**语法**

1. Pdf语法基于什么字符集？

PDF文件就是一个8位字节序列。这些字节序列可以根据如下语义规则分组成若干token。一个或多个token组成一个更高一级的语义实体（主要是构成PDF文档基本数据的对象）。PDF文件可全部由可见可打印ASCII字符子集和空白符（例如Space、Tab、回车、换行）对应的8位字节来表示。但PDF文件并不局限于ASCII字符集，实际上它可以包含任意8位字节，只要遵循如下要求即可：

a.分隔对象和描述文件结构的token全以ASCII表示，所有保留关键字和标准字典中的键值名也以ASCII表示。

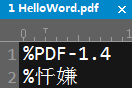
b.特定类型对象的数据值可以但不必全为ASCII，但出于可读性，最好用ASCII表示。在实际中，一些本身就是

二进制的数据，例如图像，出于压缩和效率，最好直接以二进制表示。

c.包含二进制的PDF文件必须在传输和存储过程中保留所有原始字。

2. 基于HelloWord.pdf说明Pdf文件结构？

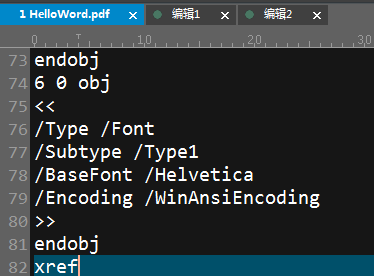
a.文件头：



这个文件头表明了这个pdf文件使用的版本号，后面的乱码则表面了在该pdf文件当中包含有二进制数据，所以在版本号之后紧跟了4个字符的二进制数据，由于转码的问题，所以显示的是表示中文字符的乱码。

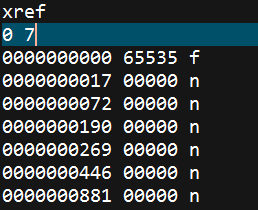
b.文件体

除了文件头，在xref标志之前的所有内容都属于文件体



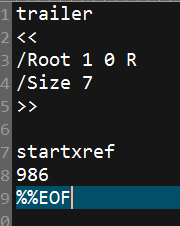
c.交叉表

交叉应用表包含一些信息，使得应用程序可以随机访问文件中的任何间接对象，而无需读取整个文件。helloword.pdf示例中包含了一个交叉应用节，其含有一个子节，子节中有7个项，其中6个项是有效项，1个项是空闲项。

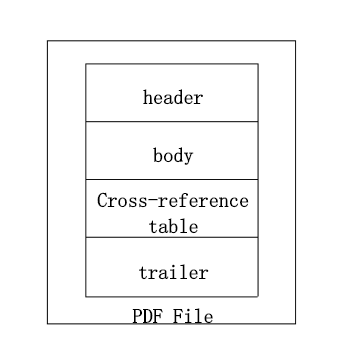


d.文件尾

一个PDF文件的尾部使得一个PDF应用程序快速定位它的交叉应用表和某些特定对象的位置。PDF应用程序应该从PDF文件的尾部开如解析。PDF文件的最后一行只包含一个文件结尾标志%%EOF。结尾标志的前面两行分别是关键词startxref和最后一个交叉应用节的偏移地址（从文件开始算起）。关键词startxref前一行是一个字典对象，其由一个关键词trailer和紧随其后的由《》括起来的一系列键值对组成。如下图helloword.pdf所示的文件尾：



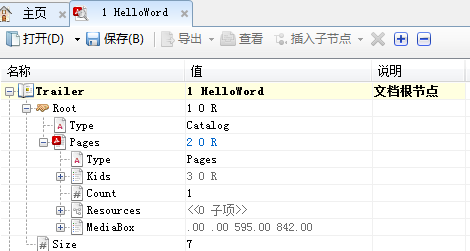
pdf大致的文件结构如下图所示：



3. 基于HelloWord.pdf说明Pdf文档结构？

文件结构反映的是二进制数据在存储介质上的存放格式，是一个物理概念；而文档结构反映的是信息的一种逻辑上的层级关系，是一个逻辑概念。

1. 文档目录



4. 基于HelloWord.pdf说明支持的基本类型？

5. 工具使用PdfPatcher、UltraEdit