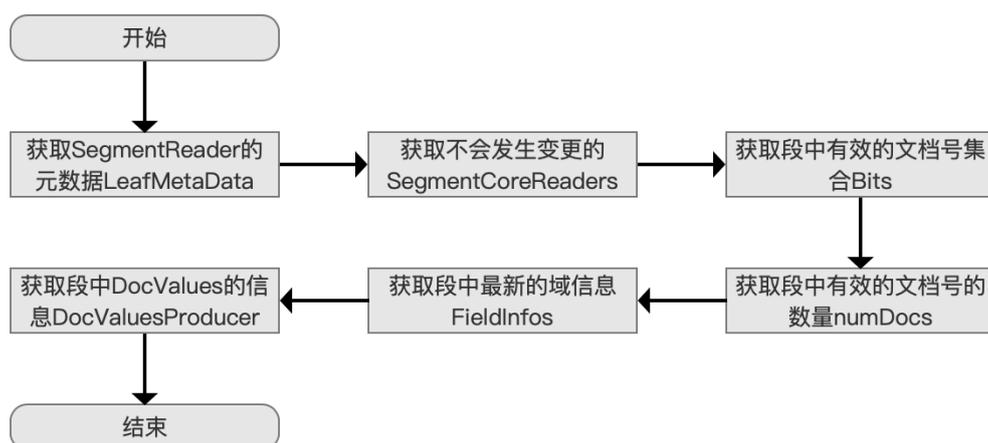


# SegmentReader (二)

本文承接[SegmentReader \(一\)](#)，继续介绍生成SegmentReader的剩余的流程。

## 生成SegmentReader的流程图

图1:



## 获取段中最新的域信息FieldInfos

FieldInfos描述了段中所有域的信息，它对应的是[索引文件.fnm](#)中的内容，在[索引文件之fnm](#)的文章中详细介绍了，这里不赘述。

在[SegmentReader \(一\)](#)的文章中我们说到，在图1中的流程点 获取不会发生变更的SegmentCoreReaders，SegmentCoreReaders中已经获得了一个FieldInfos，为什么这里还要获取段中最新的域信息FieldInfos呢：

- 同样地在[SegmentReader \(一\)](#)的文章中我们说到，如果一个段中的索引信息发生更改，那么变更的索引信息会以其他索引文件来描述，即[索引文件之liv](#)、[索引文件之.dvm](#)、[.dvd](#)、[索引文件之.fnm](#)，其中DocValues类型的索引发生更新时，会以[索引文件之.dvm](#)、[.dvd](#)、[索引文件之.fnm](#)来描述变更的索引
- 所以如果段中没有DocValues类型的索引变化时，那么我们就可以完全复用SegmentCoreReaders中所有的信息（见[SegmentReader \(一\)](#)），即可以完全复用下面的信息：
  - StoredFieldsReader：从[索引文件fdx&&fdt](#)中读取存储域的域值的索引信息
  - FieldsProducer：从[索引文件tim&&tip](#)、[索引文件doc](#)、[索引文件pos&&pay](#)中读取域的索引信息
  - TermVectorsReader：从[索引文件tvx&&tvd](#)读取词向量的索引信息
  - PointsReader：从[索引文件dim&&dii](#)中读取域值为数值类型的索引信息
  - NormsProducer：从[索引文件nvd&&nvm](#)中读取域的打分信息
  - FieldInfos：从[索引文件fnm](#)读取域的信息

- 那么如果段中没有DocValues类型的索引变化时，当我们通过DirectoryReader.openIfChange()获取最新的StandardDirectoryReader时，能获得比直接调用DirectoryReader.open()有更高的性能，其实就是大大降低了读取索引文件的I/O开销
- 那么如果段中DocValues类型的索引发生了变化，我们就需要重新读取索引目中的.fnm文件来获得最新的域信息FieldInfos

如何判断段中的DocValues类型的索引发生了变化？：

- 通过索引文件之segments\_N中的字段来获得，如下图所示：

图2：

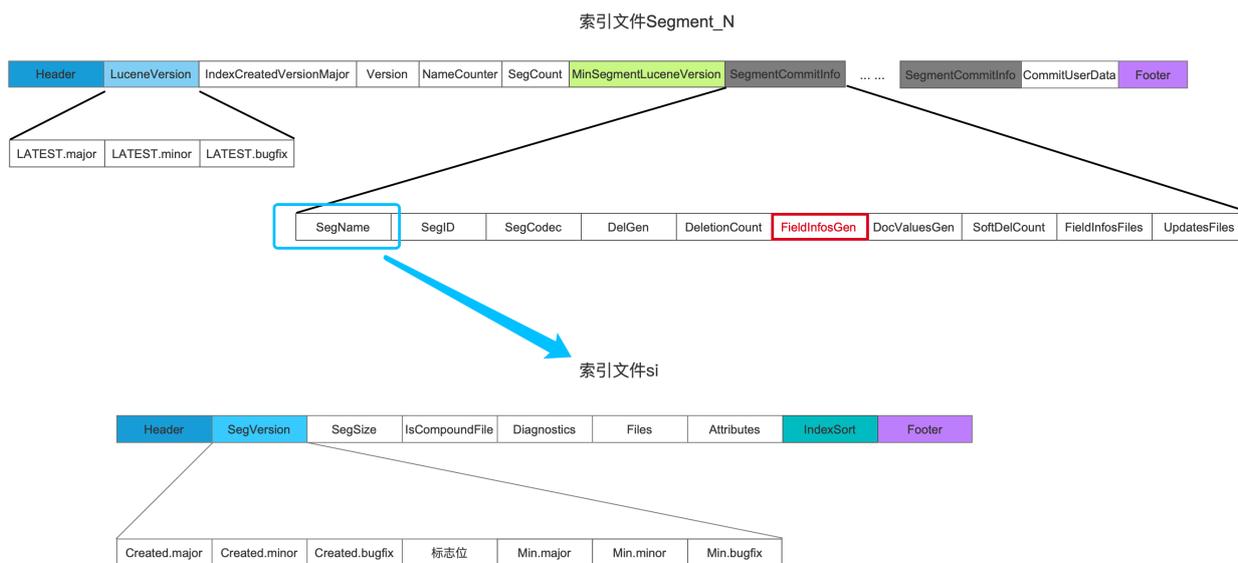


图2中红色框标注的FieldInfosGen的值如果不是 -1，那么说明段中的DocValues类型的索引更新了。

## 获取段中DocValues的信息DocValuesProducer

DocValuesProducer描述了DocValues的索引信息，它通过索引文件.dvd&&dvm获得，在这个流程点我们关注的是如何读取索引文件.dvd&&dvm。

下图描述的是包含了DocValues索引信息的一个段在索引目录中包含的索引文件，并且这里未使用复合索引文件：

图3：

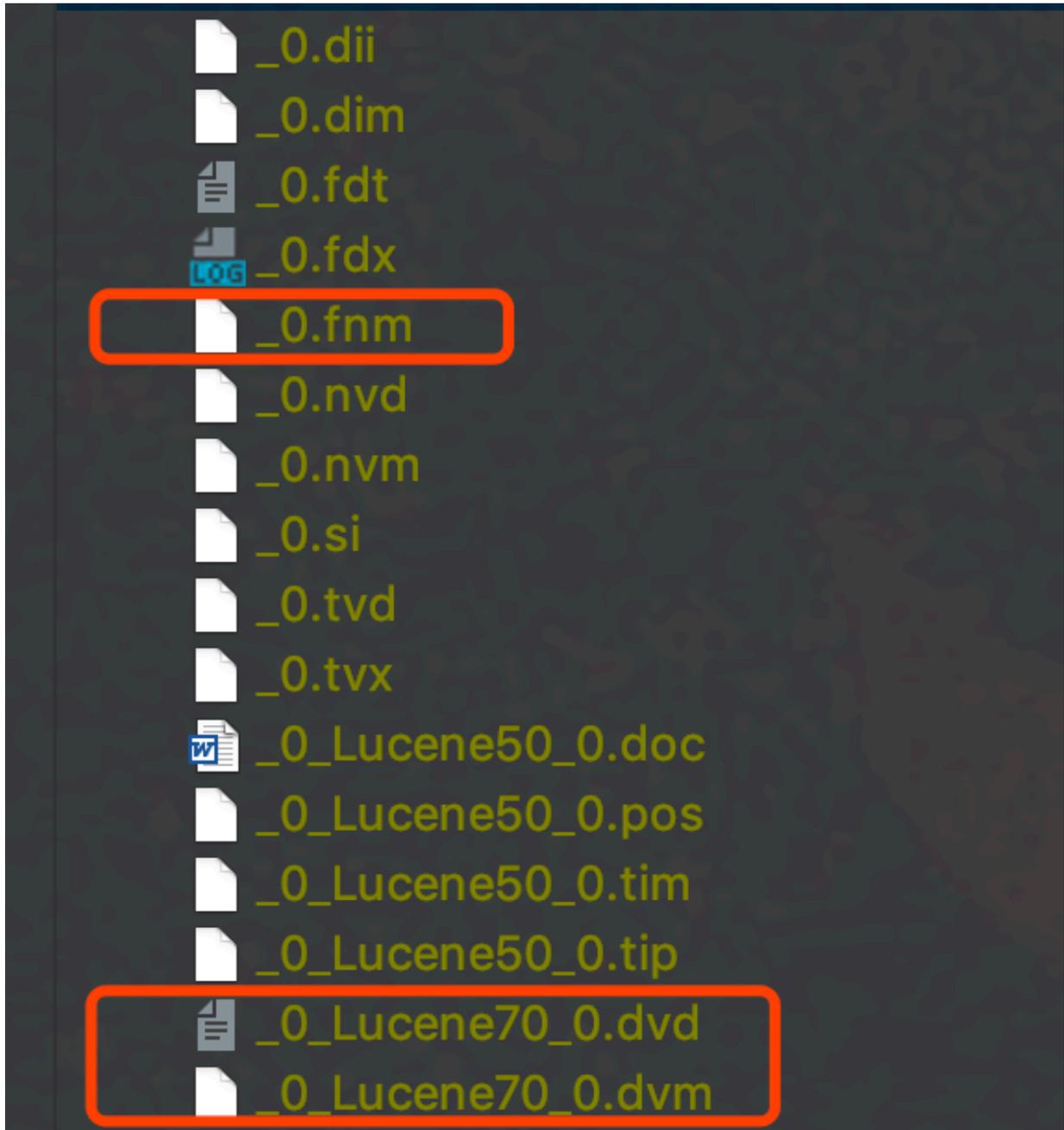
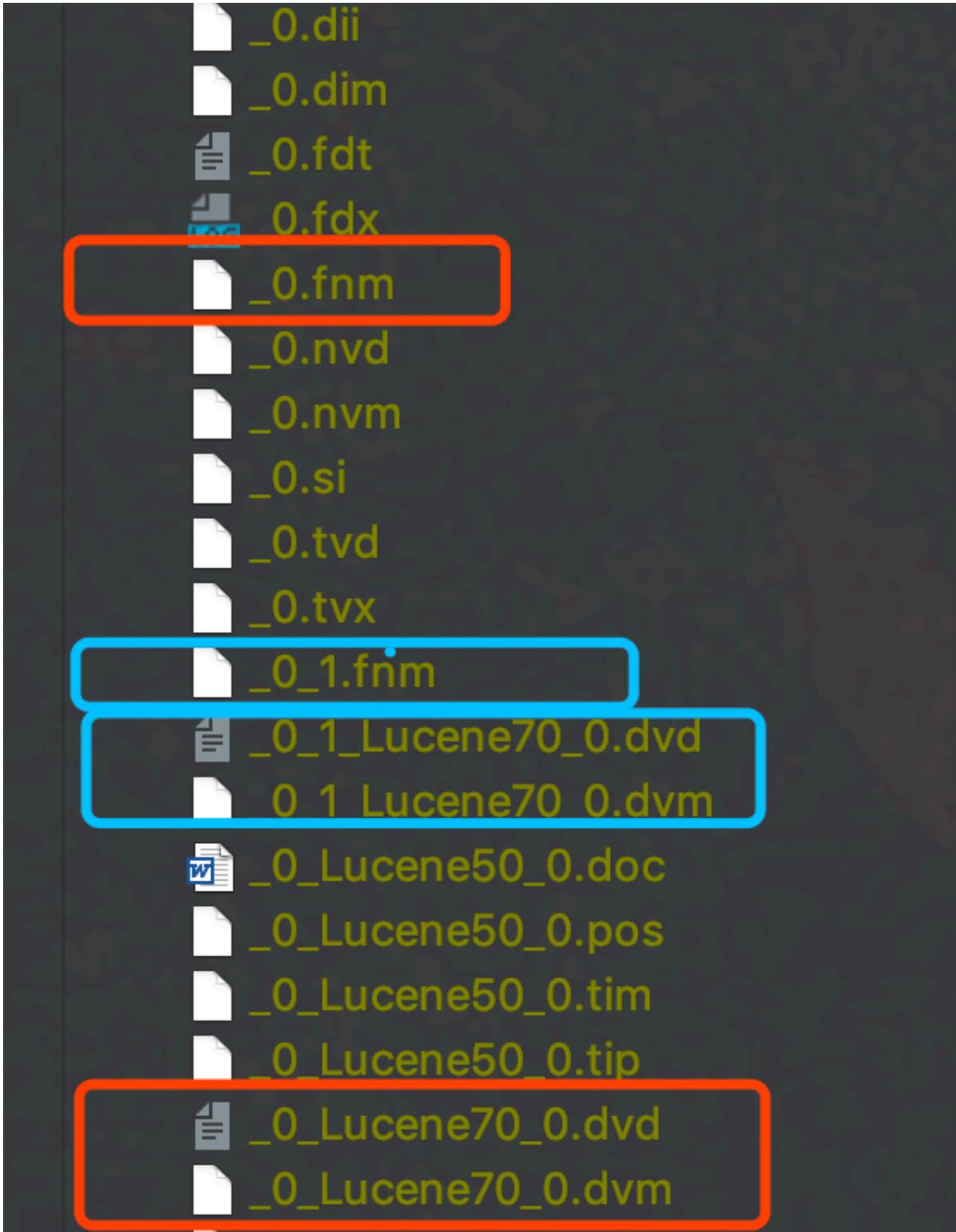


图3中红框标注的索引文件即描述了该段中的DocValues类型的索引信息，那么我们通过读取这两个文件就可以获得DocValuesProducer。

如果该段的DocValues类型的索引信息发生了变更，那么该段包含的索引文件如下所示：

图4：

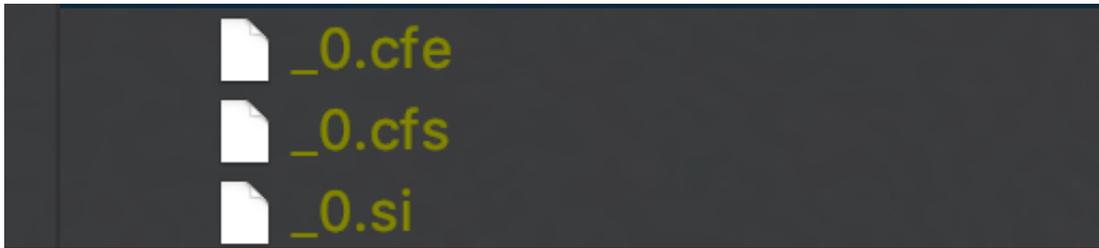


正如我们上文所说的，当段中的DocValues类型的索引信息发生了变更，其变更的内容用[索引文件之.dvm](#)、[.dvd](#)、[索引文件之fnm](#)来描述，即图4中用蓝框标注的3个索引文件“\_0\_1.fnm”、“\_0\_1\_Lucene70\_0.dvd”、“0\_1\_Lucene70\_0.dvm”。

如果我们使用复合索引文件建立索引能更直观的看出DocValues类型的索引信息发生了变更后，索引目录中的索引文件的变化。

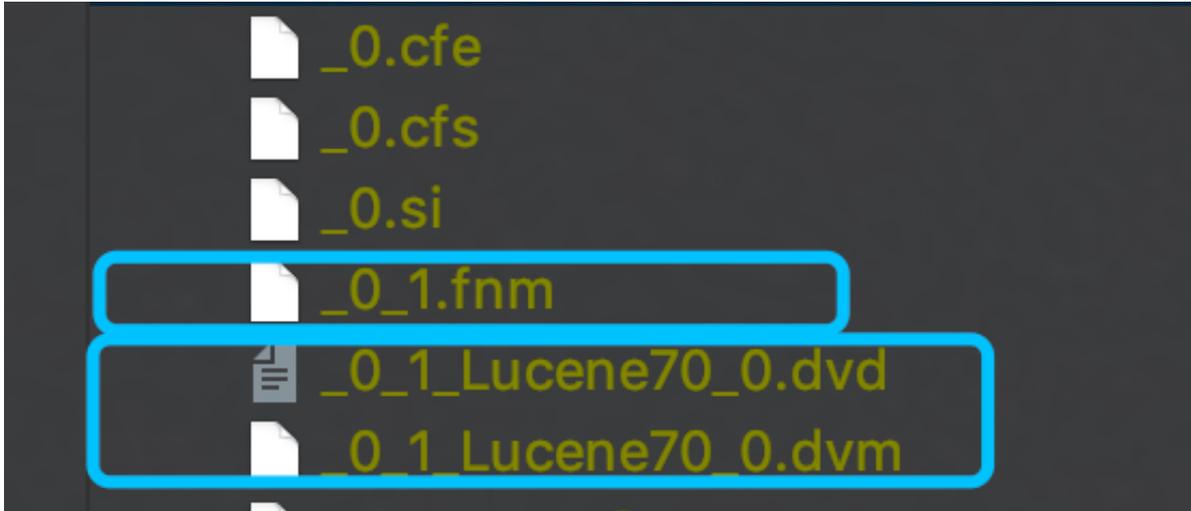
图5为使用复合索引文件的一个段在索引目录中包含的索引文件：

图5：



如果该段的DocValues类型的索引信息发生了变更，那么该段包含的索引文件如下所示：

图6：



不管是否使用复合索引文件，如果该段的DocValues类型的索引信息发生了变更，那么该段中就会包含旧的.dvd、.dvm索引文件文件（“\_0\_Lucene70\_0.dvd”、“\_0\_Lucene70\_0.dvm”）以及新的.dvd、.dvm索引文件（“\_0\_1\_Lucene70\_0.dvd”、“\_0\_1\_Lucene70\_0.dvm”），那么当我们获取这个段对应的SegmentReader时就会读取新的.dvd、.dvm索引文件。

## OpenIfChange()方法

在调用该方式时，如果发现某个SegmentReader（我们称之为旧的SegmentReader）需要更新（见[近实时搜索NRT（三）](#)），那么我们需要获得一个新的SegmentReader，我们会先完全复用旧的SegmentReader中的SegmentCoreReaders、DocValuesProducer，然后根据图1中的Bits以图4中蓝框标注的索引文件作部分的更新。

## 结语

无

[点击](#)下载附件