**介入程序需求：**

第一步：

1：原始的WORD, 按照原始WORD自动分页的情况来加<P#>（#代表页号）   
 相当于第一页的所有段落都加<p1>

第二页的所有段落都加<p2> 依次类推。

2：如果原始WORD有分页符PageBreak的情况，以分页符为标准来加<p#>

3：WORD版本里边有固定的图像，抓取图像保存成.JPG格式，对应的位置加标记<img ###-##.jpg> 第一个###代表页号，第二##代表当页的第几个图片，

例如 第10页的第一张图片 <img 010-01.jpg>

原始==：原始--1.doc

通完==：通完--1.txt 我们自己检查

第二步：

目录.xml：把目录表的ID点加好如下所示：



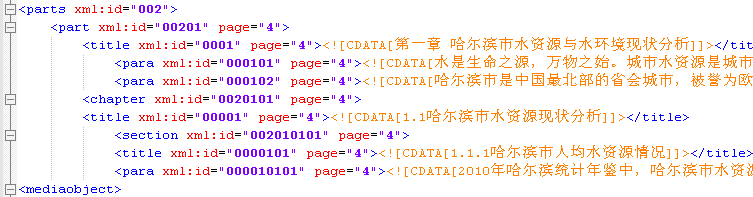
1: 根据目录表做的ID点和检查完的文件（通完-1.TXT）模糊匹配上对应的章节点的ID号，匹配不上的留标记，

2：我们自己检查：标记没有匹配的，手动加好，把摘要，参考文献放到对应的标记里

第三步：

1：按整篇的parts，把底下章节全篇再匹配

2：所有<para>的找最近一级别，按这级把所有的段，排序



成品样例：最终成品--1.xml