

# 搜索排序 = 相关性排序?

原创 XG数据 WePlayData 2020-03-25

提起排序，可能很多Ranker条件反射的脑海里蹦出相关性。是的，相关性是衡量doc与用户检索需求的匹配，其在排序中的地位永远是第一位的。但仅仅做好相关性就够了吗？本文介绍搜索引擎还会对doc从其它哪些维度来排序。

**1. 时效性：**当用户在阿里发财报的当天去搜“阿里财报”时，之前发布的财报旧文章从相关性上来说肯定没问题，而用户是想了解最新的财报，所以需要有限把新文章排到前面，这里称为时新性维度，其是衡量搜索引擎对突发性、周期性事件的快速响应程度。时效性检索在总搜索次数中可能占大头，毕竟都是热点事件导流的搜索需求。

**2. 权威性：**耳熟能详的PageRank便是为此而生，更泛的去理解，权威性可以理解成内容质量，如果用户看了一篇高质量的文章、权威的官方解答，相比于看了一篇标题党文章，其体验肯定是截然不同，这里成为权威性，其是衡量搜索引擎对内容质量的理解程度。同时内容质量也影响着搜索生态是否能健康发展，百度医疗搜索便是一个活生生的例子。

**3. 多样性：**多样性又可分为话题多样性和资源类型多样性。话题多样性指当用户输入泛query时，其特点是无明显意图，但存在较多潜在意图。比如搜“糖尿病”，比较理想结果是top结果把关于糖尿病的病因、症状、预防、药物...等都能覆盖，则是一个好的话题多样性。资源类型多样性是指，相比于都是文章类的结果，混合着文章、视频、音频等的搜索结果用户体验会更好。

**4. 场景性：**场景性则是具体搜索业务绑定的，也可理解成规则调序。比如置顶、插入在固定位置、根据检索和资源类型预定义一些槽位映射等。尤其当出现类似新冠肺炎的突发事件时，短时间排序模型很难立即调整，通常需要规则调序先顶上。

时效性、权威性和多样性的前提都是必须保证满足相关性。相关性排序保证相关>不相关，而时效性排序保证相关&新>相关&旧>不相关；权威性排序保证相关&高质量>相关&低质量>不相关；多样性排序保证相关&多样>相关&重复>不相关。搜索排序是个相互影响的复杂系统，不能通过一个端对端模型就能解决。因此后续会依次介绍搜索排序如何从这些维度在哪些模型返回令用户满意的结果。

## 相关阅读

1. Query理解 - 搜索引擎“更懂你”
2. 搜索引擎新的战场 - 百度、头条、微信
3. 搜索引擎的两大问题（1） - 召回