

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

宋 阳 严 平 曹 彤

(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)

摘 要 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 Web 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

IMPLEMENTING WEB LITERATURE RETRIEVAL SYSTEM WITH ASP, SQL SERVER 2000 AND THE QUERY OPTIMIZATION

Song Yang Yan Ping Cao Tong

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

Abstract This paper introduces the fulfilment of web literature retrieval system with ASP, SQL Server 2000 and the query optimization. By 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.

Keywords ASP SQL Server 2000 Web literature retrieval system Query optimization

0 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。

1 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。

ASP 是一种基于服务器端的动态网页技术, 它提供了一种使镶嵌在 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 Windows 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 Windows 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 浏览器 / Web 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“沙区农田、草地土壤风蚀防治技术研究”课题 (2002BA517A10); 国家自然科学基金项目 (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。

后分别输入登录 ID 与密码。

3.3 设计 Web 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 Macromedia 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”、“Keyword”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。

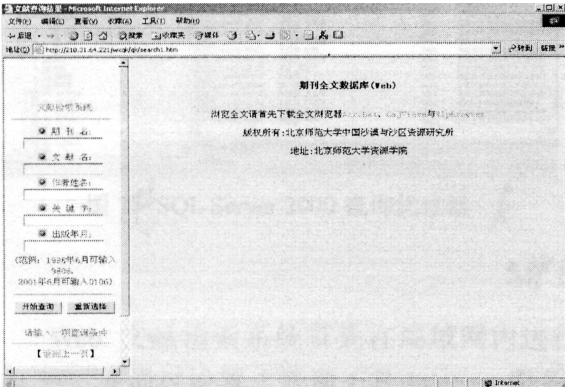


图 5 用户检索页面

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 后台数据库,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
    Dim qkname bookname author pub_date keywords
    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 wc SQLstr 指定条件栏位总数、通用字元、SQL 字符串
    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
    bookname=Request Form ("bookname") author=Request Form ("author")
    pub_date=Request Form ("pub_date") keywords=Request Form ("keywords")
    SQLstr="select * from book where " 执行 SQL 查询语句
    If Len (qkname) > 0 Then SQLstr=SQLstr & "qkname like '" & qkname & wc & "'
    If cond_count>1 Then SQLstr=SQLstr & "and " cond_count=cond_count-1
    End If
    End If
    If Len (bookname) > 0 Then SQLstr=SQLstr & "bookname like '" & bookname & wc & "'
    If cond_count>1 Then SQLstr=SQLstr & "and " cond_count=cond_count-1
    End If
    If Len (author) > 0 Then SQLstr=SQLstr & "author like '" & author & wc & "'
    If cond_count>1 Then SQLstr=SQLstr & "and " cond_count=cond_count-1
    End If
    If Len (keywords) > 0 Then SQLstr=SQLstr & "keywords like '" & 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;UID=guest;PWD=guest;database=Literature"
    执行 SQL 指令,并将结果储存于 Recordset 中
    set rs=adocon.Execute (SQLstr)
    If rs.EOF Then
        Sub Output_Literature() 显示查询结果 (程序代码略)
        End if
    End sub
```

```
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 " cond_count=cond_count-1
End If
End If
If Len (Pub_date) > 0 Then SQLstr=SQLstr & "Pub_date like '" & wc & Pub_date & wc & "'
If Cond_count>1 Then SQLstr=SQLstr & "and " cond_count=cond_count-1
End If
End If
If Len (keywords) > 0 Then SQLstr=SQLstr & "keywords like '" & wc key-
words & 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;UID=guest;PWD=guest;database=Literature"
执行 SQL 指令,并将结果储存于 Recordset 中
set rs=adocon.Execute (SQLstr)
If rs.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“Open”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。

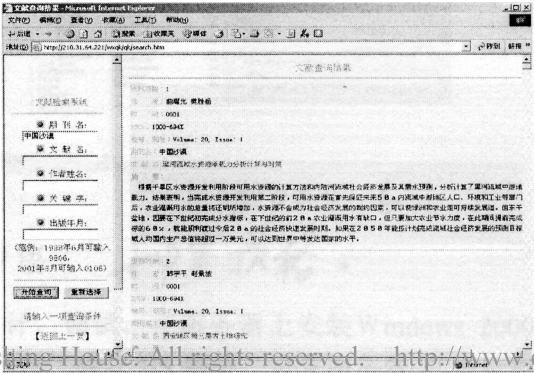


图 6 文献查询结果页面

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。

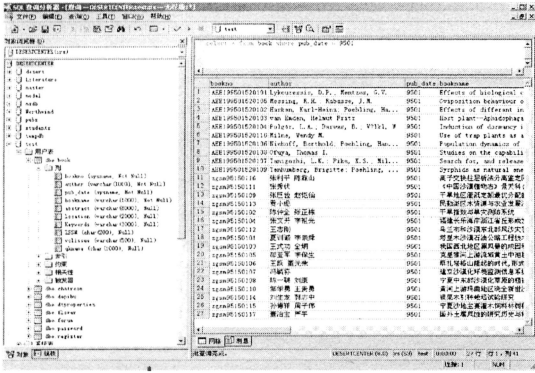


图 7 SQL Server 2000 查询优化器

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 datetime (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“所有任务\管理索引”,选择“Okname”、“Book name”、“Author”、“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)

数据库, 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 and 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、科技创新、医疗保健和百姓健康信息咨询提供数据共享和信息服务。