

一种校内网资源搜索引擎的设计与实现

汪益民, 杨春节

(安徽农业大学 网络中心, 安徽 合肥 230036)

摘要: 搜索引擎是目前人们寻找网络资源最常用的方式, 本系统依照搜索引擎的设计原理, 采用多 Agent 方式, 通过对索引模块参数设置实现对校内网不同的资源进行统一整合, 建立校内网资源的搜索引擎。该搜索引擎为学校内网提供资源共享的新方式, 提高校内资源的利用率, 减轻校内网带宽的压力。

关键词: 搜索引擎; Agent; 校内网资源; 校内网带宽

中图分类号: TP393

文献标识码: A

文章编号: 1674-7720(2011)15-0018-02

The design and implementation of campus network resources search engine

Wang Yimin, Yang Chunjie

(Network Center of Anhui Agricultural University, Hefei 230036, China)

Abstract: Search engine is the most common way to seek network resources, this system was designed in accordance with the principles of search engine with multi-Agent approach, established search engine of school network resources by setting the index module parameter of school network to achieve an unified integration of different resources. It offers schools a new way to share network resources, improve the utilization of campus resources, reduce pressure on the school network bandwidth.

Key words: search engine; Agent; resources of campus network; campus network bandwidth

随着高校校园网的快速发展, 网络作为信息的传播平台正在高速向前迈进。学校校园网中各种资讯、软件、多媒体资源越来越丰富并且呈现出爆发式的增长。以本校为例, 目前软件资源容量为 300 G、音乐资源容量为 4 T、电影资源容量为 9 T、学习资源容量近达 1 T 等, 这些信息资源帮助人们更好地完成工作与学习、丰富业余生活、提供更加快速方便的下载方式。但是这是数字资源分布在不同服务器与存储中, 如何更加有效地利用这些海量的校内网资源成为信息管理者深思的问题。人们在互联网上查找信息都会用到搜索引擎来搜索需要的信息, 通过搜索引擎的模糊匹配机制, 能很容易找到想要的信息。对于学校信息中心信息管理人员来说, 如果能在校内网中引入一套搜索引擎机制来解决这个问题, 就能使广大师生能快捷便利地使用这些信息而且减少重复从外网下载、降低校内网出口带宽的压力、充分提高校园网内部资源的利用率的问题。

1 需求分析

搜索引擎原理, 通常指的是收集了因特网上几千万到几十亿个网页并对网页中的每一个词(即关键词)进行

索引, 建立索引数据库的全文搜索引擎。当用户查找某个关键词的时候, 所有在页面内容中包含了该关键词的网页都将作为搜索结果被搜出来。在经过复杂的算法进行排序后, 这些结果将按照与搜索关键词的相关度高, 依次排列。

本系统主要目标是仿照搜索引擎的原理实现对校内网信息资源的全面搜索。这些资源可能是不同类型, 可能分别存在不同服务器中且服务器的类型也不近相同, 如有 FTP 服务器、HTTP 服务器等。对于这些不同的环境要求, 要求本系统要完成以下需求:

- (1) 索引建立要满足不同的服务器类型, 如 http、ftp 等。
- (2) 索引建立时要区分文件的类型。
- (3) 定时(手工)删除本机器的索引, 并重新建立指定目录的文件索引。
- (4) 搜索能支持模糊查询。
- (5) 提供用户搜索、下载功能, 记录关键字频率及文件访问次数。

对于以上需求, 本系统采用多 agent 模式, 对不同服

《微型机与应用》2011 年第 30 卷第 15 期

服务器,分别安装索引建立模式(agent),再通过对参数的配置实现对不同服务器中资源的索引建立。

2 系统分析

2.1 系统结构

本搜索引擎结构如图1所示,软件的设计与开发主要为两个部分:(1)索引的建立。主要负责在后台数据的建立,采用客户端与服务器结构(C/S结构);(2)Web的查询。主要负责向用户提供友好界面的查询服务,采用浏览器与服务器结构(B/S结构)。通过开发索引建立程序,将文件存放的服务器(FTP、HTTP等)进行搜索,分析文件的名称、大小、类别、位置等并将这些文件信息合并服务器IP、访问的前缀等建立对应的索引添加到数据库服务器中。

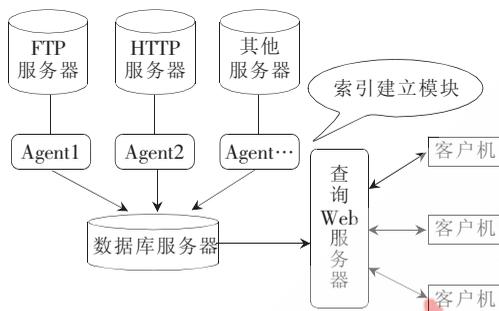


图1 搜索引擎结构图

2.2 数据库设计

为了方便数据库中文件信息的处理,数据库设计了两个基本数据表,一个存放索引建立模块,将文件的索引放入表1中;另一个表主要存放用户搜索的相关信息,如表2所示。如果搜索数据非常庞大时,要对基本数据表按不同的规则进行分割,成立多个子表(可以按照服务器类型、文件类型或者服务器所属网络等分类),实现对资源的高效率搜索。

表1 ServerFiles

| 列名 | 数据类型 | 长度 | 允许空 | 备注 |
|----------|-------|----|-----|--------------------|
| ID | 序号 | B | 8 | 自增 |
| Ip | 服务器IP | C | 20 | 服务器地址 |
| Front | 前缀 | C | 100 | 访问地址的前缀 |
| folder | 目录 | C | 100 | 文件所在目录,形如“影视/电视剧/” |
| Filename | 文件名 | C | 120 | 文件名称(含扩展名) |
| Filetype | 文件类型 | C | 10 | 扩展名 |
| filesize | 文件大小 | C | 30 | |

表2 SearchMemo

| 列名 | 数据类型 | 长度 | 允许空 | 备注 |
|---------|-------|--------|-----|----------------|
| ID | 序号 | Bigint | 8 | ServerFiles.id |
| KeyWord | 搜索关键字 | C | 30 | |

2.3 系统设计

索引建立模块(agent)的界面上参数主要包含服务器IP地址、服务器前缀、选择搜索的路径、设置定时时间

等。服务器相关的文件是以海量规模存储的,为了提高索引建立的效率,系统采用后台的存储过程来实现对数据库的操作。

索引建立模块是整个系统的核心,通过递归算法,将服务器中指定文件夹中所有文件(或者指定类型文件)的信息依次提取,将文件物理地址转换为逻辑地址并形成对应关系,按照数据库表结构预先设定的规则放入数据库对应表中,实现对文件服务器中文件索引的建立。搜索的用户程序采用B/S结构,用户通过浏览器搜索关键字实现对资源的查询。搜索的关键字采用对所在分类文件夹和文件名中模糊匹配,将搜索的结果以类别、文件名、文件类型、文件大小等信息来展现。

2.4 主要核心代码

(1) 搜索文件代码

```
if(FindFirst(PathName+"*.*", faAnyFile, sr)== 0)
{
do
{
if((sr.Attr&faDirectory)&&sr.Name!="."
&&sr.Name!="..")
//判断是否是目录,并排除目录“.”和“..”
{
//mem->Lines->Add("目录:"+sr.Name);
//调用函数本身,进入子目录
BrowDir(PathName+sr.Name+"\\", kind);
}
}while (FindNext(sr) == 0);
FindClose(sr);
}
```

(2) 调用存取过程代码

```
ADOStoredProc1->ProcedureName="sp_insert";
ADOStoredProc1->Parameters->ParamByName("@ipstr")->
Value=Trim(Form1->edip->Text);
ADOStoredProc1->Parameters->ParamByName("@frontstr")
->Value= Trim(edfront->Text);
ADOStoredProc1->Parameters->ParamByName("@folderstr")
->Value =PathName.SubString (edpath ->Text.Length () +2,
PathName.Length()-edpath->Text.Length());
ADOStoredProc1->Parameters->ParamByName("@filenamestr")
->Value=sr.Name;
ADOStoredProc1->Parameters->ParamByName("@fileypestr")
->Value= ExtractFileExt(PathName+sr.Name);
ADOStoredProc1->Parameters->ParamByName("@filesizestr")
->Value= sizestr;
ADOStoredProc1->Prepared;
ADOStoredProc1->ExecProc();
```

内网资源搜索引擎对于类似校园网内部资源的整合有很重要的作用,它的重要性表现在能将校园网的各

种资源以统一的搜索引擎的方式提供给用户,方便用户的查找与使用,极大地提高了这些资源的利用率。本文介绍的校内网搜索引擎已经在我校产生的明显的效果,随着推广的深入,广大师生越来越多地都在使用它来寻找自己所需要的资源,极大提高了工作效率并减轻了校园网出口带宽的压力。

参考文献

- [1] 郝金隆,王成良.原创优先的搜索引擎排序算法[J].计算机工程,2008,34(18):85-86.
- [2] 任勇旗,唐毅.以客户为中心的搜索引擎优化研究[J].

图书馆学研究,2009(1):44-45.

- [3] 吴断,施水才.一种结合超链接分析的搜索引擎排序[J].计算机工程,2004,30(15):143-145

(收稿日期:2011-03-07)

作者简介:

汪益民,男,1980年生,硕士研究生,工程师,主要研究方向:嵌入式系统开发与设计。

杨春节,男,1980年生,硕士研究生,工程师,主要研究方向:农业机械化。

