

06 | 如何用Elasticsearch构建商品搜索系统？

搜索这个特性可以说是无处不在，现在很少有网站或者系统不提供搜索功能了，所以，即使你不是一个专业做搜索的程序员，也难免会遇到一些搜索相关的需求。

搜索这个东西，表面上看功能很简单，就是一个搜索框，输入关键字，然后搜出来想要的内容就好了。搜索背后的实现，可以非常简单，简单到什么程度呢？我们就用一个 SQL，LIKE 一下就能实现；也可以很复杂，复杂到什么程度呢？不说百度谷歌这种专业做搜索的公司，其他非专业做搜索的互联网大厂，搜索团队大多是千人规模，这里面不仅有程序员，还有算法工程师、业务专家等等。

二者的区别也仅仅是，**搜索速度的快慢以及搜出来的内容好坏而已**。今天我们就以电商中的**商品搜索**作为例子，来讲一下，如何用 ES(Elasticsearch) 来**快速、低成本**地构建一个体验还不错的搜索系统。

理解倒排索引机制

这个倒排索引怎么构建的呢？当我们往 ES 写入商品记录的时候，ES 会先对需要搜索的字段，也就是商品标题进行分词。分词就是把一段连续的文本按照语义拆分成多个单词。然后 ES 按照单词来给商品记录做索引，就形成了上面那个表一样的倒排索引。

为什么倒排索引可以做到快速搜索？我和你一起来分析一下上面这个例子的查找性能。这个搜索过程，其实就是对上面的倒排索引做了二次查找，一次找“苹果”，一次找“手机”。注意，整个搜索过程中，我们没有做过任何文本的模糊匹配。

ES 的存储引擎存储倒排索引时，肯定不是像我们上面表格中展示那样存成一个二维表，实际上它的物理存储结构和 MySQL 的 InnoDB 的索引是差不多的，都是一**颗查找树**。对倒排索引做两次查找，也就是对树进行二次查找，它的时间复杂度，类似于 MySQL 中的二次命中索引的查找。显然，这个查找速度，比用 MySQL 全表扫描加上模糊匹配的方式，要快好几个数量级。

