

基于 B/S模式的科技文献管理系统的 策略和初步实现^{*}

梁春燕 李晓霞 聂峰光 郭 力 杨章远

(中国科学院过程工程研究所计算机化学实验室 北京 100080)

【摘要】 通过对国内外文献管理工具的现状与发展趋势进行分析,并结合当前文献信息的网络化特征,进行了文献管理系统基于 B/S(浏览器/服务器)模式下的策略研究,并讨论了相关的实现技术。本文设计了一个基于 B/S模式的文献管理系统框架,通过对用户的权限控制,以及对不同类型的科技文献信息的管理、批量导入和导出,使用户可以通过 Internet对文献信息进行有效的检索、维护、使用和共享,并进行了初步的系统(RefManager)开发和实现。

【关键词】 文献管理 Internet B/S模式 文献管理系统 权限控制 批量导入和导出 RefManager

【分类号】 G35 TP393

The Strategies and Realization of B/S Based Scientific Reference Management System

Liang Chunyan Li Xiaoxia Nie Fengguang Guo Li Yang Zhangyuan

(The Institute of Process Engineering, Chinese Academy of Sciences, Beijing 100080, China)

【Abstract】 With more and more scientific documentation are available electronically in Internet, reference management system is becoming an essential tool for daily researchers. In this paper, the strategies and the corresponding technologies to build a reference management system based on B/S (Browser/Server) scheme are discussed. The status and trend of the development of current reference managers are first overviewed, then a general framework of B/S based reference management system is designed, which mainly consists of four functional modules: user authority control, data import, management and export. RefManager, a preliminary reference management system based on the framework, is developed, which permits automated import of searching results in bibliographic databases such as EI Engineering Village and can export the references in the thesis format defined by the author's institute.

【Keyword】 Internet B/S Reference management system Authority control Import& Export RefManager

1 引言

中国科学院过程工程研究所从 2001年开始进行基于网络的文献管理系统的研究工作^[15]。基于网络的文献管理系统可以通过 C/S(客户端/服务器)模式和 B/S模式两种方式来实现。本文将主要介绍基于 B/S模式下的系统设计以及开发的初步结果。基于 B/S模式的文献管理系统,使用户可以通过浏览器直接访问 Internet上的服务器,方便地进行信息的管理和使用,并使系统的维护和更新更为方便和及时。

1.1 文献管理系统介绍

随着 Internet的发展,文献信息的发布和获取与过去有了很大的不同,文献信息的电子化和网络化越来越普及,Internet正在成为人们获取和交流信息的基本平台。许多科技期刊的全文可以通过网络直接获取^[1],例如重庆维普的中文科技期刊全文数据库^[2];著名出版商 Elsevier的 Science Direct电子期刊^[3];国内外各大图书馆也都在网上提供馆藏目录查询(OPAC)服务,例如中国国家图书馆^[4],美国国会图书馆^[5]等;还有许多综合性的专业站点提供包括会议信息、专业数据库、专利等各种丰富实用的资源,例如著名的化学化工虚拟社区 ChemWeb.com站点^[6]等。通过 Internet获取信息的方式也造成了信息的海量

收稿日期: 2003-11-18

^{*} 本文得到国家自然科学基金委项目(20160120230)的资助。

性和异质性,使得传统的文献管理方式难以满足信息的存储、组织、检索和使用等各方面的需要,从而迫切需要一种工具对信息进行有效的管理。

文献管理系统就是一种使用户(尤其是科研工作者)可以对各种类型的文献信息进行有效收集、管理、检索以及按照不同的格式进行输出的工具。如图1所示,对信息的收集可以通过“手工导入”和“自动导入”的方式来进行,“手工导入”就是通过拷贝、粘贴或直接输入数据进行文献信息的输入,“自动导入”就是通过系统提供的工具自动地将信息批量输入到系统的数据库中;对信息的管理,包括可以通过各种检索方式从数据库中找到所需的文献信息等;对信息的输出,可以将用户选择的信息按照选定的格式通过“自动导出”的功能进行自动输出,或者自动生成电子文档的参考目录。除此以外,由于如今的科研工作往往不是孤立进行,而是基于一个团队(Team)的,这就要求文献信息应该在整个团队中进行共享和交流。因此,文献管理系统应该是基于网络平台的,通过对用户的权限控制等方式,来实现文献的共享和交流。图1也说明,对科研工作者来说,方便有效的文献管理系统显得尤为重要,它贯穿于科研工作的整个过程中,与每一个阶段都密切相关^[7]。

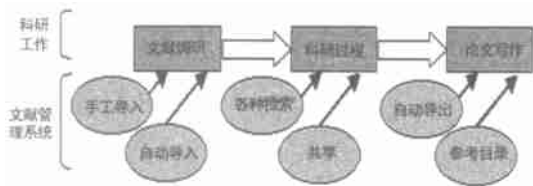


图1 文献管理系统和科研工作的关系

1.2 文献管理系统的现状

(1) 国外开发的文献管理系统

文献管理工具的开发在国外80年代初就已经开始。目前此类软件主要有: Papyrus, EndNote, Bibloscape, Procite, Reference Manager等,它们都各有自己的特点。英国 Bristol 大学的 Maggie Shapland对这些软件进行了比较详细的测试和比较^[8]。所有这些文献管理工具一般都实现了基本的功能,如对文献的导入、管理、检索和导出,只是实现的方式和侧重点各有不同。其中, Papyrus^[9], Endnote^[10]和 Bibloscape^[11]分别具有一定的代表性,应用也比较广泛。其主要性能比较如表1所示。

Papyrus目前只有DOS版本,起步较早且功能较为全面。其特点是可以对文献格式进行灵活的编辑,而且可以对文献的自动导入过程进行监控; EndNote是如今在学术界比较流行的文献管理系统,功能全面,性能比较优越。其优点是可以与Word字处理器进行紧密集成,可以在Word文档编辑时,

进行文献的引用和自动生成参考目录^{[12][13]}; Bibloscape的功能更为全面,而且与EndNote相比价格便宜、版本更新快、Web技术支持好,在一定程度上代表了文献管理系统的发展方向。它实现了几乎所有与文献管理系统有关的功能,其中有许多地方值得借鉴。同时, Bibloscape具有C/S模式和B/S模式两种版本。但其B/S版本实现的功能有限,只是用来作为对C/S版本的补充。

值得一提的是,这些软件基本都只支持英文,对中文的支持即使有,也不够完善。在用户进行中文文献的输入、管理和输出时,经常会出现乱码。

表1 Papyrus, EndNote和Bibloscape的性能比较

| | Papyrus | EndNote | Bibloscape |
|--------------|---------------------------------|-------------------------|------------|
| 适用操作系统 | Dos/Macintosh / Windows (Dos界面) | Dos/Macintosh / Windows | Windows |
| B/S版本 | 无 | 无 | 有 |
| 支持文献类型 | 8 | 16 | 20 |
| 文献描述的用户自定义 | 无 | 有 | 有 |
| 检索性能 | 好 | 不好 | 好 |
| 自动导入过程监控 | 有 | 无 | 无 |
| 导入导出格式的用户自定义 | 有 | 有 | 有 |
| 重复文献检测 | 有 | 有但不好 | 有 |
| 在文档中直接使用文献 | 无 | 有 | 有 |
| 与Web的连接 | 无 | 无 | 有 |

(2) 国内相关的软件

作者的调研结果显示,国内还没有类似的软件。只有专门用于图书馆管理的软件,如北京丹诚软件有限公司^[14]开发的图书馆管理系统以及一些功能单一的检索和管理软件,如北京丹诚开发的Z39.50前端软件Ztrans,可以使用户通过Z39.50协议从国内外的有关数据库中检索和获取文献。

1.3 文献管理系统的发展趋势

随着计算机和Internet相关技术的发展,一方面使人们对文献管理系统的功能有了新的需求,另一方面技术的发展也为新功能的实现提供了可能。我们认为文献管理系统的主要发展趋势有:

(1) 导入功能方面: 使用户可以从任何可以得到的资源中进行自动的文献信息导入。如直接从Internet上批量导入文献信息; 提供图书馆的馆藏目录查询接口(如使用Z39.50协议), 直接从文献管理系统中检索并下载馆藏书目; 提供相关的接口,从其它文献管理系统中,自动导入文献信息; 使用户可以通过自定义规则,将已有的以某种格式存在的文献信息导入到文献系统中等。

(2) 导出功能方面: 使用户可以在各种情况下方便地使用文献信息。在提供国内外主要文献格式的基础上,用户可以对文献格式进行自定义; 与字处理器进行整合,使用户在文档的编辑写作中,方便地对文献进行引用和生成参考目录。

(3) 管理功能方面: 使用户可以方便高效地管理文献信

息。如对文献全文的存储和管理;支持多语种;支持相关文献的互连;支持对图片和网上资源的存储和管理;支持多种排序方法;支持多种检索方式;针对特定的用户群,如科研工作者,将文献按照特定的方式进行组织和管理等。

(4) 共享功能方面:使用户可以方便高效地与他人共享文献信息。基于网络平台;建立用户权限机制;实现数据库的共享机制。

2 基于 B/S模式的文献管理系统的策略研究

通过对国内外文献管理工具的现状与发展趋势进行分析,本节对文献管理系统所应具备的基本功能进行了总结和概括,研究基于 B/S模式的文献管理系统所应具备的特征,讨论相关的实现技术。

2.1 文献管理系统的基本功能

文献管理系统的功能和特点在于使用户可以对文献进行有效的收集、管理和使用,于是导入、管理、导出这三方面的功能就构成了文献管理系统整个应用框架的核心。其基本功能框架如图 2所示。

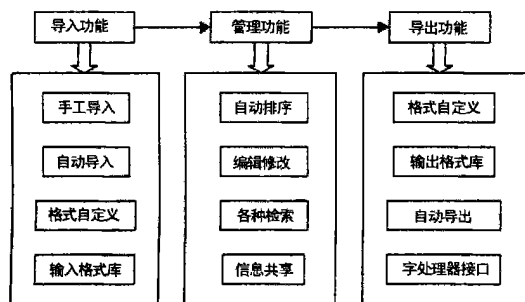


图 2 文献管理系统的基本功能框架图

2.2 基于 B/S模式的文献管理系统的特征

基于 B/S模式的文献管理系统可以有效方便的实现文献信息的共享和交流。由于它是基于网络并通过浏览器来提供相应的功能,必须将它置于网络的环境下进行考虑,因此基于 B/S模式的文献管理系统必须具有以下特征:

(1) 用户权限控制:对用户根据某种策略进行权限控制,使不同的用户具有不同的权限;同时,对数据库中的信息进行权限标记,对应不同的操作权限。

(2) 数据库同步控制:通过采用某种机制,使多个用户可以同时操作同一个数据库,同时保持数据库的一致性和完整性,并有效防止死锁的发生。

(3) 系统的安全性:使用防火墙等对系统进行保护,防止病毒或黑客袭击;对数据库进行备份,保证数据的安全性,使系统在崩溃或其它意外状况下,可快速有效地恢复。

(4) 系统的可访问性和可用性:应根据用户的使用习惯和系统提供的功能,设计友好的用户界面和系统结构;应该对系统进行测试,保证系统对不同的浏览器有良好的兼容性;应

尽可能提高系统对用户操作的响应速度。

2.3 基于 B/S模式的文献管理系统的实现方式

为了实现基于 B/S模式的文献管理系统的相关功能,需要用到网络、数据库、信息检索和计算机编程等方面的知识和技术。下面分别从几个方面进行讨论:

(1) 文献信息的导入和存储方面:由于系统收集的文献信息来自不同的资源,具有很明显的异质性和多样性。因此需要对这些不同的信息进行整合^[16]。具体的整合方法,主要有两种:一是可以通过为不同的格式设计相应的 API(应用编程接口)来进行文献的处理,其效率较高,但灵活性比较差;二是通过定义不同的处理模板,使其反映不同格式的文献与数据库的映射关系,来达到对异质信息的正确处理,其实现起来比较灵活,而且可以允许用户对模板的定制,因此是最常用的解决方式。

(2) 文献信息的导出方面:也可以通过模板来对数据库中的信息进行不同格式的输,以及允许用户对模板的自定义。为了使用户可以在文档写作的同时使用文献信息,可以使用 add-in 技术,将文献管理系统与常用的字处理器(如 Word)进行整合,通过它们之间的交互来实现文献信息的即时使用;或者在系统内部提供一个文档编辑器,使用户可以在系统内完成文档的写作和对文献的引用。

(3) 文献信息的管理方面:使用数据库的相关技术,对文献信息的存储格式进行合理的设计和实现;使用信息检索方面的技术,设计和实现对信息的各种检索。

(4) 文献信息的共享方面:利用数据库本身的功能来实现对数据库的权限标记和同步控制,以及对用户权限的分配,同时使用合适的编程方式实现用户权限的认证。如使用 PHP 中 Session 机制。

(5) 文献信息的传送方面:考虑到网络速度和带宽,使用缓冲、分页等相关技术来提高系统的响应速度。

3 基于 B/S模式的文献管理系统的框架设计和初步实现

3.1 系统总体框架设计

通过对文献管理系统的调研和分析,我们对基于 B/S模式的文献管理系统进行了一个完整的框架设计,如图 3所示。整个框架主要由四个模块构成,分别是:用户权限控制、导入、管理和导出模块。导入、管理和导出模块分别又包括了若干子模块。用户首先通过权限控制模块进入文献管理系统,并获得不同的操作权限,从而实现对系统中的共享和私有信息的保护;导入功能模块使用户可以方便地提交各种文献信息,并对这些信息进行自动的整合和存储;导出功能模块使用户方便地使用系统中的信息,生成各种格式的文

献参考目录;管理功能模块为用户提供了多种编辑、检索和批量处理数据的工具。这几个功能模块之间互相关联、互相作用,共同构成了一个完整的文献管理系统

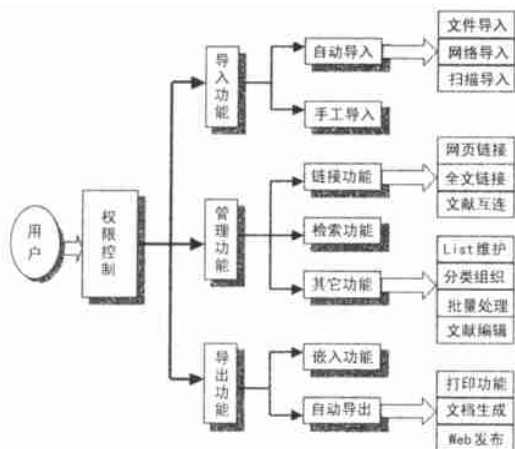


图 3 基于 B/S 模式的文献管理系统总体功能框架图

3.2 初步实现

根据对基于 B/S 模式的文献管理系统总体功能框架的设计,我们在过程工程虚拟研究中心^[17]的平台上初步设计并实现了基于 B/S 模式的文献管理系统(命名为: RefManager)。具体的系统配置、数据库设计和实现的功能如下所述:

(1) 系统配置

操作系统: Linux;

Web 服务器: Apache;

数据库: MySQL;

系统开发语言: C PHP JavaScript 等。

(2) 数据库设计

RefManager 的数据库结构包括四个方面的内容,即用户的管理,文献信息的管理,文献库描述信息的管理和导入导出格式的管理。

① 用户的管理: 将用户的相关信息存储在数据库的用户管理表中,以实现用户对用户的管理和权限控制。

② 文献信息的管理: 定义一个通用类型表和通用字段表。其中通用类型表包含了数据库中的所有字段,通用字段表则对这些字段的性质进行了定义。可通过提取类型的不同字段来定义不同的视图,显示给用户不同的信息。实现对不同类型文献信息的存储、管理和检索。

③ 文献库描述信息的管理: 对每一种不同的文献类型分别用两个表进行定义。一张表用来定义该类型的性质,另一张表用来定义该类型所包含的字段性质(如字段的中英文表示等)。同时每个字段的标识符同通用字段表保持一致。

④ 导入和导出格式的管理: 对不同的导入和导出格式分别用不同的数据表进行管理。

(3) 实现的功能

此系统经过一定时间的开发,已经初步实现了全部的基本功能。下面分别从几个方面来进行阐述:

① 权限控制 用户通过网络访问文献管理系统时,首先进行登录,通过权限控制模块进入文献管理系统,同时获得不同的操作权限。RefManager 将用户分为三种类型: 管理者、注册用户和非注册用户。不同的用户类型对系统中的信息具有不同的操作权限。管理者负责整个系统的管理和维护,具有最高的权限级别,对所有的数据都具有浏览、编辑、删除、检索、导入和导出等各种操作权限;注册用户具有中等的权限级别,可以创建自己的数据并对自己的数据具有全部操作权限,对系统的共享数据只有浏览、检索和导出的权限;非注册用户具有最低的权限级别,只能对共享数据进行浏览和检索,不具备任何其它的操作权限。

用户登录后,可以在主界面中选择所使用的功能,进入相应的功能模块;也可以通过重新登录来改变注册身份,进行不同的操作。

② 自动导入 自动导入功能,使用户可以根据不同的格式批量导入数据。用户可以通过选择导入格式,将本机上具有一定格式的文献数据文件,按照不同的文献类型自动导入到文献管理系统的数据库中。RefManager 已定义的导入格式包括: 本文献管理系统提供的导出数据格式 (Refmanager- Showall); 常用文献管理系统的导出数据格式 Endnote- showall 格式; EI Engineering Village 2 的详细导出格式和重庆维普期刊的文献信息导出格式等。

同时,用户在导入之前,可以创建自己的文献库,并选择其中的一个作为要导入数据的目标文献库。同时,用户也可以选择处理重复文献的方式,可以将导入过程中出现的与文献库中数据相重复的文献丢弃、忽略或导入到特定的重复文献库中。

③ 管理维护 管理和维护功能模块,作为系统的核心模块,具有多种功能,包括管理、编辑和检索功能以及其它功能,其中对导入和导出功能也进行了一定的扩充。

其中管理功能包括在不同文献库之间进行灵活的切换,以及对文献库的操作。前者可以极大地方便用户对不同文献库的编辑和检索,后者为用户提供了以文献库为单位的批量操作工具;编辑功能包括对文献的浏览、更改和删除,以及文献的添加——即文献的手工导入;检索功能包括对文献的简单检索、递进检索、不同文献类型的分类检索、布尔检索、模糊和精确检索、英文的大小写敏感检索以及对文献的自定义检索等各种检索功能。

同时管理和维护模块还提供了对检索结果的批量删除和批量导出的功能、检索结果的多种排序方式、分页显示和跳转显示、检索结果界面的显示字段可自定义性等,使整个管理系统更加灵活和方便。系统的管理和维护界面如图 4 所示。



图 4 系统的管理和维护界面

④ 自动导出 自动导出模块使用户可以将所选择的文献信息以

不同的文献格式进行自动输出,并以不同的文件格式进行存储。其中,系统提供的文献导出格式包括:本系统自定义的文献导出格式(Ref-manager-Showall),通过此格式用户完整地数据重新导入到系统中来,实现信息的可循环性;论文格式(根据《中科院过程工程研究所研究生学位论文撰写规则》中关于参考文献的若干规定进行定义),通过此格式,可将用户论文中所涉及的文献以规定的文献参考目录形式提供给用户。系统提供的输出文件格式包括:TXT文本文件和HTML文件,可分别用于信息的存储和发布。

4 结 论

基于B/S模式的文献管理系统可以为用户提供一个方便快捷的管理、维护、使用和共享交流文献信息的平台。本文通过对国内外文献管理系统的调研,并分析其现状与发展趋势,进行了基于B/S模式的文献管理系统的策略研究和相应的框架设计,这对文献管理系统的开发和实现具有一定的指导意义。同时我们对基于B/S模式的文献管理系统(RefManager)的初步实现也进一步证实了本文中所提出的策略和框架的可行性和实用性。需要提出的一点是,由于基于B/S模式的相关实现技术与C/S模式相比,有着一定的局限性,文献管理系统的用户自定义等功能的具体实现,尚需进一步的研究。

RefManager系统基本实现了用户的权限控制,以及文献的导入、导出和管理维护功能。同时,通过实际的应用对系统的实用性进行了测试。

(1) PEVRC的初步应用:2002年4月,基于网络的文献管理系统RefManager在过程工程虚拟研究中心(PEVRC)^[17]中进行了初步的测试和应用。在这次应用中只提供给用户浏览和检索共享文献库的权限。

(2) 局域网上的试用:系统开发之后,放在过程工程所的局域网中供大家试用,并在实际使用中不断地解决出现的问题和增加新的功能。

总之,通过各种测试和实际使用,证明本文中提出的基于B/S模式的文献管理系统的相关策略和框架设计具有可行性,在此基础上开发的RefManager系统 also 具有很好的实用性和发展前景。

参考文献:

- 1 李晓霞,郭力. Internet上的化学化工资源. 北京:科学出版社, 2000
- 2 重庆维普资讯有限公司. 重庆维普中文科技期刊全文数据库. <http://159.226.100.28/>. (Accessed July. 8, 2003)
- 3 Elsevier. ScienceDirect. <http://www.sciencedirect.com/> (Accessed July. 8, 2003)
- 4 中国国家图书馆. <http://www.nlc.gov.cn/>. Accessed July. 8, 2003)
- 5 The Library of Congress. The Library of Congress Online Catalog. <http://catalog.loc.gov/>. (Accessed July. 8, 2003)
- 6 Elsevier Ltd. ChemWeb.com. <http://www.chemweb.com/> (Accessed July. 8, 2003)
- 7 Catherine Beverley, Andrew Booth, Louise Falzon. Managing References. 2001. <http://www.shef.ac.uk/uni/academic/R-Z/scharr/ir/refman/manref/index.htm>. (Accessed July. 8, 2003)
- 8 Maggie Shapland. Evaluation of Reference Management Software on NT (comparing Papyrus with ProCite, Reference Manager, EndNote, Citation, GetARef, Bibloscape, Library Master, Bibliographica, Scribe, Refs). 1999. <http://www.cse.bris.ac.uk/~ccmjs/rmval99.htm>. (Accessed July. 8, 2003)
- 9 Research Software Design. Papyrus. <http://www.researchsoftwaredesign.com/> (Accessed July. 8, 2003)
- 10 ISI ResearchSoft. EndNote. <http://www.endnote.com/>. (Accessed July. 8, 2003)
- 11 CG Information. Bibloscape. <http://www.bibloscape.com/>. (Accessed July. 8, 2003)
- 12 ISI ResearchSoft. EndNote Manual. 2002
- 13 Bjoerner, Susanne. Bibliography formatting software for managing search results. EndNote plus and EndLink. Online (Wilton, Connecticut), 1995, 19(5): 38-43
- 14 北京丹诚软件有限责任公司. <http://www.datatrans.com.cn/>. (Accessed July. 8, 2003)
- 15 贾红阳. Internet化学化工信息服务系统平台研究及应用. 博士论文. 北京, 2002
- 16 贾红阳,梁春燕,郭力,李晓霞,杨章远,许志宏. 各种文献信息的Web整合. 科学数据库及其信息技术第六届学术讨论会. 2002
- 17 中科院过程工程研究所. 工程研究虚拟研究中心. 2002 <http://pevrc.ipe.ac.cn/>. (Accessed July. 8, 2003)

(作者 E-mail: maria_cyliang@yahoo.com.cn)

(上接第34页)

参考文献:

- 1 李丽宾. 图书馆一卡通系统. 大学图书情报学刊, 2002(4)
- 2 王林. “一卡通”图书馆自动化的新领域. 现代图书情报技术, 2002(4)

- 3 杨天军. 校园管理一卡通系统在图书馆的应用. 江苏图书馆学报, 2002(5)

(作者 E-mail: tzhaox@jnu.edu.cn)