基于 ASP、SQL Server 2000 实现的 Web文献检索系统及其查询优化

(北京师范大学资源学院,北京师范大学环境演变与自然灾害教育部重点实验室 北京 100875)

介绍了基于 ASP、SQL Server 2000实现的 Web文献检索系统及其查询优化。通过调试、运行,Web文献检索系统运行正 常,查询效率高。基于 ASP、SQL Server 2000实现的 W eb文献检索系统简单、安全、稳定,维护方便。

关键词 ASP SQL Server 2000 Web文献检索系统 查询优化

IM PLEMENTING WEB LITERATURE RETRIEVAL SYSTEM WITH ASP, SQL SERVER 2000 AND THE QUERY OPT IM IZATION

Song Yang Yan Ping Cao Tong

(College of Resources Science and Technology, Key Labora tory of Environmental Change and Natural Disaster the Ministry of Education of China, Beijing Normal University, Beijing 100875, China)

This paper introduces the fulfillment of web literature retrieval system with ASP, SQL Server 2000 and the query optimization By Abstract debugging and running the web retrieval system, the system runs soundly, with a high efficiency of query. The web retrieval system based on ASP and SQL Server 2000 is simple safe stable and easy to maintain

K evw ords ASP SQL Server 2000 Web literature retrieval system Query optim ization

引 言

随着文献信息量的剧增和 Internet技术的发展,文献信息自 动化管理已成为当今文献信息管理的一种趋势,文献信息网络 化变得日益重要。文献数据库建设是文献信息自动化管理的基 础和核心。基于文献数据库的 Web文献检索系统可以极大地 提高人们检索文献的效率。本文以"中国沙漠数据库文献查询 系统"为例介绍了基于 ASP、SQL Server 2000 实现的 Web文献 检索系统及其查询优化。

技术基础

Web文献检索系统以Windows 2000 Server作为服务器开发 平台,选择支持 ODBC (开放式数据库互连)接口的 SQL Server 2000作为后台数据库;以微软的 IIS 5.0(Internet信息服务器) 作为 Web服务器;选用 ADO(Active X数据对象)作为访问 Web 数据库的接口,将 ASP(动态服务器主页)与 ADO 相结合,建立 和提供包含数据库信息的主页内容。通过执行关键字查询的 SQL命令,让用户在 Web浏览器端进行文献的检索操作,并可 利用 Vbscript等脚本语言来控制数据库的访问,实现查询结果 的输出显示等功能。

使镶嵌在 HTML页面中的脚本程序得以运行的环境,可支持多 种脚本语言。它不但可以进行复杂的数据库操作,而且生成的 页面具有很强的交互性,允许用户方便地控制和管理数据。 ADO是微软在 OLE DB(对象链接和嵌入数据库)标准的基础上 定义的一种跨越操作平台,面向对象,与语言无关的数据库应用 编程接口,可以连接多种支持 ODBC的数据库。它语法简单,易 于理解,其中只包含了几个不多的功能强大的对象,易于使用。 把 ADO与 ASP结合起来访问 Web数据库是一种理想的 Web数 据库访问的解决方案[1]。

SQL Server 2000 是微软公司发布的一种性能优越的面向客 户机 服务器的关系型数据库管理系统。它与 W indows NT/ 2000等操作系统紧密集成。这种安排使 SQL Server 2000能充 分利用操作系统所提供的特性, 创建运行在 W indows NT /2000 等多种操作系统平台上的平稳和高性能的数据库应用程序,并 对大文本和多媒体数据提供了很好的支持。

2 系统结构

2.1 逻辑结构

Web文献检索系统在逻辑上可划分为三层模式:Web游览 器 /W eb服务器 数据库服务器 (图 1)。其实质是将客户端的应 用程序移植到 Web服务器上,由 Web服务器担任数据库服务器

收稿日期: 2004-09-29。基金项目: "十五"国家科技攻关项目 自然科学基金项目(30371191)资助。宋阳,博士生,主研领域:干旱区资 源环境与网络数据库。

的客户端,向服务器端发 送数据和应用请求以及存 放返回结果,而真正的客 户端浏览器在安装标准的 TCP/IP协议和 Web服务

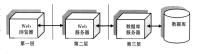


图 1 Web文献检索系统三层 模式逻辑结构图

器建立连接后,只需调用存放在 Web服务器上的应用请求页面 和结果显示页面,而无需关心应用程序。因此,实现这种模式的 核心技术就是Web服务器与数据库服务器的连接技术和数据 库访问技术[2]。

2.2 物理结构

Web文献检索系统的物理结构是在三层逻辑模式结构的

基础上设计的(图 2)。考虑 到 ASP内含在 W indows 2000 Server的 IIS 5.0 中, 而逻辑 上的 Web文献检索系统三 层模式结构在物理上可以合 并为两层,即将第二层的

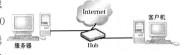


图 2 Web文献检索系统 的物理结构图

Web服务器与第三层的数据库服务器放在一起。因此,选择一 台配置较高的服务器(较高的 CPU处理速度,较大的内存,大容 量的硬盘)即可满足服务器设计的需求。在这台服务器上安装 Windows 2000 Server作为服务器开发平台,安装 SQL Server 2000作为后台数据库。随着数据库表中数据量的逐渐增大,可 添加磁盘阵列以增加服务器的存储空间。用一台普通配置的计 算机 (客户机)作为 Web数据库前端的开发平台,并作为 Web 浏览器测试对后台数据库的访问。

开发讨程

3.1 数据库的设计

利用实体一关系模型确定实体一关系图(ER图)及实 体一关系图中的属性。把每个实体都表示为关系(二维表),实 体的属性也就成为关系的属性。关系建立后,根据规范化准则 进行检查,如果需要,就将一个关系表分解成两个或多个关 系表。

具体设计时,首先启动 SQL Server 2000的企业管理器,在 SQL Server组中新建一个 SQL Server注册 —— Book。打开 Book 注册,右击"数据库"文件夹,选择"新建数据库"选项,建立一个 新的数据库—— Literature。展开 Literature库,右击"表"图标, 选择"新建表"选项,建立一个新表 —— Book表(图 3)。然后右 击 Book表,选择"设计表",完成对数据库结构(表 1)的设计。



图 3 文献数据库的建立

表 1 文献数据库的结构

| 字段名称 | 中文意义 | 数据类型 | 长度 | 主关键字 |
|-----------|-----------|-----------|-----|------|
| Bookno | 文献编号 | char | 30 | √ |
| Booknam e | 文献名 | char | 100 | _ |
| Author | 作者 | char | 50 | _ |
| Abstract | 摘要 | varchar | _ | _ |
| Keyword | 关键词 | char | 50 | _ |
| Pub_date | 发表时间 | datatim e | 8 | _ |
| ISSN | 国际标准期刊码 | char | 30 | _ |
| Volissue | 卷号、期号 | char | 50 | _ |
| Qkname | 文献出处(期刊名) | char | 200 | _ |
| Location | 文献的保存路径 | char | 200 | |

在数据库结构表中,因为 Bookno(文献编号)是唯一的,所 以将它作为主关键字。文献编号的命名规则是:期刊名首写字 母,年号,卷号,期号,文章序列号。如 ESP2003060301表示 Environmental Science and Policy期刊 2003年第六卷、第三期的第 一篇文章。为方便数据库的维护,用"记事本"记录了期刊名称 与缩写的对照表。在 SQL Server文献数据库表中保存的只是文 献信息,并不存储文献本身。文献以. pdf. vip. caj等格式存储 在服务器硬盘的其他目录下。为检索得到文献全文,还需要在 文献数据库中设置 Location字段,用来记录文献的存储位置,方 便系统链接得到文献全文。一旦定义完数据库,分配完物理存 储空间后,就可以向数据库中录入数据了。数据录入完成后还 要对数据的一致性、完整性进行验证。录入数据的文献数据表 如图 4所示。



图 4 录入数据的文献数据表

3.2 SQL Server数据库的连接与 ODBC 数据源的建立

数据库设计好后,在服务器端使用 ODBC(32位)数据资源 管理器为数据库创建数据源名 (DSN)并建立与 SQL Server数据 库的连接。具体步骤如下:

- (1) 在 SQL Server数据库中建立新登录。具体方法为:在 SQL Server中选择"安全性\登录\新建登录"。建立新的用户登 录 ID并输入密码。如:建立新的用户登录 ID: guest
- (2) 在要建立 ODBC 数据源的数据库中建立新用户。具体 方法为:在 SQL Server中选择"数据库\需要建立 ODBC数据源 的数据库\用户\新建用户"。在用户登录列表中选择一个用 户,如:数据库\Literature\用户\新建数据库用户:guest
- (3) 在 W indows 2000 Server的 "程序"中选择"管理工具\ 1994-2021 China Academic Journal Electronic PQBBSN数据调心系统,RRYing建立copRc数据源的选择建立nRA het Server数据源,在"SOL Server使用何种方式验证用户登录 ID" 中,选择"使用用户输入登录 ID和密码的 SOL Server验证",然

后分别输入登录 ID与密码。

3.3 设计 W eb 文献检索页面

为了方便用户检索文献,需要设计一个友好的 Web文献检索页面。在设计时,因为是静态网页,所以使用 Marcomedia Dreamweaver MX进行设计。设计好的页面(Search htm)如图 5所示。用户检索页面分成左、右两栏。在左栏的 form 表单中设置了"期刊名"、"文献名"、"作者姓名"、"关键字"和"出版年月"几种检索方式,用户可以选择不同的检索方式使用关键字进行文献检索。它们分别对应后台 SQL Server文献数据库 Book表中的"Qkname"、"Bookname"、"Author"、"Keywork"和"Pub_date"字段。它们在 form 表单中的名称分别为:"qkname"、"bookname"、"author"、"keywords"和"pub_date",以便在 ASP查询程序中调用。当用户选择其中一项输入检索条件并按下"开始查询"按钮后,表单就会调用 Query asp程序,执行文献检索,文献查询结果显示在右栏。

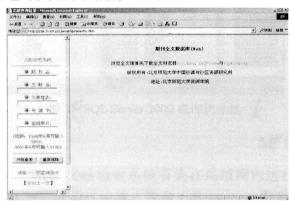


图 5 用户检索页面

3.4 创建 ASP查询程序: Query asp

为了处理用户通过 Web文献检索页面提交的检索请求,连接 SQL Server 2000台后数据库,执行 SQL查询并将查询结果显示输出,还需要创建 ASP查询页面。ASP查询程序部分代码如下。

Sub Output_book() 建立数据库连接,显示查询结果

 $D\ \mbox{im}\ q\ k\ n\ am\ e\ booknam\ e\ au\ tho\ r\ pub_\ date\ k\ eyw\ o\ rd\ s$

定义期刊名称、文献名称、作者姓名、出版年月及关键字

Dim cond_count we SQLstr 指定条件栏位总数、通用字元、SQL字串 cond_count=4 记录指定条件的栏位总数

wc= "%" 指定 SQL指令中的通用字元

qkname=Request Form("qkname") 取得表单栏内容

 $booknam \, e = R \, equest \; \; Form \; (\; "booknam \, e") \qquad author = R \, equest \; \; Form \; (\; "author")$

pub_date = Request Form ("pub_date") keywords = Request Form ("kdywords")

SQLstr="select" from book where" 执行 SQL查询语句

If Len (qkname) > 0 Then SQLstr = SQLstr & "qkname like" & wc &qkname & wc &""

 $\label{eq:cond_count} If cond_count \ge 1 \ Then \ SQLstr = SQLstr \& \ ''and'' \ cond_count = cond_count^-1$

End If

End If

If Len (book) and Parall Solving Solving Solving Solving Command Exception Country House book name & we & ""

If $cond_count \ge 1$ Then $SQLstr = SQLstr \& "and" cond_count = cond_$

count-1

End If

End If

If Len(author) ${>}0$ Then SQLstr=SQLstr& "author like" & wc & author & wc & """

If cond_count>1 Then SQLstr = SQLstr & "and"cound_count = cond_count-1

End If

End If

If Len(Pub_date) \geq 0 Then SQLstr=SQLstr & "Pub_date like" & wc & Pub_date & wc & """

If Cond_ count > 1 Then SQLstr = SQLstr & "and" cond_ cound = cond_ count-1

End If

End If

If Len(keywords) >0 Then SQLstr=SQLstr& "keywords like" & wc keywords & wc & """

End If

SQLstr=SQLstr & "order by bookno"

set adocon = Server CreateObject("ADODB Connection")

建立 ADO Connection 然后以 DSN-less方式登录到 SQL Server

adocon Open "dsn = Book; U \mathbb{D} = guest PW \mathbb{D} = guest database = Literature"

执行 SQL指令,并将结果储存于 Recordset中

set sal = adocon Execute(SQLstr)

If sal EOF Then

Sub Output Literature() 显示查询结果(程序代码略)

End if

End sub

在 ASP查询程序中,首先使用 Request Form ()语句从 Web 文献检索页面中取得表单栏内容,并存入对应的变量中。编写 SQL查询语句,为文献检索做好准备。接着,使用 "Server CreateObject"建立连接的对象,并使用 "Ope"命令打开待访问的数据库—— Literature库。然后设定 SQL命令,开始执行访问查询数据库的动作。最后,将执行结果储存于 Recordset中并通过Output_Literature()函数显示查询结果。例如:用户选择 "期刊名"检索选项,在 "期刊名"检索项中输入 "中国沙漠"后按下"开始查询"按钮,检索执行页面如图 6所示。在文献查询结果页面中显示了符合检索条件的文献的相关信息:作者;时间; ISSN;卷号、期号;期刊名;文献名。它们分别显示了后台 SQL Server文献数据库 Literature表中的 "Author"、"Pub_date"、"ISSN"、"Volissue"、"Qkname"、"Bookname"字段的内容。如果用户想浏览全文,可点击"文献名"的链接。



图 6 文献查询结果页面

4 查询优化

查询优化是为了提高数据查询的效率和性能。在网络数据库系统中,系统的性能主要受 I/O, CPU 和网络瓶颈的影响。为了提高系统性能,必须精心设置服务器、数据库、进程并减少数据库的冲突,消除瓶颈,使系统的处理时间最短^[3]。

4.1 服务器与操作系统的设置

SQL Server 2000数据库内核使用了一个基于成本的查询优化器自动优化向系统发出的数据查询操作^[4]。通过查询分析、索引选择和连接选择三个阶段来实现 SQL Server 2000的优化过程(图 7)。Windows 2000 Server支持 FAT和 NTFS两种文件系统。尽管 NTFS有许多高级的功能,但是 SQL Server 2000管理自己的数据访问不使用这些高级功能。因此,在系统开发时,将 SQL Server 2000安装在服务器单独划分的 FAT分区上以获得最快的文件访问速度。为了提高系统性能,尽量减少分配给操作系统的内存和高速缓存,不使用不必要的驱动程序与写缓存;合法设置参数,避免网络错误。



图 7 SQL Server 2000 查询优化器

4.2 使用合适的数据类型

在 SQL Server 2000数据行中,有定长列和变长列之分。变长列比定长列的开销大。数据行的定长列需开销 4字节,而对于变长列,第一个变长列的开销为 5字节,其余每个变长列增加 1字节开销。在 SQL Server 2000中,含有空值的列作为变长列处理,因此,从性能角度考虑,指定为非空 (NOT NULL),并建立默认值。在表 1中, Pub_ date(发表时间)数据类型定义为 datatine(日期型),长度系统定义为 8位。其余字段都是字符内容,为了提高系统性能,将 Abstract字段定义为 varchar(变长字符型)。其存储长度随输入数据的实际长度而变化,但最大长度不得超过 8000字符。其余字段定义为定长型。

4.3 优化索引

在 SQL Server 2000中有两种索引:簇索引和非簇索引。合理使用簇索引和非簇索引都可以加快处理速度,但占用的空间和维护的开销不同。可以通过调整填充因子 (FILL FACTOR)来设置索引选项。通过优化索引可以加快数据库访问的速度 [5]。在条件表达式中经常用到的不同值较多的列上建立索引;对不经常连接的字段由查询优化器自动生成索引。具体设置时,在 SQL Server 2000中,打开 Literature库,右击 Book表,进入"所有任(金)管理索引之"。选择"Qknetten","\$Perkmen Electronic thor"、"Keyword"和"Pub_date"五个字段建立索引,并设置"填充因子"=80(图 8)。

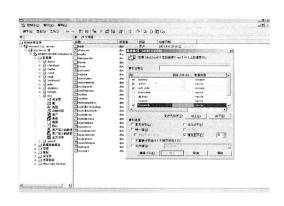


图 8 设置索引

5 结 语

Web文献检索系统首先在局域网内进行设计、调试。在服务器上安装 Windows 2000 Server 并通过设置 "管理工具\Internet服务管理器"建立 IIS服务,设置服务器 IP地址和相应的服务器主目录。在服务器上安装 SQL Server 2000 企业版作为后台数据库,在客户机上安装 SQL Server 2000 客户端,并建立同服务器的连接,可直接操作服务器上的数据库。这样在局域网中就可以通过客户机直接访问服务器。在客户机上编写好相应的查询页面和 ASP程序再上传到服务器的主目录,然后在客户机上执行,调试。效果同在服务器上直接进行操作一样。通过调试运行,Web文献检索系统运行正常,查询效率高。基于ASP、SQL Server 2000实现的 Web文献检索系统简单、安全、稳定,维护方便。

参考文献

- [1] 廖信彦, Active Server Pages应用大全—— ASP 与数据库的整合 [M],清华大学出版社, 2000
- [2] 罗亚, 胡昌盛, "网络数据库系统的开发策略 [J]", 《现代计算机》, 2000, 4:15~22
- [3] 钱文波,谢金宝,"SQL Server数据库性能优化技术 [J]",《微型机与应用》,1999,18(3):7~9.
- [4] 向中凡等, SQL Server 2000 实用操作与技巧 [M], 西安电子科技大 学出版社, 2001.
- [5] 王爱军,基于数据库查询过程优化设计[J]",《电子科技大学学报》,2003,32(2):192~194

(上接第3页)

据库,基于医药卫生科学数据共享服务网培训少数民族地区医师和群众。

7 展 望

"十一五"期间,在国家科技基础条件平台——科学数据共享工程项目的带动下,我们将进一步整合国家、地方和军队医药卫生科学数据共享资源,充分调动数据拥有单位和数据拥有者的积极性,形成医药卫生领域科学数据共享氛围。建立起该领域科学数据资源体系、标准化体系、技术支撑体系、数据服务体系和综合保障体系。形成物理上合理分布、层次分明,逻辑上高度统一、充分共享的医药卫生科学数据共享网,包括6个物理中心、40个左右主体数据库和300个左右数据库和若干个数据集或数据集系列,学科覆盖面达到80%。为政府卫生决策、科技的新、医疗保健和百姓健康信息咨询提供数据共享和信息服务。