

倒排索引源于实际应用中需要根据属性的值来查找记录。这种索引表中的每一项都包括一个属性值和具有该属性值的各记录的地址。由于不是由记录来确定属性值，而是由属性值来确定记录的位置，因而称为倒排索引(inverted index)。带有倒排索引的文件我们称为倒排[索引文件](#)，简称[倒排文件](#)(inverted file)。

例如: 先将很多文章切词, 得到切后的词与来源文章的编号维护成列表, 这样根据某个词就能快速找到文章. 这就是核心

列表中也可以保存单词出现的次数, 来源文章等等信息, (百度不知道都存了什么, 反正肯定很牛逼)

以单词“拉斯”为例，其单词编号为8，文档频率为2，代表整个文档集合中有两个文档包含这个单词，对应的倒排列表为：{(3;1;<4>), (5;1;<4>)}, 其含义为在文档3和文档5出现过这个单词，单词频率都为1，单词“拉斯”在两个文档中的出现位置都是4，即文档中第四个单词是“拉斯”。

原文：<https://blog.csdn.net/u012965373/article/details/39118483>