑,其列表显示方式符合用户使用习惯;功能按钮标记全面清晰。由于CiteSeer的检索针对内部网站,响应速度快,且检索得到的信息经过筛选,大大减少冗余。

Google Scholar建立在Google技术基础上,是直接面向科研检索需要的学术资源网络检索工具。其检索结果虽以列表显示摘要,但检索结果未经整合,且网页的动态性会导致死链和网页内容的变化,影响用户的使用。显而易见,CiteSeerX检索界面的设计更加符合用户习惯。

(2) 检索结果

CiteSeerX可以检索到文献的引用和被引情况,参考文献按照引文被引次数降序排列。通过参考文献题名链接获得该被引文献的摘要和引文,深蓝色题名表示有全文,浅蓝色则没有全文收录;另一方面,通过循环搜索法,可以进行从任一篇文献为入口的越查越新和越查越旧的引文检索。CiteSeerX的检索结果只著录题名、作者和出版时间,一般可在一页列出所有参考文献。

Google Scholar的检索结果也给出了文献的被引次数和相关文献,标记一篇文献的历史改动版本,题名后还明确标示是否可以下载全文PDF。但是通过实地使用发现,Google Scholar由于文献的标引著录信息太长,不能在一页将一篇文献的所有参考文献列全,在网页的频繁切换中,用户不仅不能很快地找到所需信息,还易造成“检索疲劳”、“网页迷航”,甚至放弃检索。但值得借鉴的是Google Scholar链接指示目标网页,可以有效规避数字时代的知识产权纠纷。

(3) 检索内容

CiteSeerX专注于计算机科学和信息科学领域，检索内容专一，一方面使得其专业性强，另一方面又导致其检索内容面窄。尽管Google也对CiteSeerX做了标引，但是后者立足于更加专业的搜索，作为一个全文检索的接口，它允许按照文件类型、引用次数、所在领域检索，但目前仍做不到通用搜索引擎那样可以对任何领域进行搜索。

Google Scholar的选择内容可以更具有广泛性和多元性，可以整合到数据库所没有收录到的一些信息。但Google Scholar注重全面性、全球性，因而在专业性要求上势必产生此消彼长的连带影响，这也体现了查全率与查准率矛盾性的一面。虽然二者都是利用巡视软件和自动索引文档系统对信息资源进行整合和组织，但Google Scholar强大的软硬件优势使得它的检索结果

具有较高的时效性。

(4) 文献排序

CiteSeerX为用户提供的文献排列依据有引用次数、相关度、日期(降序)、日期(升序)、最新排行等，可以满足不同用户的查找需要。Google学术搜索不以引文检索作为直接目的，只在检索专题文献的基础上提供了“文献被引频次”的附加服务。由于Google Scholar检索结果按论文被引用频次排列，排在列表上方的文献往往发表年份较早，从而使用户难以浏览到最新文献。且Google Scholar检索结果排序不全基于被引用频次，因而检索结果排列标准不一致，影响用户对检索结果的判断^[5]。

综合上面的分析，现将两者综合对比的内容归结如表1。

通过以上四个方面的比较，可以发现CiteSeerX在

表1 CiteSeerX与Google Scholar对比简表

检测系统 评价标准	CiteSeerX	Google Scholar
检索界面	简洁清晰，功能齐全	简单清楚，功能展示单一
检索结果	列表显示，网站内操作	页面冗繁，链接到网站外
检索内容	专业性强，主要针对计算机科学与信息科学	通用广泛，任何领域的历史与概况均可检索
文献排序	引用次数、相关度、日期(降序)、日期(升序)、最新排行	并不严格按照文献被引频次排列，结合相关度

检索界面、检索结果、文献排序上优于Google Scholar，但在检索内容上需要借鉴Google Scholar的广泛性和对知识产权处理的先进思想与技术，才能在发挥既有优势的基础上扩大影响力，吸引更多用户。

4 CiteSeerX系统评析

4.1 CiteSeerX的优势

(1) 更新快，对科学文献及其影响的揭示全面

传统引文索引如SCI、EI、ISTP等都为正式出版物，文献的发表周期过长，从研究人员投稿到索引编制耗时数月到一年不等。而CiteSeerX作为一个网络实时搜索引擎工具，充分发挥了Web的优势，将收录范围扩展到开放环境，能及时发现网上新出现的学术文献，并进行自动标引入库，即时反映所有网上学术文献类型的引证脉络，从而更新更快地反映引文的信息影响^[6]。

(2) 更科学的相关文献检索

相关文献检索是研究人员了解其研究主题发展脉络、最新动向不可或缺的途径。不同于Web of Knowledge采用的引文耦合的方法，CiteSeerX采用计算机算法，并提供同被引检索相关文献。同被引技术是文献耦合概念的创新和逆向思维的发展，加之近年来信息可视化及网络寻址定标技术的发展使同被引技术得到了新的发展。因此，CiteSeerX要比Web of Knowledge更为科学，并且更具发展潜力^[7]。

(3) 提供友好的学术交流平台

CiteSeerX充分发挥网络大众的力量，它所提供的附加网络服务能很好地促进科研人员就专业内问题进行正式或非正式的评论，发表意见和最新的研究成果，提供了一个宽松、自由的交流平台。Feedback栏目还允许用户针对CiteSeerX的服务或任何方面提出建议，以促进网站的发展和完善，这在传统封闭式的商业数据库中是鲜见的。

(4) 免费服务促进学术交流

传统引文数据库的价格非常昂贵,导致大量学术机构购买无力。基于开放获取的理念,CiteSeerX所提供的文献检索、作者检索、表格检索、全文下载、社交平台、公告板、引文统计排行、标签评注、文献版本改动与观测、个性化收藏等服务都是免费的,这将促进引文分析思想的普及,其所建立的引文索引机制也将为更多的人所采用,如南开普顿大学的Citebase,从而有利于学术论文的传播。

4.2 CiteSeerX的不足与改进建议

(1) CiteSeerX收录内容还需拓宽

较之商业数据库和Google Scholar, CiteSeerX收录的学科范围窄,还不能提供综合性学科内容的引文索引。在Google Scholar中输入某一研究领域的关键词,就能从返回结果中分析得知该研究领域的重要著作、论文、专著等,以及从事该领域研究的主要学者,通过这些著作的年代分布、相互引用情况等,大致了解该研究领域的历史和概况。而CiteSeerX也不能做到这一点。因此,CiteSeerX有必要拓宽文献信息的收录领域,加大分析的力度,多维度揭示学科历史与发展现状。

(2) 引文标注不够全面准确

由于完全依靠机器自动操作,目前CiteSeerX还存在不能准确分辨子字段、无法区分同名作者、引文在文

献中若无标识则不会被标引等情况。因此加强抽取引文信息的算法、完善规范数据库的建设以区分同名作者都应是CiteSeerX下一步的着力点所在。

(3) 支撑技术有待完善

CiteSeerX将会朝着模块化、可扩展、自管理的方向发展,还将探讨如何利用新的智能数据挖掘和信息提取算法来改进新的索引,增强文献获取功能,拓展和加强文献、新数据、元数据资源、活跃的镜像、网络服务的自动与协作化获取的功能^[8]。

虽然处于初生期的CiteSeerX还存在许多不足和需要改进的地方,但它是在开放存取、促进学术信息资源传播的机制下诞生的新一代引文搜索引擎,反映出开放环境下引文标引与检索的许多新特征,这些特征也必将对传统引文数据库的发展起到借鉴作用。随着开放获取运动的蓬勃发展和ACI技术的不断完善,以CiteSeerX为代表的开放环境下的引文搜索引擎必将不断为引文标引与检索注入新的活力。

5 总结

CiteSeerX被誉为全球最大的科学文献免费全文索引搜索引擎,在开放存取环境下,促进了学术信息的交流与共享。作为搜索引擎,CiteSeerX在这样的背景下产生,其检索功能在很多方面值得借鉴,为搜索引擎的搜索机制和发展提供参考。

参考文献

- [1] CiteSeer [EB/OL]. [2011-05-08]. http://www.worldlingo.com/ma/enwiki/zh_cn/CiteSeer.
- [2] 李麟. 推动21世纪的学术信息交流: 机构知识库在开放获取运动中的作用[R]. 香港: 香港大学, 2007.
- [3] FAROOQ U, GANOE C H, CARROLL J M. Designing for e-science: Requirements gathering for collaboration in CiteSeer [J]. International Journal of Human-Computer Studies, 2009(67): 297-312.
- [4] 宋歌. 引文搜索引擎CiteSeerX设计原理及检索[J]. 中国索引, 2008(3).
- [5] 夏训明. Google Scholar的利用与评价[J]. 中国索引, 2006(5): 97-99.
- [6] 黄日昆. 网络引文搜索引擎CiteSeer评析[J]. 情报杂志, 2004(6): 77-79.
- [7] 宋丽萍. 基于Web的学术信息资源引文索引与分析体系[J]. 情报理论与实践, 2005, 28(3): 307-310.
- [8] GILES C L. The Future of CiteSeer: CiteSeerX [C]// Knowledge discovery in databases: PKDD, Berlin, 2006.

作者简介

刘莎 (1987-), 女, 武汉大学信息管理学院2010级图书馆学硕士研究生, 研究方向为信息组织、主题法。E-mail: lsha2007@163.com

Analysis of the Citation Search Engine CiteSeerX

Liu Sha / Wuhan University School of Information Management, Wuhan, 430072

Abstract: The origin of CiteSeerX has direct relationship with Open Access, E-Science and ACI. This article analyzed the searching principle and important function of CiteSeerX, and compared it with academic search engine Google Scholar, and then pointed out its advantages and disadvantages, and finally came up with some relevant recommendations for improvement.

Keywords: Citation search engine, Open access, Document retrieval, Database, Knowledge repository

(收稿日期: 2011-08-27)

“SIMPLE Science” ——基于图像的中小学简化科学教育数字图书馆

□ 刘燕权 / 美国南康涅狄格州立大学 纽黑文市 06515

王群 / 北京师范大学信息管理系 北京 100875

摘要: SIMPLE Science是一个通过利用图像处理及分析技术 (IPA) 辅助青少年学习的数字图书馆, 通过提供图片信息、教学计划、课程活动使得图像处理和分析能够作为一种学习工具得以使用, 同时也为教育工作者在教学过程中使用图像信息提供便利条件。文章对该项目的建设及现状进行了综合性的评析, 包括项目概述、数字资源及组织、技术特征、服务特点等。

关键词: SIMPLE Science, 图像处理技术, 数字图书馆, 美国国家科学数字图书馆 (NSDL), 美国国家科学基金会 (NSF), 中小学教育, 科普

DOI: 10.3772/j.issn.1673—2286.2011.12.010

1 项目概述

“SIMPLE”是“简化图像管理及处理学习环境”(Simplified Image Management and Processing Learning Environment for Science)的首字母缩写。SIMPLE Science是一个通过利用图像处理及分析技术(image processing and analysis, IPA)辅助青少年学习的数字图书馆项目, 主要致力于使IPA在中小学教育中简易可行, 使其成为一种易于使用并为主流教育和公众接受的学习工具, 同时方便教育工作者对成像数据进行扩展和更新, 并为他们提供一种通过图像数据进行教学的具有国家标准化的教育学构架。

SIMPLE Science使用的教育标准包括国家科学教育标准(National Science Education Standards, NSES)、数学教师全

国理事会(National Council of Teachers of Mathematics, NCTM)以及国家教育技术标准(National Educational Technology Standards, NETS)。为方便查看, SIMPLE Science在每一课时前都列出该门课程应达到的相关标准。此外, SIMPLE Science工作人员在主页上也提供了其他标准的链接, 方便那些有特殊需求的教育人员(Science Approach, 2010)。

SIMPLE Science是美国国家科学基金会(National Science Foundation, NSF)资助的数字图书馆项目之一, 也是美国国家科学数字图书馆(National Science Digital Library, NSDL)系列工程之一。NSF于2004年1月1日授予图像处理教育中心(Center of Image Processing in Education, CIPE)一项396,203,000美元拨款用于此项目的建设, 即为中小学教育开发基

于图像的教育学习工具。此外, 该项目也得到地理系统教育数字图书馆(Digital Library for Earth System Education, DLESE)的赞助。SIMPLE Science致力于“消除在中小学教育中使用图像处理和分析(image processing and analysis, IPA)技术中的各项障碍”。项目初期定为两年(<http://www.simplescience.org/>)。根据地理系统教育数字图书馆(DLESE)网站提出的项目工作范围, SIMPLE Science的国家资助在2006年停止后, 图像处理教育中心(CIPE)继续维护和更新网站, 确保期资源的完整性、项目的科学准确性、教学方法的高效性、文件资料的整合性、用户使用的便捷性以及各类馆藏数字资源的稳健性。

2006年, 图像处理教育中心(CIPE)董事会通过一项决议, SIMPLE Science作为国家项目定

于2007年9月终止。同时董事会授权将其知识产权转交给科学方法公司 (Science Approach, LLC)。科学方法公司继承CIPE的责任推进可视化技术在教育领域的使用,并通过CIPE继续免费提供SIMPLE Science数字图书馆中的课程。同样,作为CIPE的合作者, SIMPLE Science 网站上的所有课程也包含在DLESE数字图书馆的数据库中。

图像处理教育中心 (CIPE) 是一个建立于1992年的非盈利性组织,其目标是促进计算机辅助可视化应用于探究性学习中。CIPE一直致力于促进“计算机辅助可视化作为应用于基于探究性学习的工具”

(CIPE, 2010)。为了支持这项任务, CIPE开发了教学材料,并组织研讨会,即使用数字图像分析和地理信息系统技术作为平台应用于科学、数学和技术领域的教学。

SIMPLE Science传承了CIPE的任务目标,为中小学教育及公共大众提供图像处理和析的服务。正如美国国家科学教师协会所评价的那样,“该网站已成为一个将资源集合使用的工具,为从事初中、高中教育的工作者准备课程资料和讲授课业提供了良好的帮助,成为了一个为教师教学和课程发展所使用的资源整合地。同时它从学生学习的角度出发,为学生提供了动手实践的虚拟学习场所”。

(SIMPLE, 2010)。不过SIMPLE Science中提出的任何观点、意见、发现、结论等仅代表网站参与者个人观点。

项目主页见图1。

SIMPLE Science使得图像处理和析变得更易使用,并通过提供图像数据、课程计划、相关活动

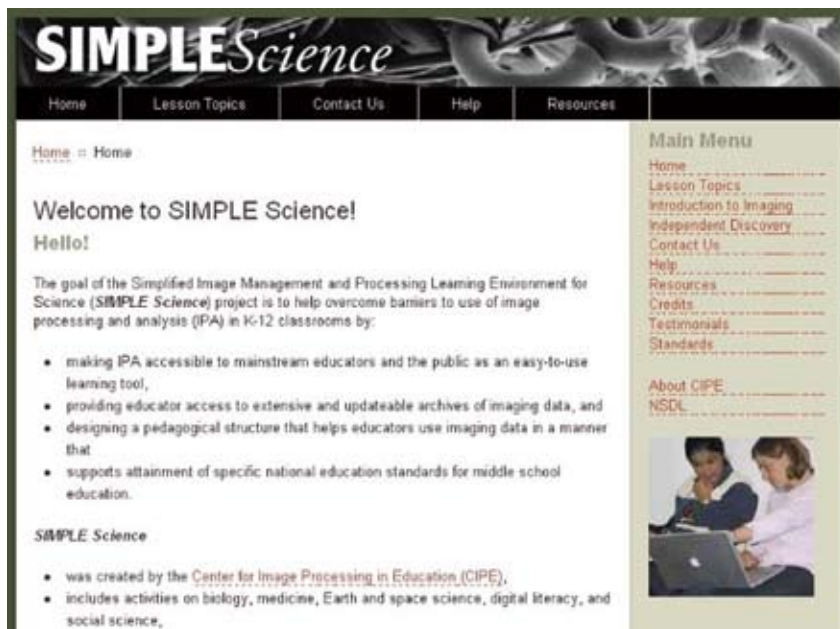


图1 SIMPLE Science项目主页

以及其他工具使其成为一种学习工具,同时借此帮助老师们能够在教学的过程中自如地使用图像数据。项目提供的资源涵盖了包括生物科学、医药卫生、地球空间科学、数字素养以及社会科学在内的主题。

2 项目资源及其组织

SIMPLE Science数字图书馆项目的意义在于为教师和学生提供更多关于成像科学方面的课外优质学习辅助资料的途径,同时多样化的资料格式满足了学习者的个性化需求。作为美国国家科学数字图书馆项目之一, SIMPLE Science在生命科学、医药卫生、物理学、工程学、数字素养以及地球空间学方面丰富了国家科学数字图书馆K-12阶段的教学资源,以多媒体的方式吸引了大量的教育工作者和学习者,并激发了他们的教学和学习积极性。

SIMPLE Science所有资源都以学生为本,为教育工作者服务,

网页上无任何广告及商业链接。接受家庭教育或者在校外接受非正式教育的学习者也可以免费使用这些资源作为学习辅助材料。教师们可以通过SIMPLE Science提供的两个教学软件Wiggins和McTighe (1998) 开发丰富理解力培养与课程设计的教学模板,以及“深入探索”主题下的图片文档,以丰富完善自己的课程。同时网站具有在线设计课程功能,方便教师随时修改、不断完善课程内容。

2.1 资源范围及种类

SIMPLE Science馆藏由六个课程主题组成: 地球空间科学、生物、医药、数字素养、数学、社会科学。每一个主题都有一系列条目,所藏文件包括视频文档、PDF文档、X射线照片或者图像文档。

SIMPLE Science资源按学科分类,所涵括的学科以及该门类下的主要内容包括:

(1) 地球与空间科学: 初级的地球科学领域中的案例学习包括雪与冰、飓风、臭氧层空洞与地球大小。

(2) 生物学: 测量温度、物体运动、与动物互动等。

(3) 医药学: 无重力状态下对物体的影响、诊断肺部疾病等。

(4) 数字素养: 将旧照片翻新、计算机侦察技术等。

(5) 数学: 坡路比率、圆周率分割、测量昆虫体长等。

(6) 社会科学: 利用计算机分析法医图片、为残疾人测量坡路比率、测量运动物体等。

2.2 站点主要区域

在主页以及其他各个页面上, SIMPLE Science提供了五个主要区域:

(1) 图像介绍 (Introduction to Imaging): 这里提供一个关于成像科学的介绍, 用户可以通过一个15分钟的视频演示来发现图像处理和分析技术 (IPA) 是如何应用不同的科学领域的, 例如医药、航天以及艺术等;

(2) 独立探索 (Independent Discovery): 包括四个按兴趣划分的类别: 独立研究、课程建设、课程帮助以及图像图书馆。这一部分将图像思维和数字化分析融合在一起作为一个工具, 帮助用户学习科学、技术、数学以及其他学科知识;

(3) 课程主题 (Lesson Topics): 这一部分涵盖了地球科学、生物、医药、数字素养、数学和社会科学, 开设内容丰富多彩, 例如“行星轨道”、“运动中的动物”、“太空失重”、“圆周率分割”。课程中的用词都十分生动有

趣。每部分结构都配以颜色醒目的图片、动画、声音以及与教学材料互动的链接;

(4) 实证 (Testimonials): 记载了来自美国国家科学教师协会、俄勒冈州、亚利桑那州等不同区域、不同学校的教师代表对网站使用的评价;

(5) 资源 (Resources): 为用户提供相关资源链接。目前网站所提供的链接有SIMPLE Science所从属的科学方法公司 (Science Approach, <http://www.science-approach.com/>) 以及图像处理教育中心 (Center of Image Processing in Education, <http://www.airedesign.com/CIPESite/>)。同时, 网站也提供了一些来自NSDL和DLESE的网站和文档的链接。

2.3 数据来源及元数据

SIMPLE Science资源主要来源于地理系统教育数字图书馆 (DLESE)、国家科学数字图书馆 (NSDL) 以及图像处理教育中心 (CIPE)。多媒体资料丰富多彩, 包括视频、网络互动、照片、动画、图片、图示、音频等。同时SIMPLE Science也是重要的信息交流平台, 将散落的资源整合后提供给用户。

根据DLESE网站给出的项目工作范围, “元数据记录都是手动输入的, 而且在提交前都经过了再次检验”。也就是说, “由于SIMPLE Science数据库中的所有资源都是可以直接修改的, CIPE的工作人员会不断检验以确保资源和元数据的准确性”。同样, 虽然CIPE持有数字图书馆元数据的版权, 但DLESE持有“调整、改变格式以及重新分配这些元数据以运行DLESE

系统和服务”的权利 (Science Approach, 2010)。

SIMPLE Science的最大特点就在于图像处理和分析技术的应用, 因此除了普通的文本信息外, 网站最大限度地使用了图片对主题进行说明补充, 使用户得到更加直观的概念。图片都以JPEG格式存储, 并附有文本标签。当用户鼠标停留在图片上, 标签即可显现, 易于用户理解。

2.4 信息组织

SIMPLE Science数字图书馆分为三层结构, 以此来帮助学生通过利用图像处理和分析 (IPA) 进行基本的探索及在各个科学领域的利用。

第一层由一些展示图像研究的案例组成。学习教程在指导学生学习案例的过程中指导学生模仿成像科学家们进行研究, 并解释成像科学领域中的一些基本概念。这一部分的副标题引导访问者“探索科学家们是如何将成像技术应用于各个领域的” (SIMPLE Science)。

第二层提供了一些来自图像处理教育中心 (CIPE) 的活动, 这些活动均基于科学指导的研究方法, 为学生更好地学习而设计。

第三层是一个“独立探索链接”, 它提供了独立的课程, 同时为进一步的研究和探索提供相关图片信息。用户可以通过下载获取这一部分的资料, 例如SIMPLE图像档案、科学实验报表格式以及能够深入研究图像处理领域的链接。SIMPLE Science网站也提供可供用户下载的教学模板、课程辅助, 这样教师们就可以在理解Wiggins和McTighe设计模板结构的基础

上完善自己的课程。此外,针对教师,网站专门开设了一个用户通过登录即可体验的在线课程制作,该部分内容可以在课程建设者标题下寻获。

3 技术特征

3.1 技术架构与支持工具

SIMPLE Science使用Java、WebImage软件为课程提供相关的图片和教学互动工具,例如直方图、对比度调节、轴点以及推导数学方程式和探索科学概念的教学工具等。通过SIMPLE Science的在线教程——图像处理入门和分析,用户可以利用该数字图书馆的IPA技术和特定功能部件协调使用来定位课程,测量并调节图片的缩放比例,校正图片以达到规定量值,聚合图片至堆栈并调动图片,修改图片量值甚至重新“创造”图片,加强图片的亮度和对比度,使用剖面图说明单个像素点是如何改变并穿越整个图片的,描绘区域轮廓以进行进一步分析,对待分析的区域在给定范围内设定值等等,同时用户也可以使用化色表改变图像外观。

(1) WebImage技术

SIMPLE Science网站使用WebImage,可被称为是一个基于网页版本的ImageJ,同时也是一个世界范围内广泛使用的图像处理程序。WebImage是一个用来提高、优化和制图的网络图片处理程序,支持GIF、JPEG、以及PNG格式,可以将图片转化成不同网络格式。

(2) 软件和硬件要求

为使用IPA和SIMPLE Science数字图书馆中的技术,网站要求Mac用户在操作系统下安装Safari

2.x,在Windows操作系统下安装IE 7.x。同时在任何操作系统下,都必须安装Java,否则WebImage无法运行。不过IE 6.x和火狐1.5x、2.x也支持WebImage,但是运行不如IE 7.x顺畅。

3.2 界面设计

SIMPLE Science页面设计简洁、便利,独具匠心。

主页包含五个主要区域(课程主题、图像介绍、独立探索、资源、实证等)以及检索菜单在每页的顶部都会出现,方便用户浏览相应资源、进行页面跳转。

同时,值得一提的还有SIMPLE Science的浏览页面,它分为顶部、右侧两个版块。顶部列出五个主要部分(主页、课程主题、联系我们、帮助、资源集合);右侧版块则以主菜单的形式在顶部版块的基础上细分了几个部分,包括科学成像介绍、独立探索、用户评价、使用标准等。这样,用户可以按照习惯选择偏重的浏览方式,非常便利。

4 服务特征

4.1 目标用户

SIMPLE Science的服务群体十分明确,主要针对美国K-12即中小學生及教师服务,不过也面向大众人群,任何对图像处理及分析感兴趣的用戶都可以在网站中找到相关资料。

4.2 服务方式

SIMPLE Science是个功能比

较健全的数字图书馆,可以为用户提供各种服务,以满足不同层次用户的需求。

4.2.1 浏览功能

用户可以通过互动主题实现对站内资源的浏览,可以点击每页顶部的标题,或者侧面的主菜单,也可以通过Back或Prev浏览之前的页面。这里也有网页资源的超链接,例如21世纪中学21世纪中学生涯展、前瞻启动、数字图像科学伙伴等。

考虑到K-12学生特点,SIMPLE Science配以图片丰富的浏览界面,还为用户提供了几种不同的途径来浏览资源,相应的链接全都列在页面左上角的“课程主题”(Lesson Topics),同样的内容也可以在页面右侧的主菜单中进行选择。

- 点击“课程主题”中选择相应课程;

- 通过Back或Prev浏览之前的页面;

- 浏览“资源”(Resources)中提供的SIMPLE Science所从属的科学方法公司(Science Approach)和CIPE链接,用户可深入研究感兴趣的问题;

- 浏览随机页面;

- 对于某个特定页面来说,用户可以通过页面右侧的主菜单中选择跳转到其他资源页面进行浏览。

4.2.2 帮助功能

SIMPLE Science的帮助功能体现在“Help”页面中,它提供各个方面的帮助信息,无论是对于普通用户还是专业学者,都有相应

的条目提供帮助信息。具体包括用户在浏览网站视频、图片时所需下载的程序帮助,软硬件支持。如需更详细的帮助信息,可以通过页面下方“Help Desk”前往Science Approach深入了解。SIMPLE Science提供了一个基本的帮助页面讨论必需的程序和链接,必须下载WebImage使用服务,同时也通过网页提供安装所需程序的指导,这里还有一个指向科学方法网站咨询台的链接,用户可以通过科学方法帮助咨询台来使用检索框或网站的

其他特色主题(SIMPLE, 2010)。

4.2.3 个性化

个性化设置是SIMPLE Science的一大特色。其特色设计之处在于将各栏目置于右侧,同时也在网站上侧也予以相应的显示,内容居中,方便用户随时调整自己所在位置。课程主题中图片、内容相对显示。图文对应,一目了然,如图2所示。

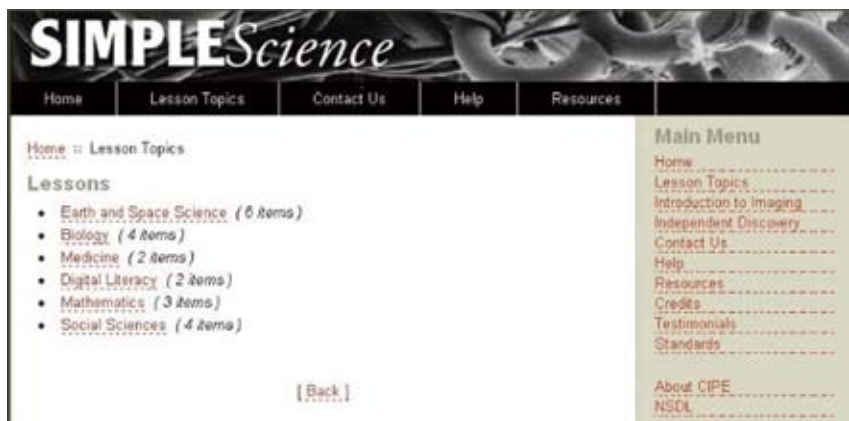


图2 SIMPLE Science课程主题

4.2.4 学习功能

SIMPLE Science 通过图像处理和分析技能(IPA)帮助学生克服课堂学习中的障碍。初接触IPA的新用户要花费一些时间熟悉该技术,不过随着用户学习课程的进行,既学习到了相关的数学和科学知识,也得以熟悉IPA技术。在学习过程中,正确的答案得到了正强化,即使答错,用户也会得到一个鼓励性的回复。每门课程结束后都附有相关的外部资料链接,例如一篇关于摄影新闻报道可信度的文章,为新闻机构更改图片数位等。

4.2.5 参考咨询功能

SIMPLE Science目前并不提供参考咨询服务,但是建议想要对图像处理进行深入研究的用户购买Discovering Image Processing的最新修订版(DIP2),DIP2包括更多课程,同时附有PDF格式的完整版用户指南,帮助教师以彩色或黑白形式打印课程和技能表单,并提供更广泛的课程支持和参考材料。

4.2.6 其他

除了上述的功能,SIMPLE

Science还拥有下载模板、创建用户课程等其他服务方式,为全球中小学教师及普通用户提供了一个可供参考的资源集合,帮助他们的工作与专业学习。

为验证SIMPLE Science在教学中所起的作用,网站曾邀请来自不同地区的学校教师参与到网站的测试中来,其中很多教师给出了可行性的建议以及对于网站的评价,具体内容可在网站右侧“实证”(Testimonials)部分查阅。

SIMPLE Science目前虽然已经停止资源更新,但仍有维护功能。网站内部的一些链接以及功能性栏目还需要通过注册页面,通过邮件联系该项目的成员,才能获得使用权限,普通用户无法直接使用。因此它的服务范围目前比较有限。

5 评价及建议

网站最值得称赞的是帮助咨询台。帮助咨询台可以通过界面顶端的“联系我们”进行使用,或者通过“My ticket”直接使用这一功能。在这里用户可以递交自己的问题或评论,用户通常可在五个工作日内得到回复。

通过对网站的试用,笔者认为SIMPLE Science在导航方式、搜索选项以及界面风格方面存在一些不足。SIMPLE Science的主页信息丰富,但由于网站目标用户为6-8年级的学生,欢迎页面图片较少,色彩不够明亮。在导航方式的设置上,SIMPLE Science为用户提供了两种选择,可以分别通过页面顶端的标签和屏幕右侧的主菜单选择所需课程,但其不足在于顶端标签和右侧的主菜单下属链接并不一致。

作为数字图书馆, SIMPLE Science最大的缺陷是缺少检索功能,使用户在查找其所藏六门课程科目内容方面感到不便。除主要课程之外, SIMPLE Science的其他资源还包括指导视频、独立探索、课程辅导、图片资料、帮助链接、技术支持链接、一个指向科学方法公司

(Science Approach) 和CIPE网站的资源页面、推荐和感谢、用户评价以及一个SIMPLE Science所有课程使用的国家教育标准链接。所有这些内容都应该可以通过一个站内路径或索引或主菜单找到。

尽管存在这些缺陷,笔者认为 SIMPLE Science数字图书馆提供

的课程内容丰富,并具有娱乐性,将教学融于娱乐之中。数字图书馆的结构也符合K-12年级学生学习数学及科学的特点及需要,而且对公众免费开放,实现了开放资源数据库。

参考文献

- [1] The SIMPLE Science [EB/OL]. [2010-05-05]. http://www.simplescience.org/component/option,com_frontpage/Itemid,38/.
- [2] NSDL.org-About SIMPLE Science Resources-The National Science Digital Library [EB/OL]. [2010-05-05]. [http://nsdl.org/search/?include_collection\[\]=2827371&verb=Search&s=10&n=10&item_num=6&identifier=http%3A%2F%2Fwww.simplescience.org%2F](http://nsdl.org/search/?include_collection[]=2827371&verb=Search&s=10&n=10&item_num=6&identifier=http%3A%2F%2Fwww.simplescience.org%2F).
- [3] Tutorial & Workshops [EB/OL]. [2010-05-05]. <http://www.isi.edu/isd/icce98/tutorial.htm>.
- [4] Science Approach - CIPE [EB/OL]. [2010-05-05]. http://www.science-approach.com/index.php?option=com_content&task=section&id=14&Itemid=54.
- [5] Science Approach - SIMPLE Science [EB/OL]. [2010-05-05]. <http://www.science-approach.com/content/view/51/54/>.
- [6] WebImage [EB/OL]. [2010-05-05]. <http://www.group42.com/webimage.htm>.

作者简介

刘燕权, 毕业于美国麦迪逊大学信息学院, 现在南康涅克州立大学任教。他近年的出版物主要集中在数字挖掘、数字图书馆、信息存储、数字化及多媒体技术、国家信息基础结构、图书馆统计及管理。E-mail: liuscsu@gmail.com

王群, 北京师范大学信息管理系图情09硕士生。

SIMPLE Science: Image-Based Learning Digital Library for K-12 Education

Yan Quan Liu / Southern Connecticut State University, New Haven, CT, USA

Wang Qun / Beijing Normal University Department of Information Management, Beijing, 100875

Abstract: SIMPLE Science is a digital library project aiming to "overcome barriers to mainstream use of image processing and analysis (IPA) in K-12 education". It attempts to make IPA accessible and easy to use, provide extensive and updateable archives of imaging data, and design a pedagogical structure that supports national education standards for middle school education. The article provides an extended review on the construction and status of the digital library, including project background, resources organization, technological structures, service features, as well as comments and suggestions made by the authors.

Keywords: SIMPLE Science, IPA, Digital library, National Science Digital Library (NSDL), National Science Foundation (NSF), K-12, Science education

(收稿日期: 2011-12-04)