

ROSA 搜索使用手册

2009-08-28



1	引言	2
2	系统实现动机	2
3	运行环境	3
4	开发环境	3
5	简单配置的系统部署	6
5.1	所需文件汇总	6
5.2	安装 Java 运行环境	6
5.3	运行系统	7
6	复杂配置的系统部署	7
6.1	数据库安装和配置	8
6.2	Tomcat 安装配置	10
6.3	部署网站	10
6.4	运行索引	12
6.5	ROSAWeb 配置	13
8	一般用户检索	16
8.1	简单搜索	16
8.2	高级搜索	16
8.2.1	文件名或文件路径查询	16
8.2.2	搜索关键字过滤	16
8.2.3	后缀名查询	17
8.2.4	逻辑与或非	17
8.2.5	单点搜索	17
8.2.6	通配符	17
8.2.7	模糊查询	18

8.2.8	短语查询	18
8.2.9	范围查询	18
8.2.10	访问控制	19
8.2.11	高级搜索	19
8.2.12	索引域的使用	19
9	用户高级功能	22
9.1	ROSA FTP	22
9.2	用户资源请求	22
9.3	用户发布资源	22
9.4	用户发布留言	22
9.5	离线站点快照	22
9.6	搜索建议	23
9.7	搜索统计	24
9.8	热门站点排名	24
10	FTP 站点管理	27
10.1	登记 FTP 站点	27
10.2	更改站点信息	27
10.3	自定义共享圈	27
11	ROSA FTP 下载	28
11.1	安装 ROSA FTP	28
11.1.1	安装 JRE	28
11.1.2	启动 ROSA FTP	28
11.1.3	ROSA FTP 的安装位置	28
11.2	ROSA FTP 的卸载	29
12	意见反馈	31

1 引言

ROSA 是一个通用意义上的开源 FTP 搜索引擎，我们借助 Lucene，Struts 及其他开源软件的帮助，设计实现了一个基于当前搜索引擎技术的 FTP 搜索引擎。相对于校内以前的搜索引擎，ROSA 的搜索灵活性更高，更符合当前用户的自定义搜索需求。并且 ROSA 采用分布式设计思路，扩展性更好。面向对象的设计和前后台设计分离，使 ROSA 的部署，升级和后续功能的实现变得较为容易。图??给出了 ROSA 搜索引擎的首页。ROSA 的校园网网址为 <http://askrosa.cn>，公网访问地址为: <http://nju.askrosa.cn>¹。



1: ROSA 首页

2 系统实现动机

现有的校内搜索引擎 Thephy 存在的问题:

1. 不支持复杂搜索，只是简单的字符串匹配。这样的搜索匹配效率不高，最严重的缺点是不能支持用户的自定义搜索，比如用户只想搜后缀名为 rmvb 的某个视频，现有的搜索引擎 Thephy 是无能为力的。

¹不同于 <http://askrosa.cn> 的 A 地址，此地址使用的 URL 转发实现

2. 搜索结果冗余。比如对于“/movie/ 蓝莓之夜 /1.rmvb”这样的资源，当搜索“蓝莓之夜”时结果会有两条：

- /movie/ 蓝莓之夜
- /movie/ 蓝莓之夜 /1.rmvb

当用户输入的关键词匹配条目很多的时候，这种冗余会带来很大的混乱，迫使用户浪费很多时间来区分条目。

3. 现有的搜索引擎不支持任何形式的排序。对于很多资源，比如软件，电影，用户往往喜欢选择最新的资源下载，在没有排序支持的情形下需要用户人工区分资源更新时间。
4. 考虑到同一关键字可能会同时出现软件，电影等资源名称中，所以分类搜索也是用户十分需要的功能。
5. 现有的搜索引擎用专门的论坛负责站点登记，当用户想把自己的站点登入的时候，要在论坛发帖，然后由管理员手工添加。这种方式过于浪费人力和时间。我们需要一个 FTP 服务器站点的自主管理平台。
6. 现有的搜索引擎无法确定结果页中的站点是否可以访问。我们希望能搜索结果页面中给出站点连接性能提示，以使用户选择。

ROSA 搜索引擎要实现的改进：

1. 使用当代搜索引擎技术，抛弃简单的字符串匹配，而使用基于分词的倒排表保存索引数据和提供搜索服务。
2. 细致的可用搜索域，可以满足用户各种搜索需求。
3. 针对常用搜索域的升序降序排列。方便用户排序搜索资源。
4. 提供针对路径名和文件名的多种搜索方式，以使搜索结果简洁美观。
5. 对资源根据文件后缀进行自动分类，并提供相应的搜索接口，方便用户进行分类搜索。
6. 自主的站点登录、管理和登出系统。方便 FTP 管理员管理站点。
7. 资源访问和站点访问统计，方便用户分享各自的兴趣和选取热点站点。
8. 站点可连接性提示，方便用户选择站点下载资源。
9. 多点下载程序，减轻用户站点选取的负担，因为用户关注的是资源而非站点。所以我们需要一个多点下载程序可以根据用户选取的资源信息同时从多个站点来分片下载资源。

3 运行环境

操作系统 由于使用 Java, JSP 和 MySQL 开发, 所以 Windows 或 Linux 皆可部署 ROSA 系统。

索引程序 JDK1.6 以上或者 JRE 1.6 以上。

数据库 MySQL-5.0 以上。

JSP 容器 Apache-tomcat-6.0.16 以上。

4 开发环境

如果需要对 ROSA 进行二次开发, 除了前面提到的运行环境之外, 还需要

开发工具 开发工具为 MyEclipse 6.0.1, VisualSVN Server 2.0.6 和 Subclipse 1.6.5。

Java 虚拟机 JDK1.6 以上

5 简单配置的系统部署

5.1 所需文件汇总

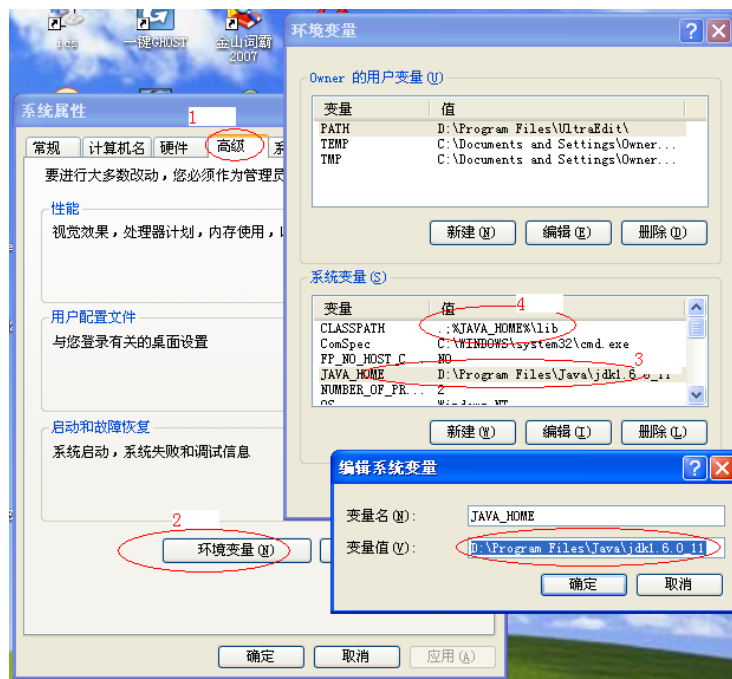
1. JRE 1.5 以上。
2. MySQL 5.0 以上数据库。
3. MySQL GUI Tools 5.0 以上的 MySQL 数据库图形管理工具。
4. Apache-tomcat-6.0.16 以上网站容器。
5. RosaWeb 网站代码。
6. RosaCrawler 爬虫程序。

简单的安装运行是指运行我们配置好的系统。我们给出上面列出的全部资料。用户无需下载 MySQL 和 Tomcat 等，但是用户的运行所需的文件是最大的。这里要注意的是，简单安装运行，仅限本地运行，如果将数据库，Tomcat 和索引程序放在不同的机器运行，需要按照复杂安装中的配置好 ip 地址信息。

5.2 安装 Java 运行环境

由于我们的系统是基于 Java 的，所以按照 JRE 是必须的，如果需要开发，则需要安装 JDK。

1. 下载并安装 JRE 1.5 以上的版本
2. 设定环境变量



5.3 运行系统

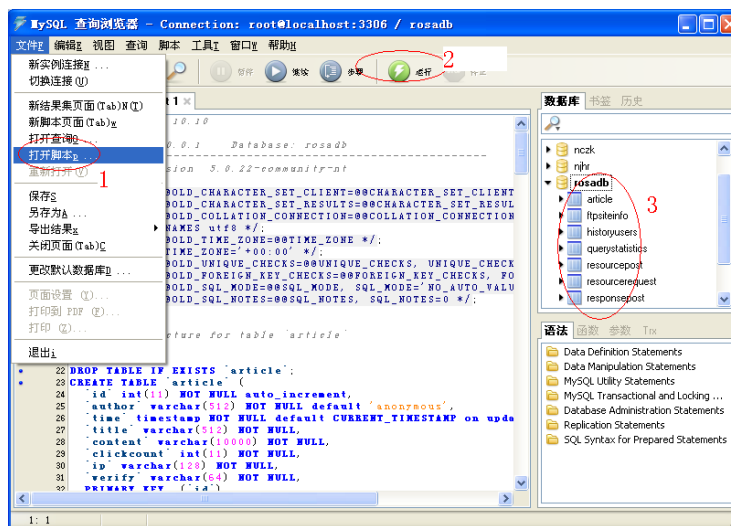
1. 运行 `mysql-5.1.34-win32\bin\mysql.exe`，启动数据库
2. 运行 `apache-tomcat-6.0.16\bin\startup.bat` 启动 Tomcat
3. 在浏览器输入 `http://127.0.0.1:8080/RosaWeb` 可以看到运行界面了
4. 点击添加站点，可以添加一个测试站点，测试用户可以根据自己的情况，添加一个 FTP 站点。
5. 运行 `Rosa\RosaCrawler\crawler.bat` 执行爬虫程序。此时，程序会访问登记的 FTP 站点来获取 FTP 资源信息，并在 `RosaCrawler\index` 中生成索引文件。
6. 在爬行结束之后(不要关闭 `crawler.bat`)，在浏览器中就可以搜索 FTP 站点的资源了。

6 复杂配置的系统部署

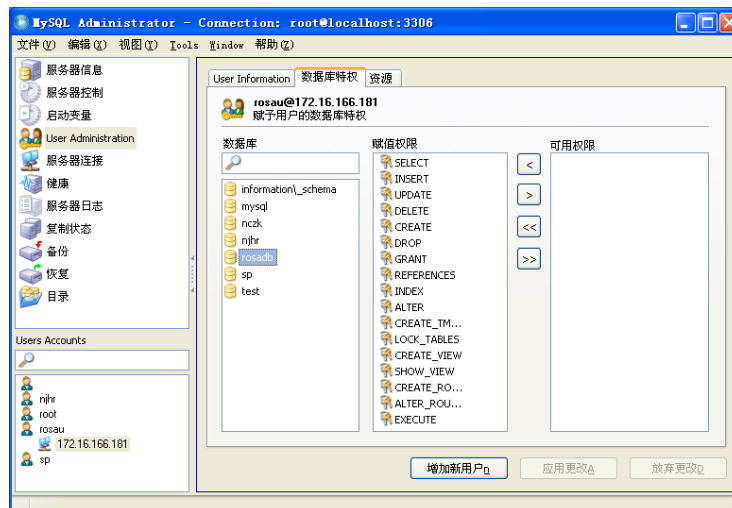
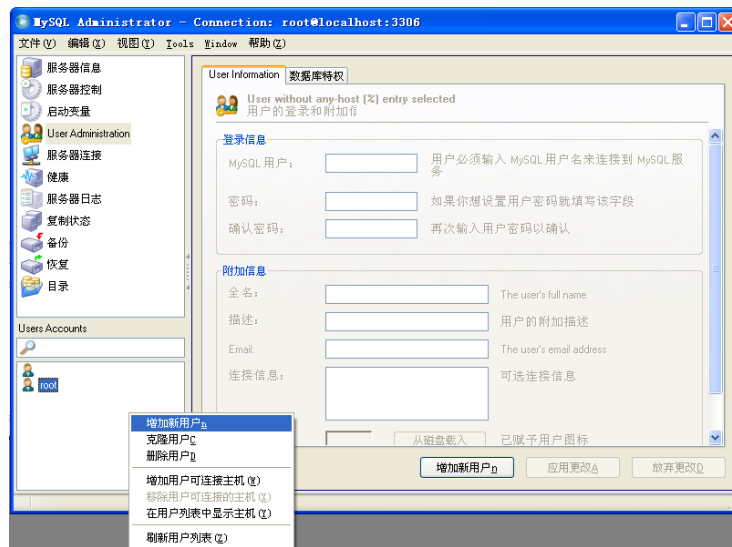
此时，需要的安装文件是最小的，仅需 Rosa 目录中的 RosaWeb 网站代码和 RosaCrawler 索引程序。

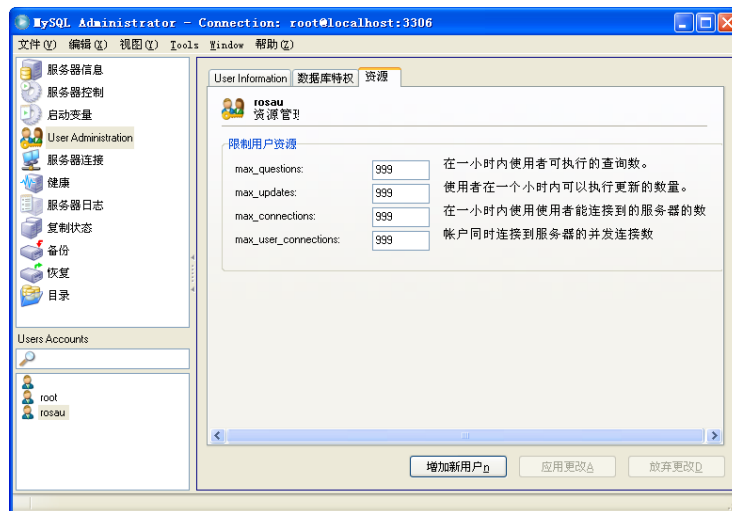
6.1 数据库安装和配置

1. 下载 MySQL 5.0 以上数据库安装，并正确配置。
2. 运行数据库。这里我们假定本地数据库，运行的 IP 为：127.0.0.1。
3. 利用数据库工具 MySQLQueryBrowser 创建数据库 rosadb（数据库名字自己可以定，但在之后的配置文件中需对应）。
4. 利用数据库工具 MySQLQueryBrowser，设置默认数据库为 rosadb，然后导入脚本 rosa.sql



5. 利用数据库工具 MySQLAdministrator 创建可网络访问的用户名 rosau，密码 rosap，（用户名密码自己可以定，但在之后的配置文件中需对应）给予此用户数据库 rosadb 的所有权限。





6.2 Tomcat 安装配置

1. apache-tomcat-6.0.16 以上安装，并正确配置。
2. 配置文件 /conf/server.xml，在加上一句 URIEncoding="UTF-8"。使之支持 UTF-8 编码的传输

```
<Connector port="8080" protocol="HTTP/1.1"
connectionTimeout="20000"
redirectPort="8443" URIEncoding="UTF-8"/>
```

6.3 部署网站

1. 将文件夹 RosaWeb 放入 apache-tomcat-6.0.16\webapps 中
2. 更改 RosaWeb 中的配置
配置文件的位置在: RosaWeb\WEB-INF\classes\resource\ProjectSetting.properties
样例配置如下:

```
# 更改数据库的地址，这里是本地运行
#database configure
database.url=jdbc:mysql://127.0.0.1:3306/rosadb
database.username=rosau
database.password=rosap
database.driver=org.gjt.mm.mysql.Driver
```

```
# 更改搜索建议索引节点的 ip 地址，索引节点是下面运行的 RosaCrawler
的运行地址，这里是本地运行
autocomplete.host=127.0.0.1
autocomplete.port=1225
autocomplete.name=ftpspellchecker
autocomplete.rows=12
```

用户可以根据这个配置来配置你的数据库名称，用户名和密码，但 IP 地址是必须修改的，建议使用 MySQL 数据库，使用其他数据库的情况虽然考虑了，但没有测试过。

3. 配置爬虫节点

配置文件位置：RosaWeb\WEB-INF\classes\resource\crawlernodes.xml
这里配置索引节点的位置，此地址是 RosaCrawler 所在机器的地址，此处假设是本地测试，所以地址为 127.0.0.1。是这里可以配置多个索引节点，这里我们就设置一个索引节点。

```
<?xml version="1.0" encoding="UTF-8"?>
<nodes version="1.0">
  <node>
    <name>rosa</name>
    <address>127.0.0.1</address>
    <port>1225</port>
  </node>
</nodes>
```

4. 运行 Tomcat，这里我们假定 Tomcat 也是在本地运行的。

5. 增加 FTP 站点

现在 RosaWeb 已经可以运行了，我们输入 <http://127.0.0.1:8080/RosaWeb> 就可以看到如下截图：



点击添加站点，可以添加一个测试站点，测试用户可以根据自己的情况，添加一个 FTP 站点。

增加站点

基本选项：

地址*：

用户名：

密码：

验证码*：

验证码确认*：

[更多高 级选项:]

管理员：

联系方式：

端口：

更新周期：

限速(KB)：

用户数：

描述：

编码：

站点位置：

访问控制：

FTP域名或IP地址

FTP用户名，默认anonymous，如果允许匿名访问，请不要改动

FTP密码，默认anonymous，匿名访问无需密码

为保证只有注册的人具有修改FTP信息的权限，用验证码来确认

验证码确认

FTP管理员，默认为UNKNOWN，注册者更具自己的情况命名

管理员联系方式，邮箱、百合账号或其他

FTP端口号，默认21

索引程序对此站点进行索引的周期，默认为1天

FTP速度限制，默认0，表示不限速

FTP用户数限制，默认0，表示不限用户数

FTP描述信息，FTP内容和作用等

FTP传输编码，默认GBK，可选的有UTF-8和GB2312

FTP站点的位置，根据您的填写，我们采用不同的访问方式

默认anybody，表示FTP可以被任何人检索到，其他字符串用户在检索时需提供此字符串

提交

注意：为了防止错误提交，提交的时候，必须保证FTP站点能够登陆进入，否则不会收录！如果登记不上，可联系我们帮您登记。

6.4 运行索引

1. 索引程序在文件夹 RosaCrawler 中。并且这里假定在本地运行。因为上面的配置文件会用到这个地址。如果地址发生改变，在配置文件中要做相应的更改。
2. 修改配置文件
配置修改如6.3节中的2，3部分一样，主要更改数据库服务器的地址和索引

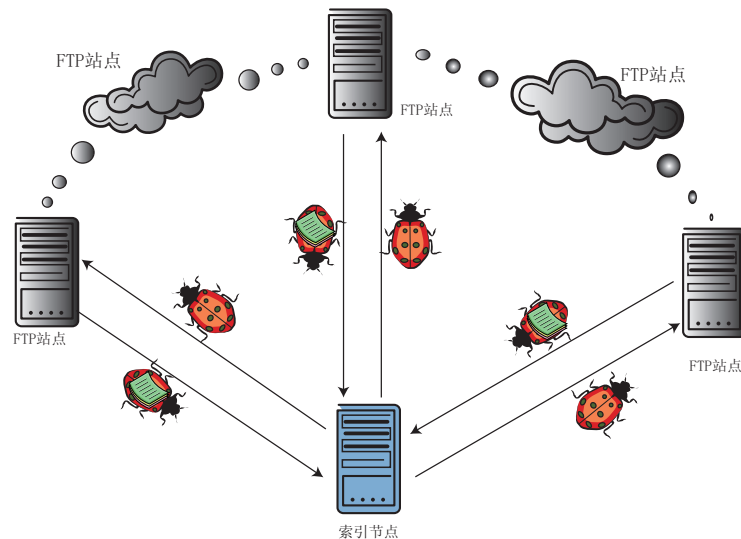
12

节点的地址。但配置文件的位置不一样，分别为：

RosaCrawler\src\resource\ProjectSetting.properties

和 RosaCrawler\src\resource\crawlnodes.xml

3. 运行 crawler.bat 执行爬虫程序。此时，程序会访问登记的 FTP 站点来获取 FTP 资源信息。



4. 爬行完成之后，需要运行 GuiHelper.bat。可以查看爬行站点的情况。也可以执行检索资源的操作。

Node 1 in 1 nodes of name:elegate

	Server	Port	Username	Password	Update	Total Fi...	Interval
1	172.16.166.169	21	wangqi...	dislab	2009-8-29	0	2
2	172.16.166.181	21	dh	dh	2009-8-29	136570	2

Node 1 in 1 nodes of name:elegate

Search: movie Go Current Page: 1 Total Pages: 1 Previous Next

Keyword: movie, hits number: 19, 30 results per page, time used: 93 milliseconds

Name	Site
movie	172.16.166.181
Movies	172.16.166.181
Movies	172.16.166.181
movie.ico	172.16.166.181
movie.png	172.16.166.181
movie.png	172.16.166.181
movie.png	172.16.166.181
movie.gif	172.16.166.181

5. 此时，用户输入 `http://127.0.0.1:8080/RosaWeb` 就可以检索登记的 FTP 站点的资源了。
6. 添加系统服务(可选)
运行 `wrapper` 目录下的 `InstallRosa_NT.bat` 可以添加为系统服务，这样系统重启的时候就不需要手动启动爬虫程序了。检查服务是否正常添加可以查看 `logs` 文件夹下面的日志。运行 `wrapper` 目录下的 `UninstallRosa_NT.bat` 可以删除系统服务。

7 针对开发者的系统部署

开发者可能会在 ROSA 源码的基础上添加功能或者进行改善，在此情形下，开发者就需要自己配置数据库和 Tomcat。另外，开发者需要使用 eclipse 或者更为强大的 MyEclipse 来进行开发。

7.1 导入工程

从源码包中导入三个工程：RosaWeb，RosaCrawler 和 RosaWebFTP。

7.2 数据库安装和配置

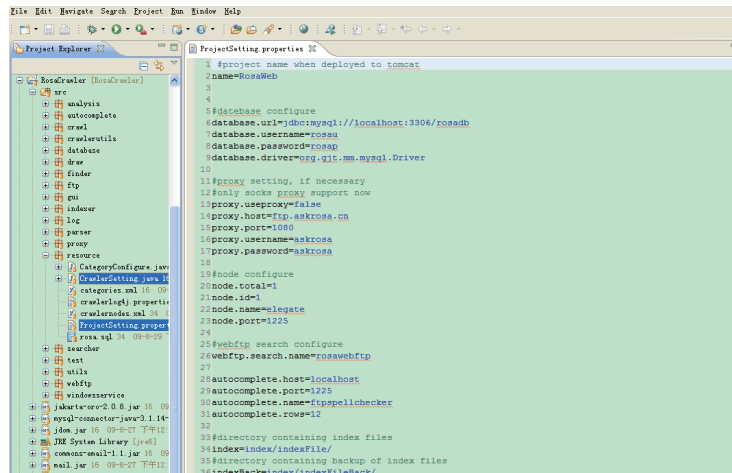
如6.1所示。

7.3 Tomcat 安装配置

如6.2所示。

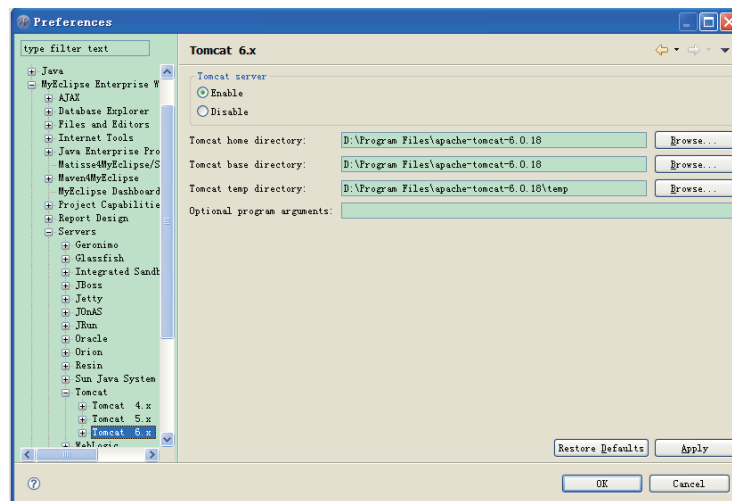
7.4 RosaCrawler 配置

下图是 RosaCrawler 中的配置文件的位置



7.5 ROSAWeb 配置

因为 RosaWeb 中引用 RosaCrawler 的配置，所以无需配置 IP 信息。但是需要配置 Tomcat 服务器的位置



8 一般用户检索

8.1 简单搜索

这个大家都不陌生，直接输入你想查的内容，一个或多个词，多个词之间用空格隔开。要注意的是这样输入的多个值在我们的查询中默认是与的关系，而非或的关系。

8.2 高级搜索

此处我们需要将搜索关键字划分成两个部分：文件名 + 限制

文件名：就是我们常说的关键词，比如你在 Google 上输入 apache 就可以得到很多跟这个词有关的网页，这部分原则上是必须的，当然你不输入这部分而只输入限制条件也是可以的，但我想你不愿意看到有上万条记录的结果。

限制：就是你对搜索结果的某种过滤，比如只在某些站点搜索，或限制搜索以某个用户名注册的站点，这部分不是必须的。

关于如何书写限制条件我们在下面会详细介绍。需要说明的是这两个部分的组合是任意的，并没有顺序性，你可以随意书写你的查询语句。看完下面的介绍我想你会相信搜索是一门艺术也是一种享受。

8.2.1 文件名或文件路径查询

用户可以根据路径或者文件名查询。路径就是 FTP 服务器上的完整路径，文件名就是最后一级的文件或文件夹名称。默认是根据文件名查询。举例来说：对于 /movie/ 蓝莓之夜 /1.rmvb。如果根据路径搜索，会有下面两条结果：

/movie/ 蓝莓之夜

/movie/ 蓝莓之夜 /1.rmvb

而根据文件名搜索只有一个结果：

/movie/ 蓝莓之夜

默认根据文件名查询，参看图??。



2: 文件名或路径名搜索

8.2.2 分类查询

☐ 视频
 ☐ 音频
 ☐ 压缩文件
 ☐ 图片
 ☐ 可执行文件
 ☐ 文本
 ☐ 文档
 ☐ 文件目录
 ☐ 源码
 ☐ 字幕或歌词

这个功能使用起来比较简单，只需要在页面上选择你需要的分类就可以了。多选表示或者关系，不选表示任意分类。比如你搜电影：老无所依，而发现结果有一堆的分卷压缩包，那么你选择文件夹会使得结果变得简洁很多。使用这个功能是比较方便的，不需要给出文件后缀，因为我们都已经帮你考虑了。分类是基于文件后缀，具体的分类依据参看表??。

1: 分类说明

分类	意义	相应的文件后缀
video	视频	rm; rmvb; mpg; mpeg; mov; mtv; dat; wmv; avi; 3gp; dmv; divx; asf; vob; flv; mkv; swf
subtitle	字幕或歌词	srt; idx; sub; smi; ssa; lrc
audio	音频	wav; mp3; ra; rma; wma; asf; mid; midi; rmi; ogg; mod; ape; aiff; au; voc; vox
document	文档	pdf; doc; docx; chm; ppt; pptx; xls; xlsx; rtf
text	文本	tex; txt; html; htm; xml; log; ini; properties; prop; url; css
program	源码	java; c; cc; c++; cpp; cxx; h; hh; hpp; hxx; lisp; perl; pas; sh; pyo; a; so; lo; la; sql; class; js; o
image	图片	gif; jpg; jpeg; jpe; ico; mng; pbm; pgm; psd; png; pnm; ps; ppm; tif; tiff; bmp; xpm; eps; dcm; dicom
compress	压缩文件	gz; bz2; tar; zip; rar; ar; ear; jar; war; iso
executable	可执行文件	bat; exe; msi; dll; cab; com; sys; deb
unknown	未知	
directory	文件目录	

8.2.3 搜索结果排序

通过点击页面上的排序方式链接可以分别根据相关度（默认排序方式），文件大小或日期进行排序。参看图??。



分类	名称 [全部 展开] [全部 合并]	站点 (校内 教育 公网)	降 升	降 升	操作
1	Friends	icpms.nju.edu.cn	4.00 KB	20090118	
2	Friends	cog.nju.edu.cn	4.00 KB	20050325	
3	Friends OST I - Friends	172.16.66.164	0.00 B	20080704	
4	Friends OST I - Friends	219.219.116.88	0.00 B	20080625	
5	Friends_Collection	172.16.66.164	0.00 B	20080704	
6	Friends_Music	172.16.66.164	0.00 B	20080704	
7	Friends_eBook	172.16.66.164	0.00 B	20080704	
8	Friends OST	219.219.116.88	0.00 B	20080625	
9	FRIENDS.WMF	margeo.nju.edu.cn	4.94 KB	20060119	
10	Friends OST II - Friends Again	172.16.66.164	0.00 B	20080704	
11	Friends OST II - Friends Again	219.219.116.88	0.00 B	20080625	
12	Friends.Cast.Say.Goodbye.Friends.on.Oprah.Show	172.16.66.164	0.00 B	20080704	
13	trouble is a friend.mp3	202.119.45.236	5.59 MB	20090729	
14	trouble is a friend.mp3	jeans236.njuftp.org	5.59 MB	20090729	
15	friend_died.wav	172.16.66.164	17.82 KB	20090414	
16	last_friend.rar	okok.njuftp.org	131.62 KB	20090209	
17	friend_died.wav	172.16.66.164	17.82 KB	20090101	
18	Friends Perfect Edition Stuff	172.16.66.164	0.00 B	20080704	
19	friends season1 music	172.16.66.164	0.00 B	20080704	
20	friends season2 music	172.16.66.164	0.00 B	20080704	
21	friends season3 music	172.16.66.164	0.00 B	20080704	

3: 搜索排序

8.2.4 搜索关键字过滤

有些关键字会导致 Lucene 分析器抛出异常，比如 “[先知].Knowing.2009.BDRip.X264-TLF” 这样的关键字由于 “[先知]” 的存在会导致程序抛出异常，这是因为 “[]” 这样的符号对在 Lucene 中表示范围搜索。一旦出现异常，搜索结果将为空，这是用户不愿看到的，所以我们队用户输入进行了过滤。比如 [] 以及 ~ 这样的 Lucene 关键字字符都会被过滤掉。但高级搜索页面中没有进行过滤，这里假定使用高级搜索页面的用户知道一些更高级的搜索方法。

8.2.5 后缀名查询

通过使用 关键字 +ext:后缀名 这样的查询条件可以针对特定的后缀名进行过滤。此处的后缀名全为小写。使用大写的后缀名将导致搜索结果为空。

8.2.6 逻辑与或非

这个其实也是比较简单的，就是说搜索的时候可以用 AND, OR, NOT，注意此处我使用的是大写形式，也就是说你在搜索里使用的时候也要是大写形式，小写形式不做处理，就当作一般的搜索词使用。ROSA 中默认的连接词是 AND。

8.2.7 单点搜索

如果你只想对某些站点的内容进行搜索的话，你可以使用这个功能，使用的方式也是非常简单的。比如 `friend server:172.16.65.79` 这样搜索语句就可以将对关键词 `friend` 的搜索限制在 `172.16.65.79` 这个站点内。如果你希望在多个指定站点内搜索，使用上面的逻辑组合就可以了，比如 `movie (server:172.16.65.79 OR server:172.16.65.105)` 就可以将搜索限制在站点 `172.16.65.79` 或 `172.16.65.105` 上。

8.2.8 通配符

这个大家可能不是很熟悉，但都应该听过。其实就是使用 `*`，`?` 这些具有特定意义的字符。`?` 代表一个或零个字符，`*` 代表 0 个或多个字符。我们对这些通配符提供了一定的支持，比如：`movie server:172.16.*.*` 这样的搜索请求我们是支持的，它讲搜索限制在以 `172.16` 为前缀的地址上。当然也可以在搜索关键词中加入通配符，比如 `fri?nd server:172.16.*.*`。这对于搜索在同一网段的资源十分有益，因为同一网段的资源往往下载速度可观。

特别提示：不建议搜索以 `*` 开头的关键词。虽然我们支持这种查询，但由于这种通配符生成的枚举项过多会导致搜索性能的下降。我们使用的是 1024 个枚举项的默认设置，如果你的搜索关键词导致枚举项的数目超过 1024 那么你的到的将是一个异常，当然我们没有将这个异常展示在你的面前而是返回一个空的结果集。

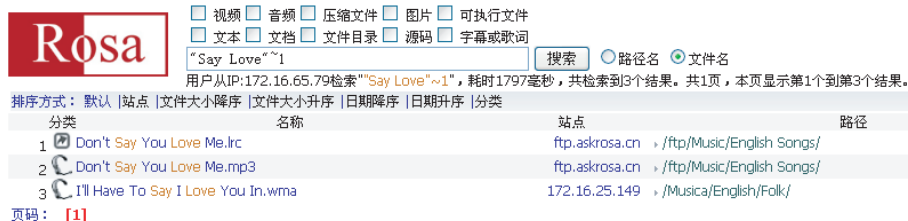
8.2.9 模糊查询

这个功能我必须讲，因为如果我不在此处告诉你的话，你可能不会发现我们的系统对这个功能有可以容忍的支持。之所以说可以容忍是因为不支持中文，但我想使用此功能的大多数情况应该是英文，所以模糊查询还是值得一提的。当你记不住某个单词的时候是不是就没办法查询了呢，你可能会考虑使用通配符，但你可能发现通配符要求你的输入字母有一定的顺序性，比如 `frien*` 和 `frine*` 的查询结果就很不相同，很可能后一个的查询结果为空(我测试了一下，结果确实为空)。而使用模糊查询，你的这个问题就很容易解决了，你可以尝试搜索 `frine~` 或者 `frined~` 这样的关键词，你会发现你真正想要的 `friends` 也在搜索结果中，并且是比较靠前的。由于关键字过滤，这个功能只能在高级搜索页面使用。

8.2.10 短语查询

这个功能其实大家也都在 Google 上使用过，比如你想查 `love me love my dog` 或 `to be or not to be` 这样的短语，直接输入可能结果不是很理想，但加上引号之后结果就比较符合你的需要了。这就是短语查询，在我们的系统中也提供了这样的支持，但这要建立在你对文件名或者文件路径比较熟悉的情形。比如：`"YOU ARE TEH LOVE.mp3"` 这样的搜索结果只有一条，而去掉括号后结果条

数为 1944 条。短语查询的另一个用法是匹配单词间隔。比如你输入”Say Love”，查询结果为空，但输入”Say Love”~1 输出结果如下：



这样的输入匹配了 Say 和 Love 的间隔小于一个词的结果。而不上~默认的是 0 个间隔，故匹配结果为空。所以在你不能很准确的把握某个短语时，使用这种加上间隔的查询会帮助你找到你想要的资源。同样由于关键字过滤，这个功能只能在高级搜索页面使用。

8.2.11 范围查询

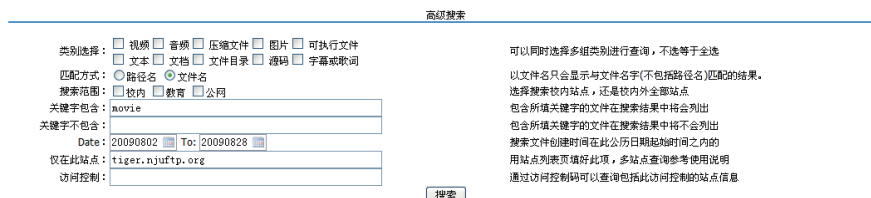
这个功能在我们的搜索中主要提供时间范围的查询，比如你想查询修改时间在 2008 年 3 月以来的有 movie 关键词的条目，可以这样搜索 movie date:[20080301 TO null]。某一端为 null 表示在这以端不加限制。同样由于关键字过滤，这个功能只能在高级搜索页面通过选择起始结束时间来使用。

8.2.12 访问控制

如果你有一个 FTP 站点，但你不将将这个站点注册上来然后能被所有人搜索，那么你可以在我们的注册页面上将访问控制设定为你想要的字串，那么在高级搜索页面中可以输入这个字串来搜索你的站点，而不输入这个字串的用户是搜索不到你的站点的。默认的字串是 anybody，所以如果你想让你的站点只在小范围共享的话请确保你的字串不要为 anybody。这个功能只能在高级搜索页面使用。

8.2.13 高级搜索

在高级搜索页面中，不对关键字进行非法字符过滤。此外用户可以组合多种搜索条件来减少结果范围，达到快速定位资源的目的。上面介绍的单点搜索，范围搜索和访问控制都要在这个页面才能使用。参看图??。



4: 高级搜索

8.3 ROSA 的索引域

8.3.1 索引域说明

2: ROSA 可搜索域	
搜索域名称	意义
id	FTP 服务器在数据库中的 ID
server	FTP 服务器域名
username	登陆用户名
password	登陆密码
location	站点位置：校内，教育网，公网
port	端口号
access	访问控制
path	文件路径域
parent	文件父目录
name	文件名称域
date	文件更新时间
size	文件大小
ext	文件后缀名
cat	文件分类
updatetime	索引更新时间

其中搜索域 access 只有使用高级搜索功能时才可使用。updatetime 域目前没有提供高级搜索接口。默认的是 10 天内得到更新的索引都可以被搜索到。此外端口号，密码对搜索的帮助不大，用户名域可以用来区分同一机器上 FTP 服务器上的不同用户的索引。比较常用的是 name 和 path 域。

8.3.2 索引域的使用

通过在搜索域后加冒号“:”然后加上条件就可以使用搜索域。比如：蓝莓之夜 cat:subtitle 可以用来检索蓝莓之夜的字幕文件。

9 用户高级功能

9.1 ROSA FTP

用户对搜索到的资源可以选择复制链接地址后使用常用的 FTP 下载软件下载，同时也可以选择使用 ROSA 提供的 ROSA FTP 多点下载工具下载。通过点击图??中标记的“启动 ROSA FTP”来启动下载工具，然后点击“操作”下方的箭头图标就可以添加下载项。图??给出了 ROSA FTP 下载过程的截图。



5: ROSA FTP 启动



6: ROSA FTP 下载

9.2 用户资源请求

在发现 FTP 检索中并没有出现所需的资源的时候，用户可以利用资源请求功能来向所有使用 ROSA 搜索引擎的用户请求资源。如果当有用户看到这个请求，并拥有其所请求的资源，用户可以上传到任何 ROSA 能搜索到的 FTP 站点。但 ROSA 爬虫程序搜索到用户请求的资源之后，会想用户发送邮件提醒资源已经可以获得。

ID	资源名	昵称	请求时间	截止时间	状态	Email	操作
1	cctv六国崛起系列	anonymous	2009-08-15	2009-09-14	未回复		删除
2	大团综	至胜先师	2009-06-24	2009-07-24	未回复		删除
3	一搜成名2	anonymous	2009-06-14	2009-07-14	未回复		删除
4	现在只想爱你	anonymous	2009-06-04	2009-07-04	未回复		删除
5	李连杰42部我有	anonymous	2009-05-09	2009-06-08	未回复		删除
6	都峰立前三季的2、3集	anonymous	2009-05-08	2009-06-07	未回复		删除

9.3 用户发布资源

上传到任何 ROSA 能搜索到的 FTP 站点。ROSA 在更新站点索引的时候，会得知请求的资源是否被上传。

9.4 用户发布留言

用于用户和用户之间，以及用户和开发者维护者之间的沟通。

ID	标题	作者	回复/点击	时间	IP	操作
1	关于FTP站点的默认密码	elegate	1/19	2009-06-04 17:41:20.0	172.16.65.78	修改
2	ftp	ldm	2/20	2009-06-03 16:55:52.0	172.16.77.21	修改
3	【资源发布】拉贝日记下载地址	anonymous	0/9	2009-05-31 20:26:32.0	172.16.166.181	修改












9.5 离线站点快照

1. 搜索结果所在位置的浏览

秒，共4个结果。第325375次检索 今天第869次 本周第4547次 本月第19775次 今年第266248次

<div><div>全部</div><div>合并</div></div>	站点 (校内 教育 公网)	降 升	降 升	操作
	ftp.askrosa.cn	14.74 KB	20070913	
	ftp.askrosa.cn	11.17 KB	20040919	查看文件所在目录
	ftp.askrosa.cn	386.98 MB	20080328	
	ftp.askrosa.cn	40.00 KB	20040224	

页码: [1] 共1页

站点快照					
<div> <div>  pub > LinuxSoft">www.thephly.nyu.edu.cn-anonymous > pub > LinuxSoft [复制链接地址] </div> <div> 点击查看 【非FTP预览】 启动Rosa FTP </div> </div>					
分类	名称	大小	时间	语言	操作
1	message	269.00 B	20070604		
2	welcome.png	324.00 B	20070610		
3	0_bcoming	4.00 KB	20090213		
4	1_LinuxCD-Linux	4.00 KB	20080417		
5	ArchLinux	4.00 KB	20080417		
6	Debian	4.00 KB	20080417		
7	FreeBSD	4.00 KB	20080417		
8	LFS	4.00 KB	20080417		
9	Magix.Linux	4.00 KB	20080417		
10	Mandrake	4.00 KB	20080417		
11	Mini.Linux	4.00 KB	20080417		
12	Minix	4.00 KB	20080417		
13	NetBSD5.3.3	4.00 KB	20080417		
14	OpenBSD	4.00 KB	20080417		
15	PCBSD1.4	4.00 KB	20080417		
16	PClinuxOS	4.00 KB	20080417		
17	RAYS	4.00 KB	20080417		
18	Redhat	4.00 KB	20081127		
19	Redflag 6.0	4.00 KB	20080417		
20	Slackware	4.00 KB	20080417		
21	Solars	4.00 KB	20080417		
22	SuSE	4.00 KB	20080417		
23	Sun.Java.Desktop.System.2-GO	4.00 KB	20080417		
24	Trustix	4.00 KB	20080417		
25	Turbolinux10	4.00 KB	20080417		
26	Ubuntu	4.00 KB	20080417		
27	XenEnterprise	4.00 KB	20080417		
28	Yogwin	4.00 KB	20080417		
29	gentoo	4.00 KB	20080417		
30	puppy	4.00 KB	20080522		

2.针对 FTP 站点的浏览

ID	地址	用户名	总文件数	视频	音频	文档	程序	HOT	更新时间	周期
1	tiger.njuft.org	anonymous	3724	1515	1	2	0	6202	2009-08-24	1
2	rock.nju.edu.cn	anonymous	31973	186	24725	32	4	3555	2009-08-26	5
3	cog.nju.edu.cn	anonymous	24138	2174	494	5591	237	2481	2009-08-25	5
2	202.119.51.71	anonymous	0	0	0	0	0	2427	2009-08-28	1
5	theophy.nju.edu.cn	anonymous	24190	159	85	206	2198	2376	2009-08-28	1
6	cs.nju.edu.cn	download	42783	1938	1840	2470	2680	1603	2009-08-28	1

9.6 搜索建议

类似于 Google suggestion 的功能，根据用户的输入给出关键字建议

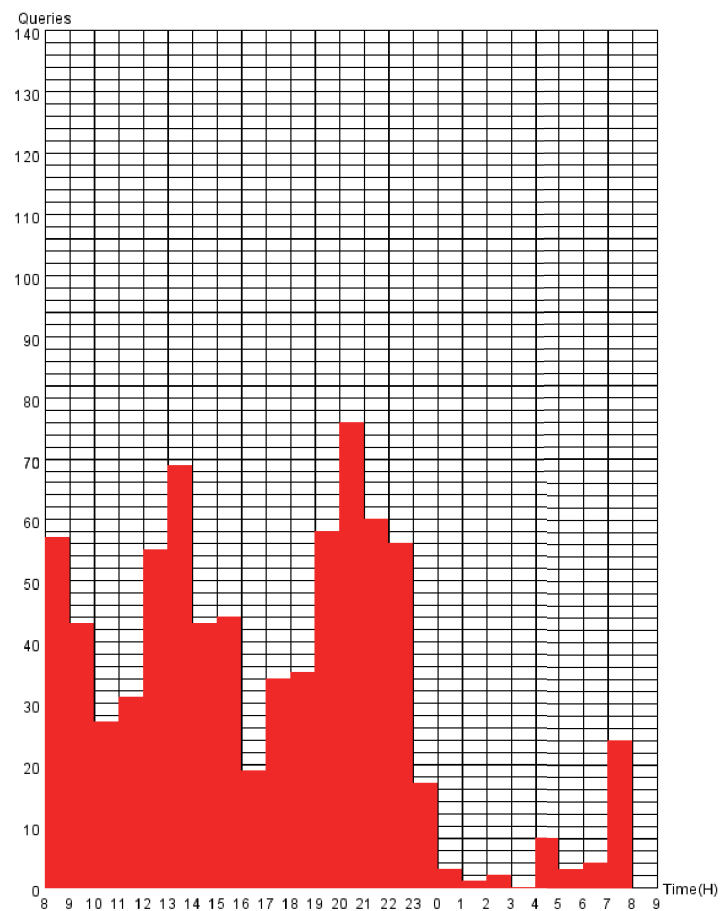
9.7 搜索统计

通过记录用户搜索历史和点击历史，我们在首页和单独的系统负载页面给出了相关的统计信息。一方面可以让用户了解群体的搜索喜好，另一方面也可以让管理员了解系统的负载情况。图??中的资源名称后的数字表示此资源的一周搜索次

数；站点名称后的次数表示此站点的历史用户点击数。图??给出了系统 24 小时的负载直方图，横坐标表示 24 时制的时间，纵坐标表示系统每小时的总搜索次数。

热门搜索：NOT! 先知(33) 电影(33) 纪晓岚(21) mov(16) offic(15) 夜店(15) anonymous(15) 变形金刚(15) window(12) friend(9)
热门站点：NOT! tiger.njuftp.org(6205) rock.njuftp.org(3584) cog.nju.edu.cn(2482) 202.119.51.71(2427) thephy.nju.edu.cn(2377) cs.nju.edu.cn(1609)

7: 热门资源和站点统计



8: 24 小时系统负载统计

9.8 热门站点排名

用户可以根据不同属性排名站点，然后选择自己喜欢热门站点来离线浏览或下载。

站点列表

1.可导入FlashFxp的XML格式下载 2. FlashFxp-3.4.0下载 3.站点:(校园网: 247 教育网: 109 公共网: 5) 站点总数:361 文件总数:659147

FTP地址:		联系方式:		站点位置:		全部		更新时间从:		到:	
ID	地址	用户名	总文件数	视频	音频	文档	程序	HOT	更新时间	周期	联系方式
1	tiger.njuftp.org	anonymous	3724	1515	1	2	0	6202	2009-08-24	1	phtiger@bbs.nju.edu.cn
2	rock.njuftp.org	anonymous	31973	186	24725	32	4	3556	2009-08-26	5	QQ:6868922
3	cog.nju.edu.cn	down	24138	2174	494	5591	237	2481	2009-08-25	5	
4	202.119.51.71	anonymous	0	0	0	0	0	2427	2009-08-28	1	
5	thephy.nju.edu.cn	anonymous	24190	159	85	206	2198	2377	2009-08-28	1	
6	cs.nju.edu.cn	download	42783	1938	1840	2470	2680	1603	2009-08-28	1	
7	media.nju.edu.cn	anonymous	45925	2974	19562	276	113	1358	2009-08-24	5	
8	sofree.njuftp.org	anonymous	1210	341	15	3	0	953	2009-08-26	1	EL8LatSPQ@bbs.nju.edu.cn
9	172.16.65.43	anonymous	-1	5545	2285	5	3	863	2009-07-18	5	
10	salon.nju.edu.cn	anonymous	-1	1367	1673	2568	456	728	2009-06-26	5	
11	ftp.askasia.cn	admin	166293	854	3326	2851	1698	556	2009-08-28	2	
12	172.16.71.139	anonymous	-1	120	1045	289	895	433	2009-07-14	5	
13	202.116.160.71	scau	0	0	0	0	0	392	2009-08-18	5	华南农业大学
14	166.111.72.5	outofthu	-1	573	676	6931	2081	388	2009-05-07	5	清华大学自动化系(中央主

10 FTP 站点管理

10.1 登记 FTP 站点

增加站点

基本选项：

地址*：

用户名：

密码：

验证码*：

验证码确认*：

FTP域名或IP地址

FTP用户名，默认anonymous，如果允许匿名访问，请不要改动

FTP密码，默认anonymous，匿名访问无需密码

为保证只有注册的人具有修改FTP信息的权限，用验证码来确认

验证码确认

[更多高 级选项:]

管理员：

联系方式：

端口：

更新周期：

限速(KB)：

用户数：

描述：

编码：

站点位置：

访问控制：

FTP管理员，默认为UNKNOWN,注册者更具自己的情况命名

管理员联系方式，邮箱、百合账号或其他

FTP端口号，默认21

索引程序对此站点进行索引的周期，默认为1天

FTP速度限制，默认0，表示不限速

FTP用户数限制，默认0，表示不限用户数

FTP描述信息，FTP内容和作用等

FTP传输编码，默认GBK，可选的有UTF-8和GB2312

FTP站点的位置，根据您的填写，我们采用不同的访问方式

默认anybody，表示FTP可以被任何人检索到，其他字符串用户在检索时需提供此字符串

提交

注意：为了防止错误提交，提交的时候，必须保证FTP站点能够登陆进入，否则不会收录！如果登记不上，可联系我们帮您登记。

10.2 更改站点信息

使用说明 更新历史 访问统计 Rosa FTP 资料下载 加入收藏

站点列表

3. 站点: (校园网: 247 教育网: 109 公共网: 5) 站点总数: 361 文件总数: 659147

站点位置: 全部 更新日期从 到 搜索站点

音频	文档	程序	HOT	更新时间	周期	联系方式	位置	操作
15	1	2	0	6203 2009-08-24	1	phtiger@bbs.nju.edu.cn	校园网	修改 复制
86	24725	32	4	3558 2009-08-26	5	QQ:6868922	校园网	修改 复制
74	494	5591	237	2481 2009-08-25	5		校园网	修改 复制
0	0	0	0	2427 2009-08-28	1		校园网	修改 复制
50	85	206	2108	2277 2009-08-28	1		校园网	修改 复制

10.3 自定义共享圈

FTP 站点管理员为自己的站点设定一个访问控制码，只有拥有这个访问控制码的用户才能检索到此站点的资源，用于用户之间私有资源的共享。

FTP传输编码，默认GBK，可选的有UTF-8和GB2312

FTP站点的位置，根据您的填写，我们采用不同的访问方式

默认anybody，表示FTP可以被任何人检索到，其他字符串用户在检索时需提供此字符串

提交

注意：为了防止错误提交，提交的时候，必须保证FTP站点能够登陆进入，否则不会收录！如果登记不上，可联系我们帮您登记。

11 ROSA FTP 下载

11.1 安装 ROSA FTP

11.1.1 安装 JRE

因为 ROSA FTP 是 Java 程序，所以需要安装 JRE 或 JDK 才可以使用。用户可到 Java 官方网站 <http://java.sun.com> 下载 JRE1.6 或者 JDK1.6。

11.1.2 启动 ROSA FTP

点击 Web 页面的“启动 ROSA FTP”。如果机器中已经安装 ROSA FTP，已经按照 ROSA FTP 的用户会直接启动 ROSA FTP，没有按照的用户会先下载，再安装。

下载之后，启动 ROSA FTP 的时候需要验证，点击确认即可：



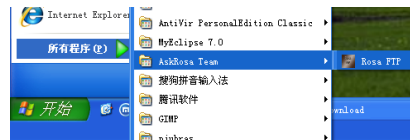
在用户需要利用 ROSA FTP 下载所检索到的资源的时候。点击下载图标进行下载。ROSA FTP 会弹出选择保存地址的对话框，选择保存地址后即可下载。

112毫秒，共4个结果。第325273次检索 今天第767次 本周第4445次 本月第19673次 今年第266146次					
[开] [全部] [合并]		站点 (校内 教育 公网)	降 升	降 升	操作
		ftp.askrosa.cn	14.74 KB	20070913	用RosaFTP下载
		ftp.askrosa.cn	11.17 KB	20040919	
		ftp.askrosa.cn	386.98 MB	20080328	
		ftp.askrosa.cn	40.00 KB	20040224	

11.1.3 ROSA FTP 的安装位置

ROSA FTP 的快捷方式

桌面快捷方式和开始菜单快捷方式如下图所示，ROSA FTP 保存在 Java 缓存中

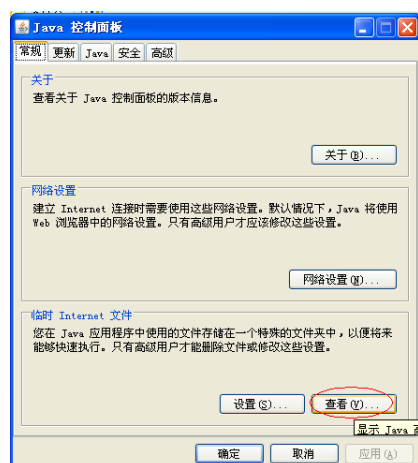


11.2 ROSA FTP 的卸载

在 Java 控制面板--添加删除程序中删除

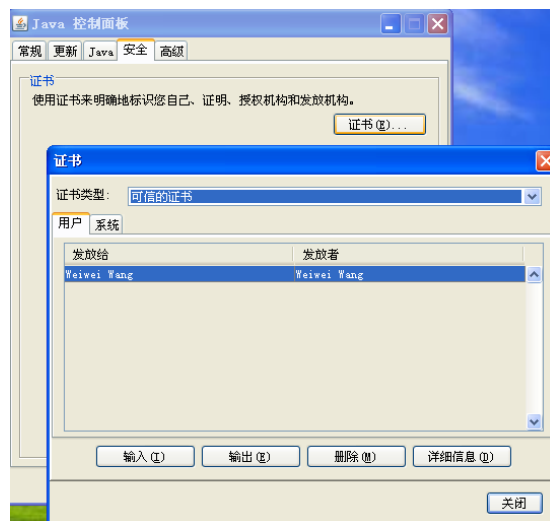


清除 ROSA FTP 临时文件





去除信任签名



如果您觉得没有必要使用 JRE 也可以删除 JRE



12 意见反馈

如对 ROSA 系统有任何意见或建议请登录 <http://nju.askrosa.cn> 留言或发邮件到 askrosateam@gmail.com。