# Report

**学号：1152740**

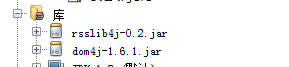
**姓名：杨磊**

**邮箱：272283475@qq.com**

## 项目概述

**本项目是基于RSS新闻源的新闻集成系统，我通过新浪RSS，人民网RSS，网易网RSS，搜狐网RSS四个新闻源来得到新闻信息，通过RSSlib解析这些数据，然后通过DOM写入XML，最后通过DOM来解析新生成的XML并且显示在网页上。此外还写了XML SCHEMA来验证生成的XML表**

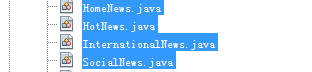
## 主要用到的外部类



Rsslib4j是用来解析RSS新闻源的

Dom4j是用来**创建**和**解析**XML的

## 解析RSS并写入XML（RSSLIB4J, DOM4J）



这四个类主要用来解析RSS信息源，这里针对HomeNews来进行解释

C:\Users\Andy\Desktop\a6.png

Document是用来写入XML的地址栏



String remoteRSS\_People="http://www.people.com.cn/rss/politics.xml";

String remoteRSS\_Sina="http://rss.sina.com.cn/news/china/focus15.xml";

String remoteRSS\_NetEase="http://news.163.com/special/00011K6L/rss\_gn.xml";

String remoteRSS\_Sohu="http://rss.news.sohu.com/rss/guonei.xml";

**设置RSS新闻源**

String Title\_People = item.getTitle();

//(4)摘要的链接

String Link\_People = item.getLink();

//(5)摘要的发布日期

String Pub\_People = item.getPubDate();

**提取新闻标题、新闻链接、新闻发布时间的RSS内容，提取其对应object**

Element publisher\_title=this.document.createElement("publisher\_title");

publisher\_title.appendChild(this.document.createTextNode("人民网国内新闻"));

Element title\_people = this.document.createElement("news\_title");

title\_people.appendChild(this.document.createTextNode(Title\_People));

Element link\_people = this.document.createElement("news\_link");

link\_people.appendChild(this.document.createTextNode(Link\_People));

Element pub\_people = this.document.createElement("publish\_time");

**创建XML文件中的节点**

elements.appendChild(publisher\_title);

elements.appendChild(title\_people);

elements.appendChild(link\_people);

elements.appendChild(pub\_people);

**将刚刚提取的RSS object存入节点**

**循环存入**

CreateXml会同时完成解析RSS和写入两个步骤，但这里的写入只是在内存中，

**函数CreateXML会执行存入数据的步骤**



## 解析并读取XML(DOM4J)



for (Iterator it = nodes.iterator(); it.hasNext();)这是对于news的循环读取

在JSP中通过SAXReader来读取目标XML，然后显示在网页上

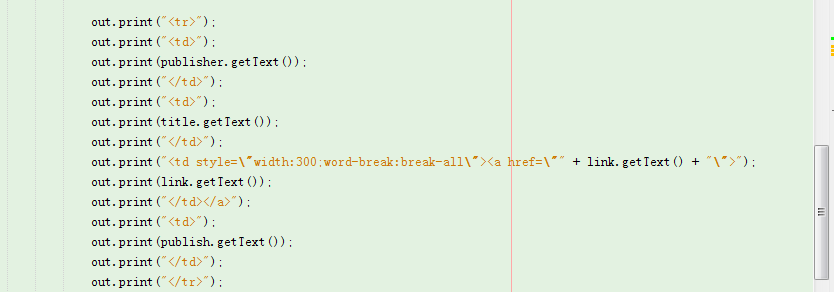
Element publisher = elm1.element("publisher\_title");

Element title = elm1.element("news\_title");

Element link = elm1.element("news\_link");

Element publish = elm1.element("publish\_time");

**读取元素的核心代码**



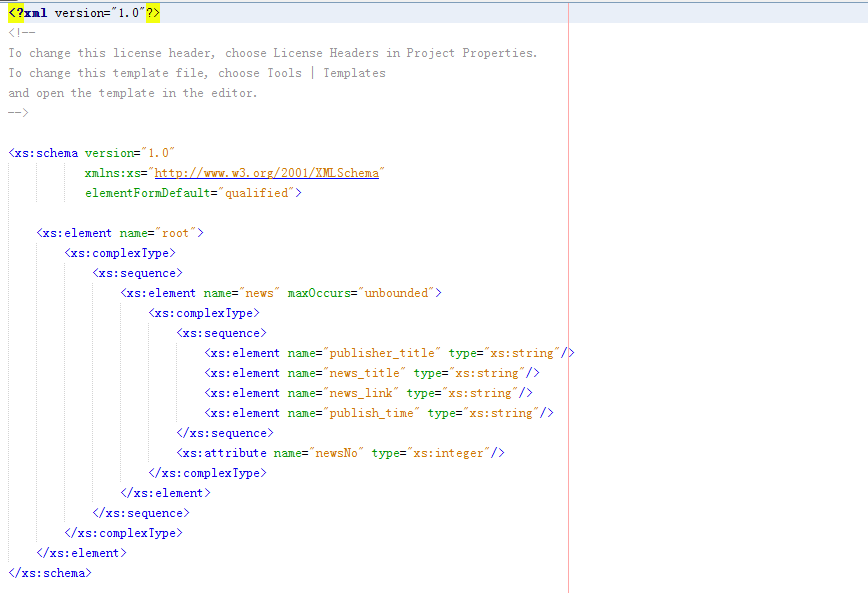
**然后循环生成表格，这样的表达形式更加清晰。**

## 搜索功能

C:\Users\Andy\Desktop\search.png

搜索功能是通过string类中自带的contains属性来完成的，通过判断新闻标题中是否含有对应的search\_key,可以完成关键字搜索

## Confirmation



通过XML Schema来对XML文件进行验证，上图为XML Schema代码，其结构分为三层，第一层为Root,第二层为news，第三层为news节点下的各种元素。

通过Dom4j自带的Validator方法来对其进行验证



## 心得体会

虽然这次只是做一个简单的RSS新闻聚合应用，但是在实际开发过程中却有相当多的问题和困难需要解决。看似简单的XML创建和读取，节点的插入往往要花费很多时间来进行调试。UTF-8与GBK的编码的转译、Web页上存储路径的选取，这些小问题却在开发过程中成了不小的难题。

整个项目简单说来就是从RSS新闻源获取信息，然后写入到本地XML，再从本地XML读取到网页这三个主要步骤。但其实现却一点都不简单。从清明放假开始我就每天花3-4小时进行编码，小问题层出不穷，加上之前对于XML只有一些存于理论的基本概念，实际开发可谓一波三折幻灭感丛生。所幸最后赶在deadline之前完成了整个项目。通过这次作业无疑加深了我对XML的理解，其中包括其创建和读取，节点的存入和读出，编码的转译，也学会了怎样利用XML Schema来验证XML文件，虽然编写时困难重重，但如今看来一切都是值得的。