

基于 Notes/Domino 的 Internet/ Intranet 数据库全文检索

桂建军 尹韶峰

(中国科学院沈阳自动化所 沈阳 110015)

【摘要】 简述了 Lotus Notes 软件及其数据库特性,介绍了 Notes 数据库构建和在 Internet/ Intranet 环境下提供全文检索的途径和方式。

【关键词】 Internet/ Intranet Notes/ Domino 数据库 全文检索

目前网上提供的 WWW 数据库检索主要采用两种方式,一种是编写 CGI 程序,利用高级程序语言访问数据库。这种方法编程难度高,工作量大,不易使用。另一种是利用 Web 服务器具有的客户功能,通过内置的数据库连接器,采用 ODBC 技术访问数据库。这种方法数据查询功能强,不需要复杂的编程工作,目前在主页上被广泛的使用。但由于其驱动程序仅支持标准的 SQL 语句,因此其主要用于对关系型数据库的查询。对于全文型和多媒体数据库,这种方法在处理文本和查询能力方面有很大的局限性。而目前已在我国广泛使用的 Notes/ Domino 群件系统,以其强大的非结构化数据库全文检索功能,以及与 Internet/ Intranet 服务器的无缝集成技术,极便于在网络环境下提供数据库全文检索。本文根据我们在工作中实践,简要地介绍一下 Lotus Notes 及其在 Internet/ Intranet 环境下数据库全文检索中的应用。

1 Notes 与 Notes 数据库概述

Lotus Notes 是美国莲花公司推出的先进的通讯处理软件和群件产品,已广泛地应用于构建企业的 Intranet。它综合了四大核心技术:强大而灵活的复合文档数据库、丰富的应用开发环境、完备的电子邮件系统和与 Internet/ Intranet 的无缝集成。具备通讯、协同操作和协同运行等多种信息管理方式,概括起来主要有以下几个特点:

- (1) 跨网络、跨平台、跨协议协同工作能力;
- (2) 自带工作流,内含强大的工作流软件开发环境;
- (3) 用文档数据库实现多媒体文档管理,并支持客户机/服务器方式;
- (4) 提供客户机/服务器结构的功能强大的邮件服务系统;
- (5) 具有丰富的管理和监控能力;
- (6) 具有完善的安全性能;
- (7) 具有极强的伸缩性及与其它软件的交叉共用性。

最近,Notes 已推出了 5.0 版。从 Notes 4.5 版开始,Lotus 公司将 Notes 分成了两个部分,一部分是客户软件,称为 Notes;另一部分是服务器,称为 Domino。目前 Domino 已转

化成一种 Internet 服务器。运行该软件就可以建立自己的 WWW 服务器,然后既可以用 Notes 客户软件,又可以用标准的浏览器来访问该服务器。

Notes 数据库作为 Notes 群件产品中的一个重要组件,它不同于普通的关系型数据库。关系型数据库虽然也能处理非结构型的数据,但其核心是以行和列组成的矩形表格为基础,系统结构非常固定,因此最适合处理结构型的数据;而 Notes 数据库虽然也可以处理结构型的信息,但它主要用来处理以文本信息为基础,可包括图像、图形等多媒体信息的非结构化型数据,它是一种适合处理文档型信息的非结构化数据库,其基本存储单元是文档。Notes 数据库中的文档类似于关系数据库中的记录,可与同一个数据库或其它数据库中的文档建立关联。相对于字段为定长的关系型数据库,Notes 数据库更适合处理一些不定长的文摘型和全文型数据库。Notes 数据库支持中英文全文检索方式。

2 Notes 数据库的构建

Notes 数据库是一个用来储存信息的 Lotus 文件,它由域(Field)、表单(Form)、视图(View)、文档(Documents)、代理(Agent)等元素组成。域是 Notes 数据库字段,用来输入信息和执行计算。表单是一个用于输入、显示或打印数据的窗口。视图是一个按特定顺序排序、用以显示数据库文档的窗格。文档是 Notes 数据库中的记录,它由数据库表单中的一个或多个能输入信息的域组成。代理是在数据库中执行一个特定任务或功能的过程。

当创建一个 Notes 数据库时,Notes 新建数据库对话框提供了数据库标题、存放数据库的路径和文件名(NSF 为 Notes 数据库的扩展名)、数据库大小(以 1GB 为增量档次,最大为 4GB)及一些内置的数据库模板等输入栏和选项。当在该对话框中输入或选定合适参数后,Notes 在当前的工作台上建立了该数据库图标,并打开了数据库的设计窗口。

2.1 创建表单

Notes 数据库的设计首先应是数据库表单设计。在数据库设计窗口,选取表单或工具栏表单设计项,进入表单设计

界面。表单设计有二个窗口:主窗口和公式窗口。其主窗口就像是一个文档编辑窗口,在该窗口可以创建域,也可以加入静态文本、图形、图像、表格、按钮等其它表单元素,并可进行版面和格式的随意编版。表单的公式窗口位于表单设计界面的下部,在该窗口可输入 Notes 函数、公式和 LotusScript 语句,用来设定表单元素的属性和为表单创建自动执行的任任务。

Notes 域分数字、关键字、时间、文本和 Rich 文本等八种数据类型,其域为不定长。Notes 文本型和 Rich 文本型的区别在于文本型域只能包含一般字符和不适用于数字运算的数字, Rich 文本型域为多信息文本域,可插入图片、图像、热点、附件或嵌入式对象。在表单设计过程中,用鼠标右击主窗口的表单元素,就可对表单和表单元素的各项基本属性进行设置和修改。当完成表单设计并存储表单后,该表单就出现在数据库设计窗口,用户可随时点击该表单进行修改。同理,可以为 Notes 数据库设计多个不同用途的表单。

2.2 创建视图

Notes 数据库至少应有一个视图。当进入数据库设计窗口,Notes 就已经为数据库提供了一个未命名的缺省视图。视图设计既可以在缺省视图上进行,也可以通过菜单栏视图设计选项新建一个视图。

视图设计窗口类似于表单设计窗口,在上部的主设计窗口上,可根据需要设计视图中的列。视图中的列可以是数据库的域名,直接显示文档中域的信息;也可以是一个函数,显示视图中文档的号码、文档创建和修改日期及文档的属性。在视图下部的公式窗口,用来输入 Notes 公式和 LotusScript 语句。像表单设计一样,在视图设计时,用鼠标右击视图设计窗口或点击菜单选项,可随时打开视图属性和视图中每个列的属性对话框,来定制该视图和视图中列的属性,如标题、别名、字体、颜色、显示和排序方式等。

完成了数据库表单和视图的设计,就基本完成了一个数据库的设计工作。当然还可以创建一些能自动执行特定任务的代理和方便操作的按钮,来提高数据库的运行效率。在数据库投入使用之前,还需要对数据库进行最后的设计:设置适当的访问权限,书写数据库帮助文档,确定数据库启动选项等。

2.3 创建文档

完成了数据库表单和视图设计,就该往数据库里输入信息,生成数据库文档。Notes 文档产生主要有两种方式:Notes 下的创建和外部数据的引入。

Notes 下数据库文档的创建是通过 Notes 代理程序和人工输入来实现的。代理是一段可自动运行的宏,它通过简单操作、公式和 LotusScript 代码来实现对数据库操作。人工输入是在数据库表单上进行的,当激活一个输入表单,进入编辑模式,该表单的域就可接受合法的数据输入。像其它数据库一样,Notes 有些域需要人工输入数据,有些域可以继承来自另一个文档或环境变量的数据;而 Rich 文本型域,由

于它是一个多信息文本域,可使用普通剪贴板,将文本、图形和电子表格等信息粘贴到域中,并可在该域中链接或嵌入对象。

Notes 还可通过文件引入功能和 LotusScript 语言把外部系统及关系型数据库的数据直接移植到 Notes 数据库中。Notes 文件引入支持三种类型的文件源: Lotus 1- 2- 3 工作表、结构化文本、制表文本。LotusScript 是一个面向对象的编程语言,它的函数和句法与 Visual Basic 十分相似,在 Notes 数据库表单的公式窗口用 LotusScript 内置的 ODBC 类可访问外部关系型数据库,与外部数据库进行数据集成。

3 Notes 数据库的共享与发布

3.1 Notes 数据库的共享

Notes 支持其数据库在网域内共享,即同一个网域的用户都可用 Notes 客户软件使用本网域内数据库资源。所谓 Notes 网域,就是使用同一个公用通讯录的一组 Domino 服务器及其客户机。在安装 Notes 软件时,Notes 首先要求输入一个网域名,并确定该网域第一个服务器就是该网域的主服务器,同一个网域内其它服务器和客户机 Notes 软件安装必须从主服务器上得到身份验证、授权和密钥。这样该网络里所有服务器、客户机和用户信息均保存在该网域的同一个公用通讯录中,确保了 Notes 网络的安全和对网络统一管理。

Notes 数据库共享主要是通过数据库的存取控制表 (ACL) 和 Notes 数据库复制功能来实现。Notes 数据库的存取控制表是用来确定用户和服务器对数据库的存取级别。而数据库的复制可使一个数据库在网域内有多个复本并保证复本信息同步更新。Notes 数据库通过 ACL 赋予服务器和用户的存取级别,来决定服务器可以复制数据库哪些信息,用户可执行数据库上的何种操作。

3.2 Notes 数据库的发布

Notes 数据库既可以在本网域内共享,还可以在 Web 服务器发布。发布后的 Notes 数据库可在 WWW 环境下接受 Internet/ Intranet 用户浏览器的访问。在 Notes 数据库发布之前,必须成功地安装并运行 HTTP 网络协议的 Internet 服务器,并对 Domino 服务器作必要的配置。如果在 Internet 上发布,还应确保你数据库存放的服务器与 Internet 建立了连接。

Domino 服务器配置,主要是设置该服务器网络名、网络协议名、服务器端口号或缺省字符集等参数。在完成了服务器的配置后,还必须保证该服务器启动了 HTTP 协议和要发布的数据库已建立了全文索引。HTTP 协议的启动,可每次在 Domino 服务器控制台手动输入命令 'Load HTTP',也可在 Notes. ini 文件 Tasks 行中加入 'HTTP' 项在启动 Domino 服务器时自动启动。Notes 数据库的全文索引则是通过 Notes 数据库属性对话框提供的创建、删除和更新全文索引选项来实现的。

(下转第 47 页)

因此,馆领导要重视馆员思想观念的革新,并订出切实可行的人员培训计划,分期分批轮训业务干部,使之尽快熟悉信息服务新手段,适应图书馆信息化、智能化、社会化和国际化发展的需要。

(2) 重视用户教育。

美国情报学家兰开斯特(Lancaster)曾经断言,“联机情报检索系统成败的关键在于使人们懂得如何使用这些设施的方法效果”。机检系统建成后,图书馆应把这种先进的检索方式介绍给用户,可以采取以下措施:

¹ 将机检常识列为新生入馆教育必修内容,使学生一进学校就能受到计算机检索技能的初步培训,为获取网上信息打下良好基础;

² 从文检课的内容和教学方式上打开突破口。加强网络基本知识和基本服务功能、超文本检索、多媒体检索、全文检索等先进检索手段以及WWW等网络检索工具知识的教育。强化检索策略分析和上机实习环节,提高学生的机检技能;

³ 举办专题讲座。在机检知识技能讲授中融入专业内容,使用户真正感受到掌握机检技能犹如为自己铺设了一座通向成功彼岸的桥梁,从而自觉地参加学习。这种讲座既是服务育人和增强用户信息意识的一种有效方式,又是馆员联系用户的纽带;

⁴ 为了促使图书馆重视用户教育,在图书馆评估中,将接受过机检用户教育的师生数占持证师生数的比例,作为衡量该馆服务质量的一项指标体系看待。

(3) 加强宣传,培养用户信息意识

图书馆应采取多种方式加强舆论宣传,不仅使用户知道机检系统的存在及机检所具备的特点和优越性,掌握必要的计算机检索知识,而且要培养用户的信息意识,才能从根本上改变设备闲置、信息资源浪费的现实。

(4) 重视数据库建设

数据库是公共检索系统的生命线,除了标准化规范化外,还应注意以下几点:

¹ 在建库方式上,要尽量利用国家图书馆出版的CNMARC光盘数据库建库,这类数据库标准规范,学科覆盖面广、数据量大、补充更新及时、检索方便。以其为主,自行按MARC格式建库为辅,是经济快捷建库的一条捷径;

² 在建库种类上,要尽快启动外文图书数据库、中外文期刊数据库和音像数据库建设;

³ 在建库特色方面,要密切注意本校的专业特色和校际之间资源共享的问题,争取做到不重复、有特色;

⁴ 在建库人力投入上,要组织协调馆内各部门的工作,在不影响正常开放情况下多投入人员,争取在短期内完成建库任务。

(5) 强化基础工作,不断完善检索系统

解决检索率问题可采取馆员指导读者检索的办法,既增大文献查全率和查准率,又在图书馆与用户之间架起沟通的桥梁,对信息的开发利用和用户培训都有促进作用;对系统功能问题应随时注意收集情况,不断改进和完善系统功能,提高用户对图书馆服务效能的满意度;在加强馆藏数据库建设的同时,通过校园网使广大教研人员和学生能在全中国乃至全球的计算机网络环境下最大限度地利用网上资源开展教学科研活动。

总之,计算机检索方式取代传统手工检索方式这项服务,解决的不仅仅是查询手段变化的问题,而且是文献信息和知识能否充分利用、信息资源共享能否顺利实现的问题。向用户提供这项服务,以弥补传统图书馆服务功能的不足,实现图书馆知识信息服务的网络化、国际化,在一定程度上反映了图书馆自动化建设的方向,值得我们努力去探索。

参考文献

- 1 李桂琴.利用计算机检索的障碍与对策分析.云南图书馆,1998,(4)
- 2 胡昌平.面向新世纪的我国图书馆网络化知识信息服务的宏观组织.中国图书馆学报,1999,(1)

(上接第45页)

Notes数据库在WWW上发布主要有二种形式,一种是直接把Domino服务器上的数据库设置成为主页。另一种就是通常使用的一种方式,在当前服务器主页上建立与Notes数据库的链接。链接的格式是:HTTP://服务器名或IP地址/数据库名。Notes能把Notes表单和视图自动转换成HTML格式。在这种方式下,可通过在Notes表单上创建一些特殊的域,在服务器主页上定制用于检索、显示和输入数据库信息的界面,构造交互式表单。

完成了Notes数据库在主页上的发布,也就实现了该数据库在WWW环境下的全文检索。Internet/Intranet用户通过浏览器就可在网上查询Notes数据库。Notes支持在WWW环境下使用布尔逻辑式和通配符检索数据库信息。

4 结束语

从以上可以看出,Notes/Domino作为一个功能强大通讯处理软件和群件产品,不仅适合开发现代化办公服务系统,用于构造企业的Intranet,其复合文档数据库、与Web服

务器无缝集成技术及强大的 workflow 软件开发环境,也适合文献情报部门在Internet/Intranet环境下提供文摘型、全文型及多媒体型数据库共享和服务。中国科学院沈阳自动化所文献信息中心利用Notes/Domino软件已先后实现了《机器人文摘》全文库,机器人文献库、先进制造技术文献库等数据库的上网查询。其在WWW环境下检索功能强,效率高,且无需复杂的编程工作,有力地推动了本所文献信息资源的共享和图书馆自动化建设的进程。

参考文献

- 1 (美)Randal A. Tamura等著. Lotus Notes和Domino Server 4.6技术大全.北京:机械工业出版社,1998,10
- 2 (美)Barbara A. Bolin等著. Lotus Notes 4.5中文版使用详解.北京:机械工业出版社,1998,8
- 3 史国祥.用Foxweb实现WWW书目查询.现代图书情报技术,1998,(2)
- 4 张汛涑. Lotus Notes编程技巧三例.电脑编程技巧与维护,1999,(2)