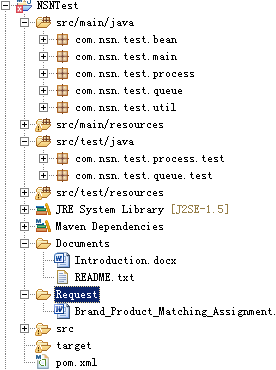
由于时间关系也为了表达的更清楚，下面将用中文描述。

主体项目结构如下



工程主要的信息来源都是以配置文件的方式产生，主要逻辑如下：

1. 工程将产生线程去处理xml文件数据，采用SAX的方式读取文件，通过xml文件里面的数据标题，用反射的方式封装好数据对象，并向将起分类放进缓存队列。
2. 缓存队列现在主要分了两个，一个为bran为空的数据队列，对于这个队列，采用的是生产者-消费者模式，开启一个线程去专门处理这个队列并输出到结果文件。另外一个队列是存放所有拥有brand的数据队列，这个需要等待全部文件处理完后才能将数据分类输出达到目标文件。
3. 在等处理XML文件的线程完成之后，开启两个线程分别去输出只有brand的数据的文件和输出所有brand以及product的文件。
4. 对于模糊匹配部分由于逻辑问题以及时间问题，未能完成。

在设想中能通过title的内容匹配出最合适的brand。但是在实际的数据中发现并不可行，如下面三条数据。

<ProductDetail brand="迪士尼" category="首页|电脑整机首页|电脑整机|笔记本配件" id="113430" model="DNC090463" price="129.0" productID="177674" siteName="360buy" title="Disney迪士尼快乐公主系列14寸电脑包-休闲包-DNC090463-粉" url="http://www.360buy.com/product/177674.html"/>

<ProductDetail brand="爱义" category="首页|家用电器首页|汽车用品|其它车载用品" id="115305" model="" price="78.0" productID="148316" siteName="360buy" title="迪士尼（Disney）“米奇\_头运”高级车用香薰（毒药）（金黄）" url="http://www.360buy.com/product/148316.html"/>

<ProductDetail brand="" category="首页|数码产品|数码附件|相机/摄像机包" id="111490" model="DSM080939 红" price="30.3" productID="96-c00-134" siteName="newegg" title="Disney 迪斯尼 数码相机包 DSM080939 红色" url="http://www.newegg.com.cn/Product/96-c00-134.htm"/>

在实际的title中，这样的数据将不能匹配

1. 还有中英文匹配问题，由于水平有限，未能给出好的解决方案





这个设想是将所有中英文的brand进行合并，并吸纳没有brand信息数据块的数据，对其进行brand分类