

# 基于 Java 编程的校内搜索软件的设计

孙文华

(南昌工程学院 网络信息中心,江西 南昌 330099)

**摘要:** 如何在各高校内部成百上千个资源站点中,快速有效地查找到所需信息成为网络用户面临的一个突出问题。为此提出了一款校内搜索引擎软件的设计,解决了当前各高校大学生对本校信息难于快速有效获取的问题。

**关键词:** Java;校内;搜索引擎;设计

中图分类号: TP319

文献标识码: A

文章编号: 1674-7720(2012)23-0008-03

## Design of search engine software in school based on the Java programming

Sun Wenhua

(Center of Network Information, Nanchang Institute of Technology, Nanchang 330099, China)

**Abstract:** How in the hundreds of thousands of resource site quickly and effectively find the required information becomes the network users face an outstanding problem. In this paper the author presents a web search engine software design. Solution to the current colleges and universities students of the school of information to effectively access problems.

**Key words:** Java; in school; search engine; design

当前,面对浩瀚的网络资源,搜索引擎为所有网上冲浪的用户提供了一个入口,所有的用户都可以借助于搜索引擎到达自己想去的网上任何一个地方。随着 Internet 的快速发展,海量信息和人们获取所需信息之间的矛盾日益明显,在信息海洋里查找信息,如同大海捞针一样,而搜索引擎的出现正好解决了这一难题。

目前高校校内信息量的不断增加,面对浩瀚的网络信息,选择资源范围广而针对性不强,很难搜到符合自己的资源。针对这种情况,本文提出了一款校内搜索软件的设计与实现方法。该软件采用 Java 语言开发,是一款简易且实用的校内搜索引擎软件。

### 1 设计原理

搜索引擎系统<sup>[1]</sup>通常是指互联网信息检索系统。本系统是建立在通过网络爬虫软件抓取大量的网络资源的基础上进行开发设计的,系统通过对抓取的网页文件进行智能提取、去标注、内容分析等处理,再经过索引加载建立索引数据库。用户可以通过搜索页面查询索引数据库,返回包含所有匹配查询关键词的网页<sup>[2]</sup>。

校内搜索引擎软件的设计原理如图 1 所示,其原理如下:(1)通过网络爬虫软件抓取网页内容;(2)从抓取

的网页文件中获取索引表,索引表内容为每个文件的序号与网址信息,得到网页文件库里面的内容即为网页文件;(3)对得到的网页文件库进行处理,使用正则表达式去除网页标注信息,得到的新内容存放在一个新的文件夹中,作为内容信息库;(4)根据网页文件库和索引表,得到一个新的文件(也称为标题和网址信息库),其里面的内容为每个文件的序号、标题和网址等信息;(5)当用户需要查找自己需要的信息时,只要输入关键词,搜索引擎软件根据用户输入的关键词在内容信息库中进行查找<sup>[3]</sup>,如果内容信息库中存在用户查找的内容,软件将根据给内容所在的文件序号,在标题和网址信息库中提取出该内容所在的标题与网址等信息,最后再加

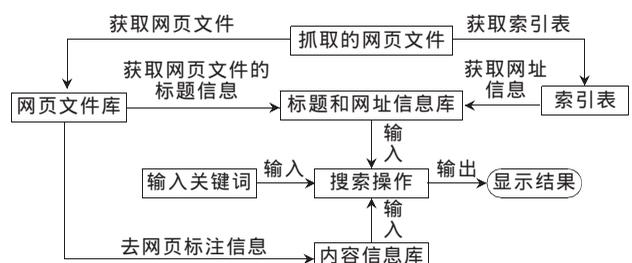


图 1 搜索引擎软件设计原理图

上内容信息库中与关键词相关的内容信息作为查找的结果显示出来。

## 2 具体算法实现

搜索引擎软件开发环境:Myeclipse 平台, 使用 Java 语言。首先可以 Myeclipse 平台新建一个 java project, 在新建的 project 中需要导入下面一些相关的文件:

```
import java.io.BufferedReader;
import java.io.File;
import java.io.FileNotFoundException;
import java.io.FileReader;
import java.io.IOException;
import java.io.InputStreamReader;
import java.util.StringTokenizer;
```

搜索引擎软件中获得搜索结果的搜索函数核心代码如下:

//根据用户输入的关键词,在相应的库中进行搜索,并返回搜索结果信息

```
private static String search(String[] str, int n, String
filepath)
{
String all=null;
File file=new File(filepath);
try{
String[] filelist=file.list();
for(int i=0,flen=filelist.length;i<flen;i++)
{
File readfile=new File(filepath+"\\"+
filelist[i]);
String upname=readfile.getName();
upname=upname.substring(0,upname.length
()-4);//去掉文件名中的.txt用于后面找网址
BufferedReader br=new BufferedReader
(new FileReader(
readfile));
String s="",ss=br.readLine();
while(ss!=null)
{
s=s+ss; //s中存放文件内容的信息
ss=br.readLine();
}
//判断当前读入的记录行中是否有
//输入的关键词,输入了几个关键词
switch(n)
{
case 1: //1个关键词的情况
int end=s.indexOf(str[0]);
if(end!=-1)
{
upname=Url(upname);
```

```
//获取存标题与网址信息
s=s.substring(end,20);
//取关键字后的20个字符
all=all+"\n"+upname+"
关键字内容:"+s;
```

```
flag=1;
break;
} else
break;
```

```
case 2: //2个及多个关键词的情况
.....
```

```
br.close(); //一定要关闭资源
```

```
} catch (FileNotFoundException e)
```

```
{
e.printStackTrace();
} catch (IOException e)
```

```
{
e.printStackTrace();
```

```
}
return all;
```

```
}
```

```
return all;
```

```
}
```

搜索函数中调用的其他主要函数代码如下:

//通过传递的文件名参数,在文件 TitleInfo 中查找得到与此文件名对应的标题信息+网址信息

```
public static String Url(String filename)
```

```
{
```

```
String filepath="D:\\test\\TitleInfo\\TitleInfo.txt";
```

```
File file=new File(filepath);
```

```
String url="";
```

```
try
```

```
{
```

```
BufferedReader br=new BufferedReader(new
```

```
FileReader(file));
```

```
String s=br.readLine(); //读入第一行信息
```

```
while(s!=null)
```

```
{
```

```
if(s.indexOf(filename)!=-1)
```

```
{
```

```
url=s.substring(s.indexOf("标题"),s.length());
```

```
//取该行标题后的信息
```

```
break;
```

```
} else
```

```
s=br.readLine(); //读入下一行的信息
```

```
}
```

```
br.close(); //一定要关闭资源
```

```
} catch (FileNotFoundException e)
```

```
{
```

