

倒排索引，是适合用于进行搜索的

倒排索引的结构

1. 包含这个关键字的 document list
2. 包含这个关键字的所有 document 的数量，IDF (inverse document frequency)
3. 这个关键字在每个 document 中出现的次数：TF (term frequency)
4. 这个关键字在这个 document 中出现的次数
5. 每个 document 的长度：length norm
6. 包含这个关键字的所有 document 的平均长度

word	doc1	doc2
how	*	*
are	*	
you		*

倒排索引并不是这么简单，里面还包括上面的结构，但倒排索引大概就是个这样的东西

倒排索引不可变的好处

1. 不需要锁，提升并发能力，避免锁的问题
2. 数据不变，一直保存在 os cache 中，只要 cache 内存足够
3. filter cache 一直住在内存，因为数据不变
4. 可以压缩，节省 cpu 和 io 的开销

倒排索引不可变的坏处：每次都要重新构建整个索引