

cfs&&cfe

索引文件.cfs、.cfe被称为复合（compound）索引文件，在IndexWriterConfig可以配置是否生成复合索引文件，默认开启。

在前面的文章中，我们介绍了其他的[索引文件](#)，而复合索引文件则是将这些索引文件的数据组合到一个文件中，这种设计的目的是为了减少文件描述符的使用。

正如上文中描述的那样，复合文件实质是索引文件的组合，意思是无论是否设置了使用复合文件，总是先生成非复合索引文件，随后在flush阶段，才将这些文件生成.cfs、.cfe文件，其中[.liv](#)、[.si](#)索引文件不会被组合到.cfs、.cfe中。

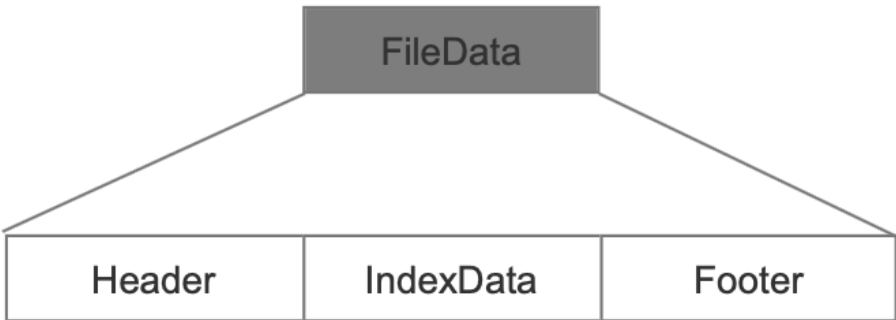
cfs文件的数据结构

图1：



FieldData

图2：

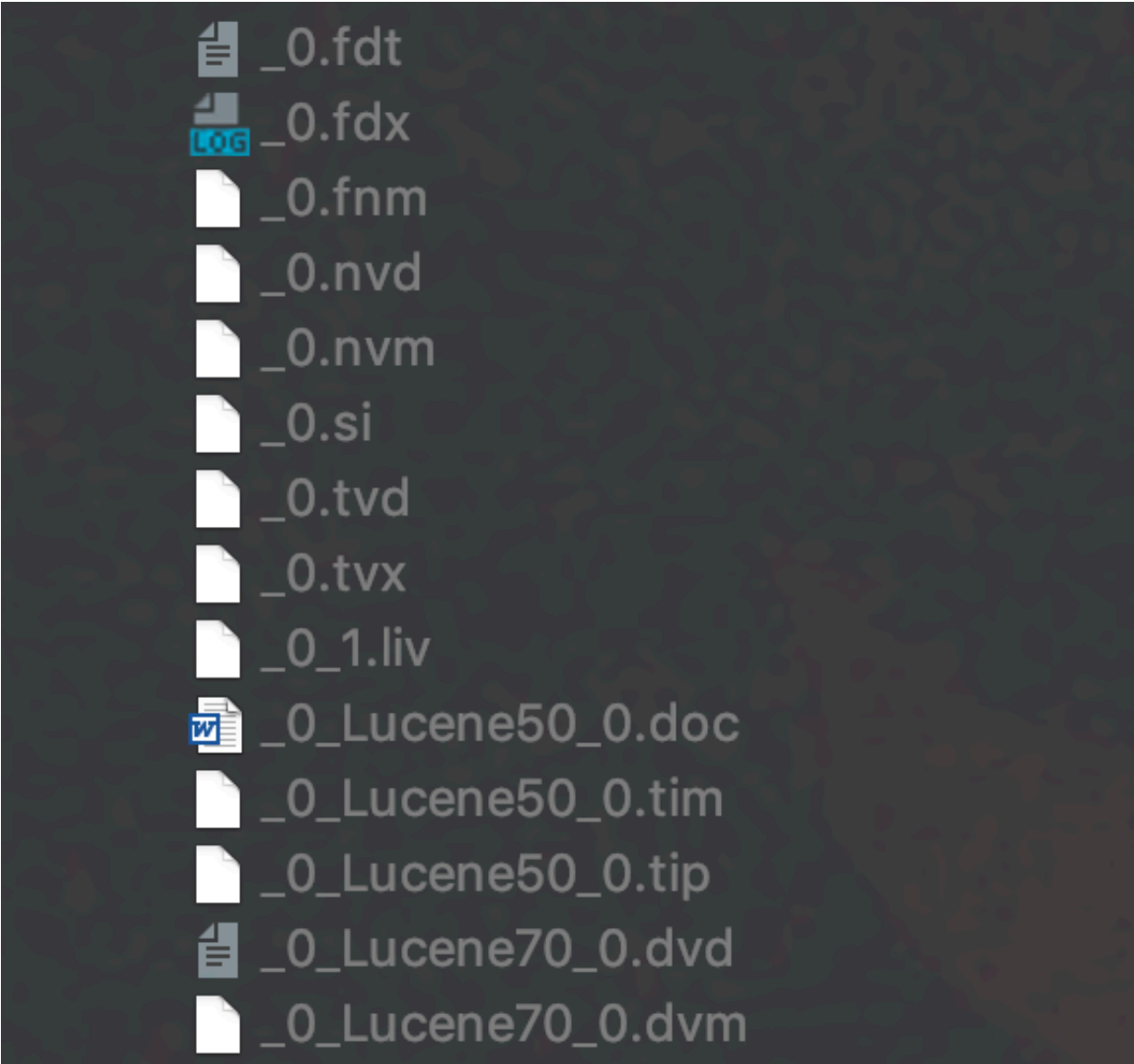


FieldData为非复合索引文件的数据。原非复合索引文件的Header、Footer会被重新计算，只保留有效数据区域IndexData，计算过程不展开介绍（暂时不感兴趣~）。

例子

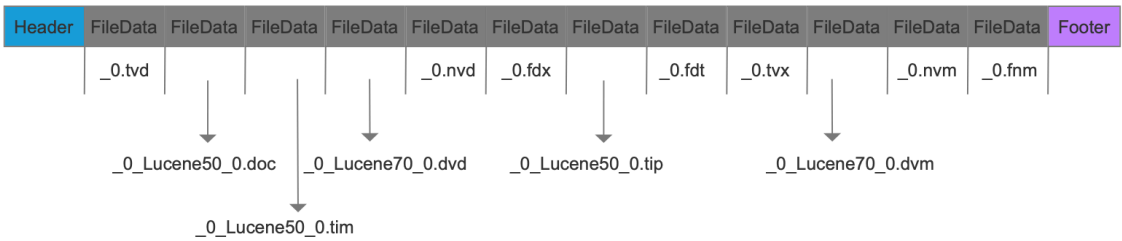
如果一个段中有以下的非复合索引文件：

图3:



生成符合索引文件后 .cfs的数据结构如下:

图4:



遍历一个Set容器，容器的key为非复合索引文件的文件名，根据文件名将其索引信息添加到.cfs文件中，故非复合索引文件在.cfs文件的排列为遍历Set的顺序。

顺序并不重要，因为在读取阶段，总是一次性的读取.cfs文件中的所有内容。

cfe文件的数据结构

图5:

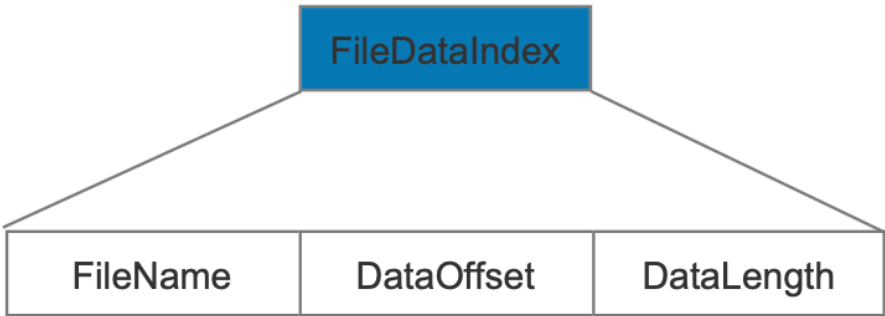


FileCount

该值描述了复合索引文件中包含的非符合索引文件的种类数量，在图3的例子中，该值就是10。

FileDataIndex

图6:



FileDataIndex描述了某个非复合索引文件在.cfs文件中的数据区域。

FileName、DataOffset、DataLength

FileName是非复合索引文件的部分文件名：

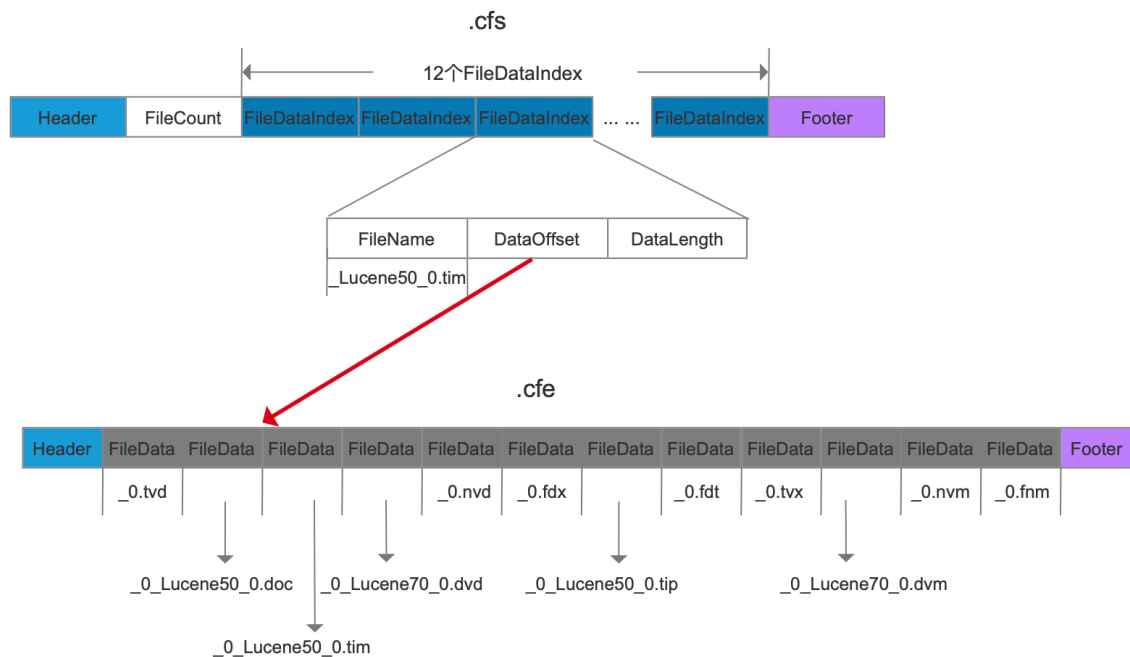
在图3中，_0.tvd跟_0_Lucene50_0.tim对应的FileName分别是 .tvd、_Lucene50_0.tim。

DataOffset为在.cfs文件中的偏移位置，DataLength为非复合索引文件的数据长度，DataOffset跟DataLength就能确定非复合索引文件在.cfs文件中的数据区域。

cfe、cfs文件的映射关系

根据图3中的例子给出以下的映射关系。

图7:



结语

复合文件的数据结构过于简单，写这篇文档的目的是作为在后面介绍flush文章时的一个预备知识。

[点击下载附件](#)