倒排索引：

在搜索引擎中，每个文档都有一个对应的文档ID，文档内容被表示为一系列关键词的集合。例如，文档1经过分词，提取了20个关键词，每个关键词都会记录它在文档中出现的次数和出现位置。

那么，倒排索引就是关键词到文档ID的映射，每个关键词都对应着一系列的文件，这些文件中都出现了关键词。

举个栗子。

有以下文档:

DocId	Doc
1	谷歌地图之父跳槽 Facebook
2	谷歌地图之父加盟 Facebook
3	谷歌地图创始人拉斯离开谷歌加盟 Facebook
4	谷歌地图之父跳槽 Facebook 与 Wave 项目取消有关
5	谷歌地图之父拉斯加盟社交网站 Facebook

对文档进行分词之后，得到以下**倒排索引**。

WordId	Word	DocIds
1	谷歌	1,2,3,4,5
2	地图	1,2,3,4,5
3	之父	1,2,4,5
4	跳槽	1,4
5	Facebook	1,2,3,4,5
6	加盟	2,3,5
7	创始人	3
8	拉斯	3,5
9	离开	3
10	与	4
..	..	..

另外，实用的倒排索引还可以记录更多的信息，比如文档频率信息，表示在文档集合中有多少个文档包含某个单词。

那么有了倒排索引，搜索引擎可以很方便地响应用户的查询。比如用户输入查询Facebook，搜索系统查找倒排索引，从中读出包含这个单词的文档，这些文档就是提供给用户的搜索结果。

要注意倒排索引的两个重要细节：

- 倒排索引中的所有词项对应一个或多个文档。
- 倒排索引中的词项根据字典顺序升序排列。