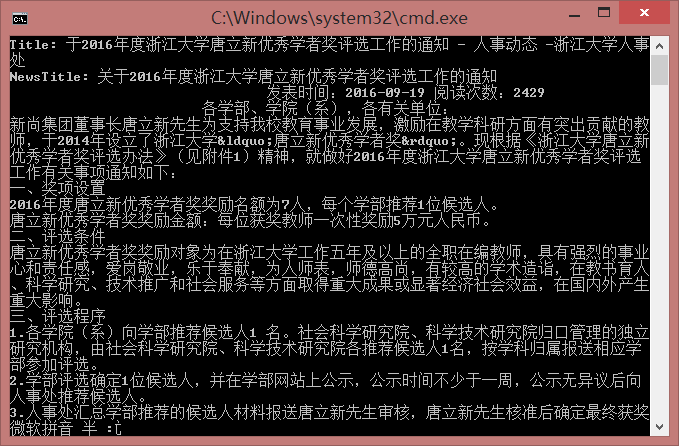
第二周作业 计算机学院网人事处网络公告节选程序

周晨恺 3140102277

程序说明：

能够从 <http://cspo.zju.edu.cn/redir.php?catalog_id=255> 中各个新闻链接中对应的新闻进行节选，删除无关部分信息内容，并将主要内容进行相关处理之后以 .txt 的格式存储在源文件根目录下。

程序运行实例：

程序使用说明：

将压缩包解压后的根目录下，修改 GetHtml.java 源代码中第 59 行的网址，就可以实现对不同通知的解析。

现在暂时只支持对目录 <http://cspo.zju.edu.cn/redir.php?catalog_id=255> 中对应的新闻进行解析和下载。

程序实现说明：

整个程序总共实现了两个类，分别是 GetHtml 类和 ParseHtml 类。其中 GetHtml 类负责对指定的 Html 地址进行链接并且将其对应的 Html 文本下载下来，保存在 String 类的变量中。 ParseHtml 类只要负责对得到的 Html 文本进行解析，并从中提取出我们所需要的内容。其中 Parse 过程通过单向的状态机实现，采用 正则表达式来去掉 html 标签。而整体的 Parse 过程则是通过直接阅读网站的源代码，通过对我们所需要的内容的标签中的类进行标定，来筛选我们所需要的内容。