Poi手册

获取关联id

XWPFDocument xwpfDocument = new XWPFDocument(inputStream);

String 关联ID =xwpfDocument.getRelationId(xwpfDocument.getRelations().get(0))

根据关联ID获取图片,XWPFPictureData

xwpfDocument.getPictureDataByID(关联ID);

POI图片

XWPFPicture

XWPFPictureData

使用org.apache.xmlbeans 对docx xml 进行索引

xmlCursor.getAllNamespaces(); //获取当前的所有命名空间，

xmlCursor.insertNamespace(frefix, Namespaces); //插入一个命名空间

InputStream inputStream = new FileInputStream(new File("E:\\dms\\document\\XPTN-9000-85-4XG8GC16GP-HR-V(1).docx"));  
XWPFDocument xwpfDocument = new XWPFDocument(inputStream);  
CTDocument1 document = xwpfDocument.getDocument();  
XmlCursor xmlCursor = document.newCursor();  
xmlCursor.xmlText();

//查询语句,注意的是XMLBean对命名空间的要求。所有XPath和XQuery请求都需要给出相应的命名空间  
String xx = "declare namespace w='http://schemas.openxmlformats.org/wordprocessingml/2006/main'; declare namespace w14='http://schemas.microsoft.com/office/word/2010/wordml'; " +  
 "./w:body";  
 //"w:body/w:p/w:r/";  
xmlCursor.selectPath(xx);  
xmlCursor.toNextSelection();  
XmlObject object = xmlCursor.getObject();  
  
System.*out*.println(object.getClass());

注意：

使用标准的XPath需要添加，saxon，saxon9-dom，依赖

<dependency>  
 <groupId>net.sf.saxon</groupId>  
 <artifactId>saxon9-dom</artifactId>  
 <version>9.1.0.8</version>  
</dependency>

######################################################

<dependency>  
 <groupId>net.sourceforge.saxon</groupId>  
 <artifactId>saxon</artifactId>  
 <version>9.1.0.8</version>  
</dependency>