|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **北京邮电大学软件学院** | **文档编号** | TDIPS-6 | **版本** | 1.0 | **密级** | 商密A |
| **项目名称** | **文本数据智能处理系统** | | | | |
| **项目来源** |  | | | | |

**支持AJAX的定址网络爬虫模块**

**源代码及清单**

(内部资料请勿外传)

|  |  |  |  |
| --- | --- | --- | --- |
| **编写：** | 刘凡凡 | **日期：** | 2011-4-25 |
| **检查：** | 刘凡凡 | **日期：** | 2011-4-25 |
| **审核：** | 吴国仕 | **日期：** | 2011-4-25 |
| **批准：** |  | **日期：** |  |

**北京邮电大学软件学院**

**版权所有不得复制**

**文档变更记录**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **序号** | **变更（+/-）说明** | **作者** | **版本号** | **日期** | **批准** |
| 1 | 初次编写 | 刘凡凡 | 1.0 | 2011-4-25 |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

目录

[1. 目录结构 4](#_Toc342663456)

[2. 文件描述 5](#_Toc342663457)

[2.1. com.liu.core包 5](#_Toc342663458)

[2.2. com.liu.ui包 5](#_Toc342663459)

[2.3. com.mywie.utils包 6](#_Toc342663460)

[2.4. 其他文件 6](#_Toc342663461)

[3. 代码清单 6](#_Toc342663462)

# 目录结构

-- src

--com.liu.core

-- abstractTimerTask.java

-- AutoSaver.java

-- BrowserApplication.java

-- BrowserMainWindow.java

-- httpGet.java

-- JsGenerator.java

-- MyCookieJar.java

-- reloadTimerTask.java

-- com.liu.ui

-- ClearButton.java

-- ConfigPage.java

-- CookieModel.java

-- CookiesDialog.java

-- Crawler.java

-- CrawlerImpl.java

-- CrawlPage.java

-- EditTableView.java

-- ExLineEdit.java

-- JsCreatePage.java

-- MDialog.java

-- MyformPage.java

-- MyWebView.java

-- SearchButton.java

-- SearchLineEdit.java

-- com.mywie.utils

-- FileHelp.java

# 文件描述

## com.liu.core包

本包之下的代码，为软件系统中实现核心功能的代码。

**包内文件描述：**

|  |  |
| --- | --- |
| 类名 | 描述 |
| abstractTimerTask.java | 在执行批量抓取的过程中，对于一次加载停滞或长时间加载不玩的情况，激发上级程序执行重新加载或直接加载下一个URL链接。 |
| AutoSaver.java | cookie保存相关设置。 |
| BrowserApplication.java | 管理浏览器配置文件信息的保存地址。 |
| BrowserMainWindow.java | 批量抓取的浏览器，能加载Javascript脚本执行翻页爬取，并保存新加载的内容。 |
| httpGet.java | 链接抓取的核心代码，实现了一些列方法抓取一定数量、一定格式的URL链接。 |
| JsGenerator.java | 生成脚本的核心代码，和用户进行交互，根据网页的具体情况，生成可以执行翻页爬取网页内容的Javascript脚本。 |
| MyCookieJar.java | 管理cookie的核心代码。 |
| reloadTimerTask.java | 具体管理执行abstractTimerTask.java的代码。 |

## com.liu.ui包

本包之下的代码，为软件系统中实现界面内容的代码。

**包内文件描述：**

|  |  |
| --- | --- |
| 类名 | 描述 |
| ClearButton.java | Cookie管理窗口的搜索栏中的清除按钮。 |
| ConfigPage.java | 主界面当中的配置信息界面。 |
| CookieModel.java | Cookie管理窗口中Cookie信息显示框。 |
| CookiesDialog.java | Cookie管理窗口界面。 |
| Crawler.java | 主界面。 |
| CrawlerImpl.java | 主界面配置信息。 |
| CrawlPage.java | 主界面当中的批量抓取界面。 |
| EditTableView.java | Cookie管理窗口中Cookie信息显示框的具体实现。 |
| ExLineEdit.java | Cookie管理窗口中搜索栏的核心。 |
| JsCreatePage.java | 主界面当中的脚本生成界面。 |
| MDialog.java | 链接抓取界面当中的各类串添加设置的显示框。 |
| MyformPage.java | 主界面当中的链接抓取界面。 |
| MyWebView.java | 脚本生成界面当中的浏览器。 |
| SearchButton.java | Cookie管理窗口中的搜索栏自动按钮。 |
| SearchLineEdit.java | Cookie管理窗口中的搜索栏。 |

## com.mywie.utils包

本包之下的代码，为软件系统中实现辅助操作的代码。

**包内文件描述：**

|  |  |
| --- | --- |
| 类名 | 描述 |
| FileHelp.java | 用以读取文件的类。 |

## 其他文件

|  |  |
| --- | --- |
| 文件名 | 描述 |
| log4j-1.2.16.jar |  |
| qtjambi-4.7.1.jar |  |
| qtjambi-designer-4.7.1.jar |  |
| qtjambi-win32-msvc2008-4.7.1.jar |  |
| yuicompressor-2.4.7.jar |  |
| bbs-mop.js | 猫扑贴贴下的帖子的翻页脚本。 |
| bbs-tianya.js | 天涯论坛下的帖子的翻页脚本。 |
| blog-163.js | 网易博客下的博文的翻页脚本。 |
| blog-sina.js | 新浪博客下的博文的翻页脚本。 |
| blog-sohu.js | 搜狐博客下的博文的翻页脚本。 |
| blog-xinhua.js | 新华博客下的博文的翻页脚本。 |
| blog-xinhua-preprocessing.js | 新华博客下的博文的翻页脚本。 |
| bufanye.js | 不需翻页的网页通用脚本。 |
| mblog-163.js | 网易微博下的主页的翻页脚本。 |
| mblog-sina.js | 新浪微博下的主页的翻页脚本。 |
| mblog-sohu.js | 搜狐微博下的主页的翻页脚本。 |
| shopping-360buy.js | 京东商城下的商品的翻页脚本。 |
| shopping-taobao.js | 淘宝网下的商品的翻页脚本。 |
| shopping-tmall.js | 天猫商城下的商品的翻页脚本。 |

# 代码清单