

基于 ASP 的书目智能检索程序的设计与实现

黎邦群

[摘要]在分析书目数据库数据结构和研究用户输入的检索关键词的基础上,利用 ASP 技术,采用 OLE DB 方式连接书目数据库,即 SYBASE,通过 SQL 语句检索 SYBASE,并将检索结果集分页返回给用户,最终实现了基于 ASP 的书目数据库智能检索程序。

[关键词]智能检索;检索程序;书目数据库

[中图分类号]G254.9

[文献标志码]A

[文章编号]1671-3982(2012)01-0053-03

Design and implementation of bibliography intelligent retrieval program based on ASP

LI Bang-qun

(Library of Huizhou College, Huizhou 516007, Guangdong Province, China)

[Abstract]Based on the analysis of bibliography database structure and user-input retrieval key words, the bibliography intelligent retrieval program was successfully designed and implemented based on ASP, by linking the bibliography database with OLE DB, namely SYBASE which was retrieved using the SQL statement, and returning the retrieval results to the users.

[Key words]Intelligent retrieval; Retrieval program; Bibliography database

随着因特网的发展,信息技术在图书馆的应用日益广泛,数字化图书馆已经成为未来图书馆的发展方向^[1]。馆藏书目检索不断发展,现已成为展现馆藏特色、拓展读者服务的一种高效率手段。由于馆藏书目数据不断增加,检索用户不断增多,且使用习惯差异较大,他们对目前馆藏书目检索程序和检索效果并不满意。从本馆 SULCMIS 系统书目检索程序的应用及反馈情况来看,主要存在检索速度慢、漏检率高、过于专业化、容错性低等问题,严重影响了读者的检索体验与检索效果,需要改进。从检索用户角度看,检索程序应具有高效性、通用性与智能性。高效性是指查得快且准,通用性是指检索方法类似通用搜索,智能性是指具备一定的容错能力。本文拟建立一个书目智能检索程序,取代原有的检索系统。

1 设计思路

[作者单位]惠州学院图书馆,广东惠州 516007

[作者简介]黎邦群(1976-)男,广东惠州人,本科,馆员,发表论文章数篇。

1.1 基本原理

ASP(Active Server Pages)是一种服务器端的网页设计技术,能很轻松地读取数据库的内容^[2]。B/S 架构的系统具有易部署、易更新、方便用户操作等优点^[3]。因此,本检索程序将利用 ASP 技术,采用 OLE DB 方式连接图书集成管理系统 SULCMIS 的书目数据库 SYBASE 中的相关表,在分析相关数据特征的基础上,设计基于 B/S 架构的书目检索程序。

1.2 数据结构

书目检索程序必须同时查询 SULCMIS 书目数据库中的 2 个表,即书目索引表 B_Brief、索书号索引表 B_CallNo。这 2 个表中可以作为检索项字段的数据结构与特征见表 1。根据待检字段的数据结构与特征,我们可以优化检索程序的 SQL 语句,提高检索的速度与准确性。

1.3 系统结构图

书目检索程序检索时,大致可以分为如下 3 个流程,如图 1 所示。即读者输入检索词,程序进行数据特征分析并执行相应的检索策略,最后将结果集分页后呈现给读者。

表 1 书目索引表 B_Brief 与索书号索引表 B_CallNo 结构

字段名	所属表	数据类型	含 义	特征分析
CtrlNo	B_Brief B_CallNo	int	图书 Id	索引; 外键
F01x	B_Brief	char(20)	ISBN 号	字符个数为 13 或 17; 以“978-7-”或“7-”打头
F200	B_Brief	varchar(255)	题名和责任者	含题名、责任者 2 个检索项
F210c	B_Brief	varchar(80)	出版社	以“出版社”3 个字符结尾
CallNo	B_CallNo	varchar(255)	索书号	索引; 以英文字符打头, 数字结尾, 且中间含字符“/”

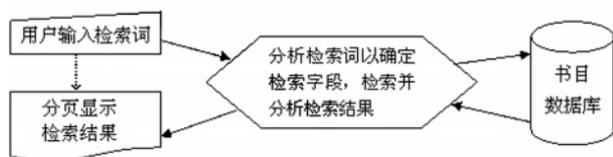


图 1 检索流程图

2 系统实现

2.1 连接数据库

程序采用 OLE DB 方式连接书目数据库 SULCMIS, 需要在服务器预装 Provider for OLE DB. 连接代码如下:

```
Set Cn=New Adodb. Connection
```

```
Cn. Open" Provider=Sybase. ASEOLEDBProvider. 2; Server Name = SULCMIS; Server Port Address = 5000; Initial Catalog = sulcmis; User ID= * * * ; Password= * * * ;"
```

2.2 分析检索词并实现跨库检索

通过分析用户输入的检索词确认用户的检索目的, 从而查询不同的字段, 避免查询所有字段。这样不但可以节约有限的系统资源, 提高检索速度, 而且可以提高检索的准确度。

假如输入的检索词为“q”, 其 SQL 检索语句如下:

```

For n=0 to Int( Ubound( Split( q , " " ) ) )
w=Trim( Replace( Split( q , " " ) ( n ) , " " , "" ) )
If Len( w ) > 0 Then
If w>0 And w<1000000 Then
Sql=Sql & " And B_Brief. CtrlNo= " & w
ElseIf ( Left( w 2 ) = " 7 - " Or Left( w 4 ) = " 978 - " )
And ( w = 13 Or w = 17 ) Then
Sql=Sql & " And B_Brief. F01x Like " & w & " "
ElseIf Right( w 3 ) = " 出版社 " Then
Sql=Sql & " And B_Brief. F210c Like " & w & " "
ElseIf Asc( w ) > 64 And Asc( w ) < 123 And ( Len( w ) < 3 Or

```

```

InStr( w , "/" ) > 0 Or Isnumeric( Mid( w , 2 , 1 ) ) Or Isnumeric
( Mid( w , 3 , 1 ) ) Or Isnumeric( Right( w , 1 ) ) ) Then
Sql=Sql & " And B_CallNo. CallNo Like " & w & " % "
Else
Sql=Sql & " And B_Brief. F200 Like " & w & " % "
End If
Next

```

假如输入 0-1000000 之间的数字, 仅检索字段 CtrlNo, 该字段为表索引。这样, 即使检索上千万条记录的书目库, 耗时也不到 1/10 秒。当用户用空格符号分割检索词并进行“逻辑与”检索图书的 ISBN、出版社、索书号(包括分类号)、题名或责任者时, 智能识别其中的题名和责任者之外的检索项, 将成倍提高检索效率和准确度。如果检索词置于英文双引号中, 则不进行分词, 仅检索题名或责任者:

```

If Left( q , 1 ) = " " And Right( q , 1 ) = " " Then Sql="
And B_Brief. F200 Like " & Replace( q , " " , "" ) & " % "

```

实现跨库检索, 去掉重复的记录, 并按图书的新旧排序。

```

Sql=" Set RowCount 5000 Select Distinct B_Brief. CtrlNo , B
_Brief. F01x , B_Brief. F200 , B_Brief. F210c , B_CallNo. CallNo
From B_Brief , B_CallNo Where B_Brief. CtrlNo = B_CallNo.
CtrlNo" & Sql & " Order By B_Brief. CtrlNo Desc"

```

检索结果还可以按照图书的出版时间等依据排序, 只需相应地更改“Order By”的参数即可。

2.3 分析检索结果

如果检索结果的记录数为 0, 程序将自动进行二次检索, 将建议结果返回给用户。二次检索仅针对题名或责任者项进行检索策略的修改, 并对检索词依据长度分类进行分词检索, 即实现其容错性。其核心语句如下:

```

If Len( w ) > 9 Then
If InStr( Sql , w ) = 0 Then Sql=Sql & " And B_Brief. F200
Like " & w & " % "

```

```
Else
For i=1 to Len( w)
Sql=Sql &" And B_Brief. F200 Like '%" & Mid( w ,i ,1)
&"%"
Next
End If
```

可以根据书目数据的特性及用户检索习惯,自行设定二次检索的检索词分类长度,尽量避免检索失败。

2.4 分页显示检索结果

用户的查询提交到检索系统中,通过对查询和索引的相似度进行计算,完成结果查询和排序,将最终结果反馈给用户^[4]。利用 Replace 函数将记录中的关键词替换突出显示,并对结果进行分页。

假如 RecordSet 对象为 Rs,页码数为 p,则其分页的语句如下:

```
Rs. Pagesize = 20
pagecount = Rs. Pagecount
Rs. Absolutepage = p
If pagecount > 1 Then
q = Server. UrlEncode( Request( "q" ) )
If p > 1 Then Response. Write " <a href=? q=" &q&" &p
=" &p-1&" >上一页</a> "
i = 1
Do While pagecount > 0 And i < p + 10
If i = p Then
Response. Write i
ElseIf i < > p And i > p - 11 Then
Response. Write " <a href=? q=" &q&" &p=" &i&" >"
&i&" </a> "
End If
i = i + 1
pagecount = pagecount - 1
Loop
If p < Rs. Pagecount Then Response. Write " <a href=? q
=" &q&" &p=" &p+1&" >下一页</a> "
End If
```

即通过 If 语句进行判断,当页码数“pagecount”大于 1 时,将检索结果进行分页呈现,每页显示 20

条书目记录。

2.5 实现效果

书目检索程序利用 ASP 技术通过 OLE DB 方式连接 SULCMIS 书目数据库 SYBASE,通过分析检索词的数据特征,检索其书目表 B_Brief 及 B_Call 中的相关字段,找到匹配的结果。首先,通过分析用户输入的检索词定位用户的检索目的,可避免书目库的全字段扫描,无疑会提高检索效率。原程序检索任意词“红楼梦”需要 3.9 秒,而该程序仅需 1 秒左右。可见,在同样的网络及系统环境下,该程序比原 SULCMIS 书目检索系统的平均检索速度提升了 3 倍多。其次,通过分词等检索策略,使程序具备了一定的容错性,优化了用户的检索体验,有效激发了读者的检索兴趣。

3 结语

本文分析了书目数据库的表结构及字段特征,并在此基础上分析了用户输入的检索词的数据特征,利用 OLE DB 方式连接 SULCMIS 书目数据库 SYBASE,从而实现与 SULCMIS 系统的无缝连接,并采取了分词等检索策略,对检索的结果集进行分页,实现了基于 ASP 的书目数据库智能检索程序。整个检索页面就一个检索框,检索方法与通用搜索引擎类似,具有易用性。从应用效果来看,该程序原理简单,结构清晰,具有一定的高效性、通用性与智能性,可供图书馆、网上书店参考。

【参考文献】

- [1] 郭 蕾. 基于 PHP 的图书馆书目检索系统的设计与实现[J]. 邯郸职业技术学院学报, 2009, 22(1): 92-96.
- [2] 成 悦, 周 建. 用 ASP 实现图书馆书目检索[J]. 南通航运职业技术学院学报, 2003, 2(4): 38-44.
- [3] 蔡昭权. 利用 Web Service+Java Script 实现无刷新的数据交换[J]. 惠州学院学报: 自然科学版, 2005, 25(3): 33-36.
- [4] 韩中元. 基于 SQL Server2008 的小型信息检索系统框架的设计与实现[J]. 电脑编程技巧与维护, 2010(10): 44-46, 58.

[收稿日期: 2011-05-12]

[本文编辑: 王 颖]