Word 文档信息获取方式

在Open Xml SDK中，WordprocessingDocument类代表了Word 文档中的一个包。要处理一份Word文档，首先需要通过文档创建一个WordprocessingDocument类的实例，然后操作这个实例从而实现对Word文档的操作。获取该实例之后，我们就能够访问包含了文本的主文档部分。每一个打开的XML包都包含若干个部分，一个WordprocessingDocument至少包含了一个主文档部分，该部分作为一个容器用来容纳Word文档中的主要文本内容，该WordprocessingDocument所代表的包也可以包含其他的一些部分。在一个Word文档中，主文档部分中的文本内容在XML包中使用WordprocessingML来标记。

要从一份Word文档创建WordprocessingDocument类的实例，我们可以调用Open（Stream，Boolean）方法，Open XML SDK中提供了好几个Open方法，我们采用具有2个参数的Open方法来获取实例，第一个参数代表一个Stream，该Stream关联了我们想要打开的Word文档，第二个参数可以取值为true或者false，代表我们打开的Stream是否是可编辑的。在我们的应用当中，由于对Word文档都是只读的，所以该参数均取值为false，表示不能修改打开的Word文档。创建WordprocessingDocument类实例的示例代码如下：

WordprocessingDocument document = WordprocessingDocument.Open(stream, false);

WordProcessingM文档的结构：

一份WordProcessingML文档的基本结构由document和body元素组成，之后再跟上一个或几个块级元素，例如p元素，代表一个段落。一个段落包含一个或几个r元素。r元素代表run，它代表一段具有共同属性的文本区域。一个run中包含一个或几个t元素，t元素包含一段文本。例如，对于只包含文本“测试文本”的Word文档来说，它的WordprocessingML 标记如下所示：

<w:document xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main">

<w:body>

<w:p>

<w:r>

<w:t>测试文本</w:t>

</w:r>

</w:p>

</w:body>

</w:document>

WordprocessingML中的各个元素在Open XML SDK里面都有相应的类与之对应，这些类都在DocumentFormat.OpenXml.Wordprocessing命名空间当中，下表列出了与document,body,p,r和t元素相对应的类：

|  |  |  |
| --- | --- | --- |
| WordprocessingML 元素 | Open XML SDK类 | 描述 |
| document | Document | 主文档部分的根元素 |
| body | Body | 段落、表格等块级元素的容器 |
| p | Paragraph | 段落 |
| r | Run | 一段具有共同属性的文本区域 |
| t | Text | 一段文本 |

要访问主文档部分的body，我们可以把存在的文档体赋值给Body引用，像如下代码所示：

WordprocessingDocument document = WordprocessingDocument.Open(stream, false);

Body body = document.MainDocumentPart.Document.Body;

操作段落

WordprocessingML中的段落

WordprocessingML中最基本的块级内容单元就是段落，段落用<p>元素进行存储。一个段落定义了一个由新行开始的独立内容区域。段落可以包含三部分信息：可选的段落属性、内联内容（比如run）,以及一些可选的校正ID，用来比较2篇文档的内容。

一个段落的属性是通过<pPr>元素指定的，段落属性包含对齐方式、边框、缩进、行距、底纹、文本方向等。下表列出了Open XML SDK中与段落操作相关的常用类与WordprocessingML元素的对应关系：

|  |  |
| --- | --- |
| WordprocessingML元素 | Open XML SDK类 |
| p | Paragraph |
| pPr | ParagraphProperties |
| r | Run |
| t | Text |

Paragraph类

Open XML SDK中的Paragraph类代表为WordprocessingML文档定义的Open XML文件格式schema中的段落（<p>）元素，使用Paragraph对象来操作WordprocessingML文档中的<p>元素。

ParagraphProperties类

在WordprocessingML中，一个段落的属性是通过段落属性（<pPr>）元素指定的，Open XML SDK中的ParagraphProperties类代表<pPr>元素。

Run类

一个Word文档中的段落一般都会包含文本，在WordprocessingML文档的Open XML 文件格式schema中，run（<r>）元素被用来界定一个文本区域。在Open XML SDK中，Run类代表<r>元素。

Text类

在<r>元素内部，是文本(<t>)元素，文本元素是组成文档内容文本的容器。在Open XML SDK中，Text类代表<t>元素。

以下代码实例化了一个Open XML SDK中的Paragraph类对象并使用该对象添加一段文本到WordprocessingML文档中：

public static void WriteToWordDoc(string filepath, string txt)

{

// 使用指定的文件路径来打开一个文档并对其进行编辑

using (WordprocessingDocument wordprocessingDocument =

WordprocessingDocument.Open(filepath, true))

{

// 实例化body对象

Body body = wordprocessingDocument.MainDocumentPart.Document.Body;

// 在一个段落中添加一段文本

Paragraph para = body.AppendChild(new Paragraph());

Run run = para.AppendChild(new Run());

run.AppendChild(new Text(txt));

}

}

上述代码执行完之后，写入WordprocessingML文档中的XML如下所示：

<w:p>

<w:r>

<w:t>测试文本</w:t>

</w:r>

</w:p>