导读 | 三步走策略，轻松搞定检索！

2020-03-23 陈东

检索技术核心20讲 进入课程 >

讲述：陈东
时长 13:15 大小 12.15M

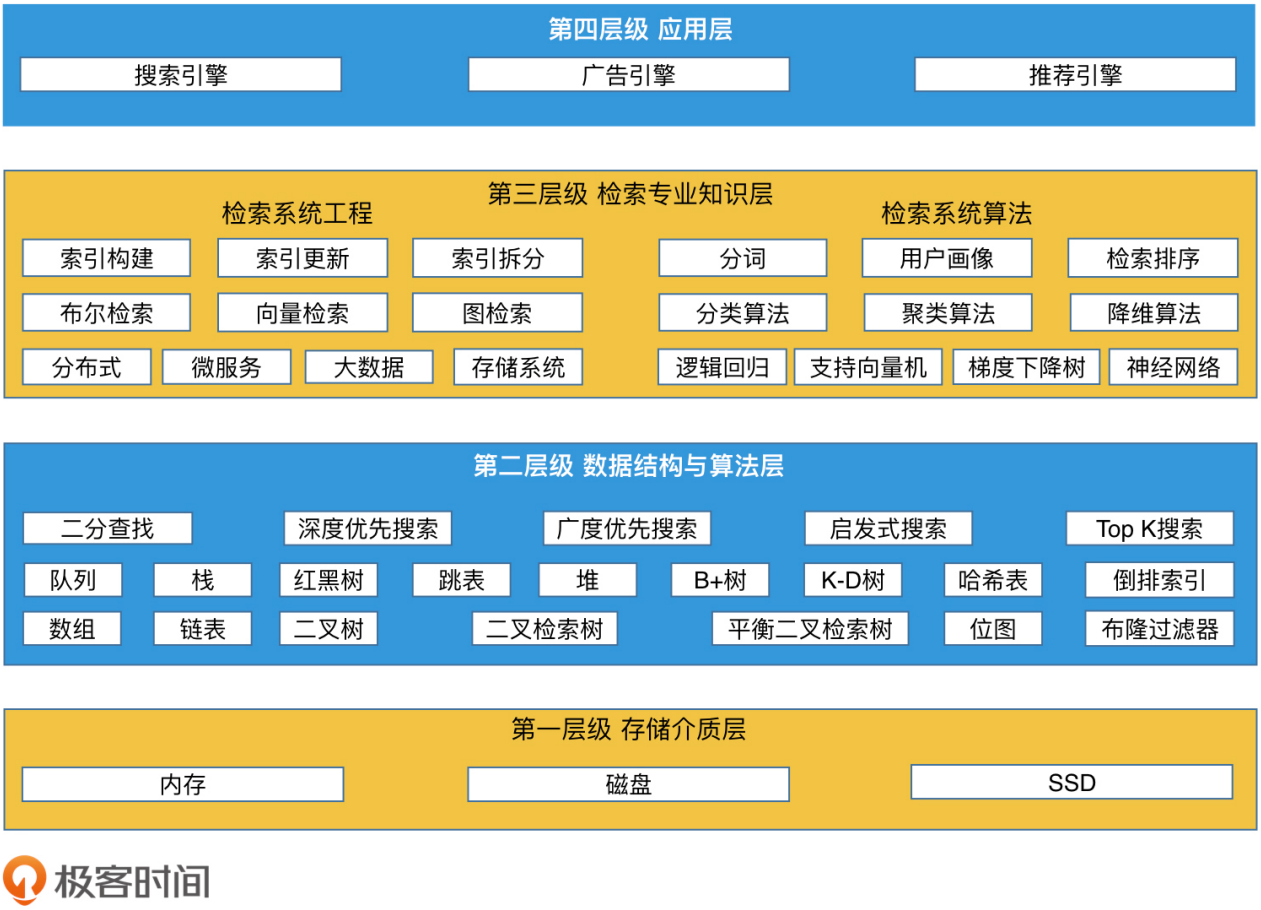
你好，我是陈东。欢迎来到《检索技术核心 20 讲》。

今天是课程导读，在正式开始学习检索技术之前，我想和你先聊聊这个专栏的学习方法，目的就是让我们后面的学习能达到事半功倍的效果。

想要高效地学习检索，我的经验是咱得先弄清楚到底都要学哪些内容，给自己一张知识地图，才能做到心中有数。

这里，我根据十多年的工作经验，梳理了和我们的工作有较强相关性的检索知识，并整理出了一张知识全景图，你可以看一看。

检索技术知识全景图



极客时间

在这张图中，我从基础到实际应用，将需要学习的检索技术分为了四个层级，我们按照从下往上的顺序依次来看。

第一层是**存储介质层**。因为检索效率的高低和数据存储的方式是紧密联系的，所以，存储介质的特性是我们需要学习的基础知识。

第二层是**数据结构与算法层**。提到“效率”，自然就离不开数据结构和算法。在遇到实际业务的时候，我们要知道如何利用每个数据结构和算法的特点，来提高检索效率。所以，这块内容我们必须要学得很扎实。

第三层是**检索专业知识层**。如果我们想实现工业界中的检索引擎，需要掌握这些检索技术。我把它们划分为两部分，分别是**工程架构和算法策略**。这些内容是我们解决常见业务问题的必备知识。

第四层是检索技术的**应用层**。检索技术在互联网中有许多应用场景，其中最常见的，有搜索引擎、广告引擎、以及推荐引擎。这些业务系统有相似的工程架构和算法部分，也分别有自

己独特的业务处理环节。学习它们的实践方法，我们可以更全面、更深入地掌握检索技术。

你可能想说，这些内容看起来可真不少啊！其实，这张图已经是我精简后的了，如果你想要全面掌握检索技术，这张图里提到的每一个方向，你都可以非常深入地去学习。

这么一看，即使有了这张全景图，学习检索依然不容易。那如果想要高效学习，我们具体又该怎么做呢？我们需要有清晰的学习路径和科学、高效的学习方法，我把它们称为“三步走策略”。接下来，我就和你具体来聊一聊。

第一步：夯实基础

万丈高楼平地起。我们刚才说过，检索技术的底层基础知识，离不开数据结构和算法。在检索领域里最常用的基础数据结构和算法，主要有：数组、链表、位图、布隆过滤器、哈希表、二叉检索树、跳表、倒排索引。

因为检索技术本质上就是将数据从存储的地方高效取出的技术，而数据是以什么数据结构存储的，会直接影响到检索效率。所以，对于这些基础知识，我们要重点关注它们的存储特点和检索效率。从中，我们不仅可以学会在合适的场景使用合适的技术，还可以掌握检索的核心设计思想，从而在代码层面提高检索效率。

比如说，数组和链表是最基础的数据结构。从存储特点上来说，它们都属于线性结构。而从检索效率上来说，在数据无序存储，并且本身结构没有做任何优化的情况下，数组和链表的检索效率是 O(n)。那想要提高检索效率，我们就需要结合它们自身的特点，将二分查找的思想加入进去，将检索效率提升到 O(log n)。这其中就体现了检索的核心设计思想：合理组织数据，尽可能快速减少查询范围，提升检索效率。在具体写代码的时候，如果我们能应用这样的设计思想，那检索效率肯定会有大幅提升。

第二步：在实践中将技术落地

多年的工作经验告诉我，工业界的技术落地和基础知识之间还是有很大差距的。解决实际问题的能力，往往是衡量一个工程师水平的标尺。如果我们想加强自己的技术经验积累，除了打好基础，更重要的是要从实践中学习工业界的解决方案。

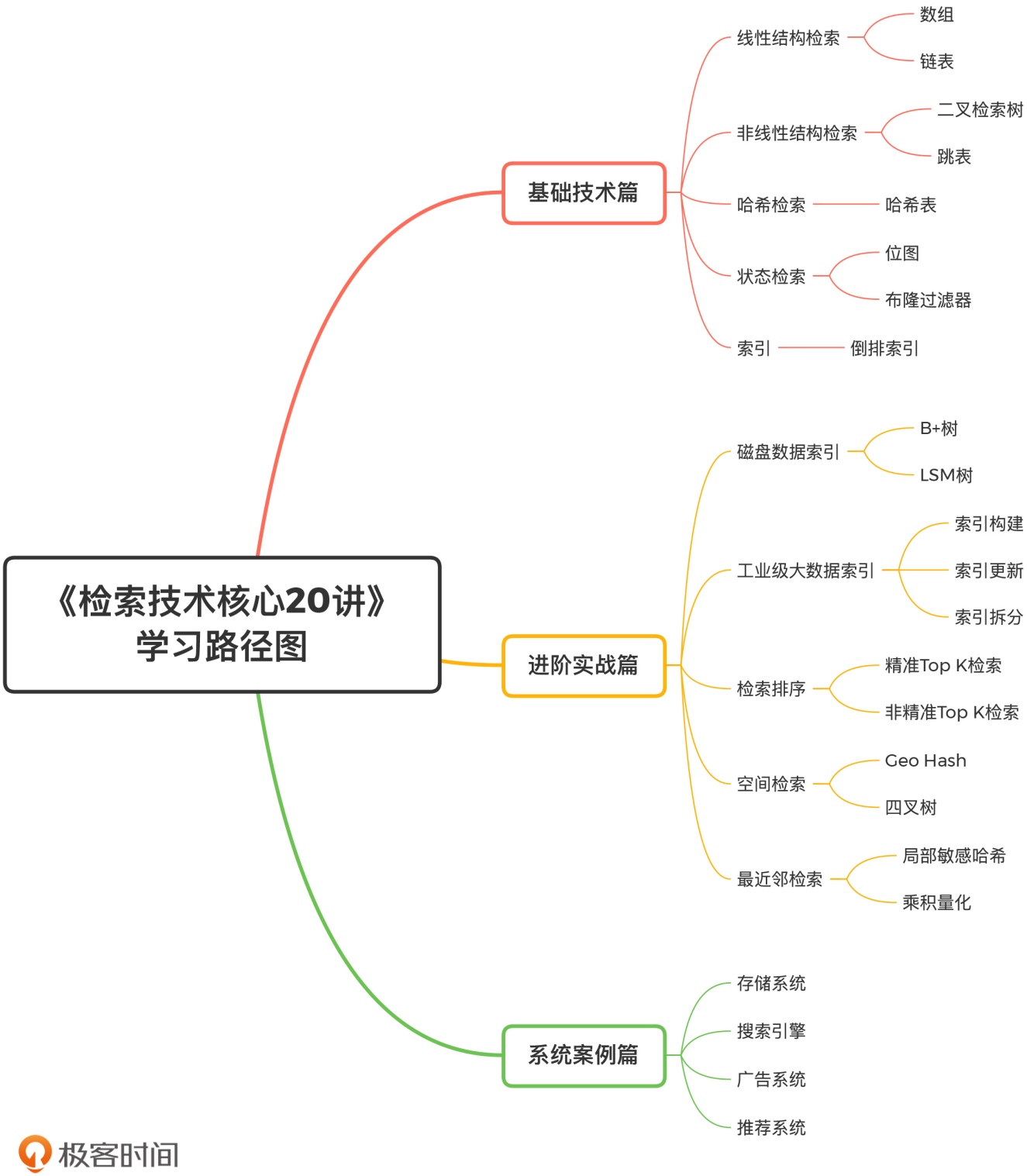
在开篇词中我们说过，检索技术是数据库、搜索引擎、广告引擎和推荐引擎等热门业务系统的底层技术。所以，这些场景中的实际业务需求，都可以作为我们学习检索技术的“题

库”。在“解题”的过程中，我们要重点关注工业界中，针对不同场景的高效检索技术，和热门业务系统中检索架构的设计方案，以及它们各自的特殊处理环节。从中，我们不仅能学到对应的行业经验，还可以了解不同行业中检索架构特点，以此来解决工作中的实际检索难题。

比如说，“刚发布的文章为什么能被搜到”，这就是一个典型的检索实战题，它主要用到的知识就是索引更新。在工业界中实现索引更新的时候，为了追求更高的检索性能，我们一般不会直接对索引加锁，而是会利用“**双 buffer 机制**”来实现索引更新。但是像搜索引擎这样万亿级网页的索引规模，无法直接使用“双 buffer 机制”来更新，需要使用“**全量索引结合增量索引**”方案来更新索引。

再比如说，当我们要在系统中使用数据库来进行存储和检索时，那么是使用关系型数据库好呢？还是选择 NoSQL 好呢？这就需要我们对于数据库的检索技术 B+ 树，和 NoSQL 中的 LSM 树有一定的了解。比如，在数据被频繁写入、较少查找的日志系统和监控系统中，我们更应该使用 NoSQL 型数据库。

知道了学习的重点在哪儿，我们就可以很容易地梳理出高效学习的路径了。本专栏就是从这样的角度出发，借助三个模块，帮助你建立全面的检索知识体系。



第三步：搭配高效学习攻略

那除了高效的学习路径，我也有一些学习方法想分享给你。我一共总结了 3 条攻略，希望它们能够帮助你更轻松、高效地学习这个专栏。

1. 多思考、多提问，善用“理解记忆法”

平时，我们想要记住一篇文章中新学的知识点的时候，很多人会说：多读几遍，你就能记住这些知识点了。但事实证明，任何不经过深刻理解的知识点都是“留不住”的，过一阵子我们就会把它忘了。即使我们短时间内没有忘记，“记住”也不等于“学会”。

那对于我们专栏来说，在面对复杂的检索知识的时候，我更建议你通过理解记忆的方式进行学习。具体的方式有啥呢？我比较推荐问答的方式。也就是说，在学习每个知识点的时候，你可以一直问自己几个问题，比如，“这个知识点要解决什么问题？”“如果不用这个方法还有其他的解决方案吗？”“使用这个方法有副作用或者限制吗？”。

慢慢地，你会发现，这种问自己问题的学习方式，不仅能帮你“学会”知识，更重要的是，它还能训练你的思考能力和理解能力。当你在学别的知识的时候，依然可以用同样的方法。

2. 建立自己的知识体系

学会了怎么“记”还不够，因为，随着我们学习的不断深入，知识点会越来越多，而不成体系的学习，往往会事倍而功半。所以，我们需要把它们有效地组织起来，有体系地学习。那么问题来了，我们该如何建立自己的知识体系呢？一般来说，我会用这么两个小技巧：**对比和拆解**。

在学习一个新知识点的时候，我们可以把它和之前学过的知识点**对比**，看看它们之间的相同点和不同点，为新、旧知识之间建立联系。

那如果这个新知识点是一个比较复杂的知识点，我们可以试着把它**拆解**成多个小知识点，拆解之后，我们依然可以用对比的方法，让这些小知识点和旧知识建立联系。

借助这两个小技巧，你就能将零散的知识点关联起来，从而形成一个自己的知识体系了。

3. 有耐心、反复学、多交流

我相信，通过前面这些方法你已经建立起了学习的信心。那这里，我就要给你泼一盆“冷水”了。在刚开始学习检索的时候，就算已经使用了前面的学习方法，我们还是会遇到一种情况：不能完全掌握学到的内容。

不过，你不用担心，这种情况再自然不过了。作为一名“检索老兵”，我想告诉你，实用的学习方法虽然能保证我们少走弯路，但想要真正掌握一项技能，反复学习是非常重要的。尤

其是在新知识点比较多的时候，反复学习就更为重要了。

在反复学习的过程中，你需要多看、多听、多思考、多对比。一方面，这能帮助你加强对检索知识的理解；另一方面，在重复学习的过程中，你也能更好地将前后的知识进行梳理，从而将新知识点融入自己的知识体系中。在这个过程中，你还可以多和同事、同行交流、讨论，弥补自己的知识盲点，做到全面地理解知识。

重点回顾

说了这么多，我已经把学习这个专栏的高手攻略，全部分享给你了。但我还是想再和你强调一些重点内容，希望能帮助你应用到自己的学习和工作中。

首先，我们可以从 4 个层级来学习检索技术，分别是存储介质层、数据结构和算法层、检索专业知识层、以及应用层。这里面的内容很多，要熟练掌握不是一件容易的事情。因此，我们需要有科学、高效的学习方法，以及清晰的学习路径。那我也和你分享了学习这个专栏的一些方法。总结来说就是“三步走策略”：夯实基础、在实践中将技术落地，最后搭配高效的学习攻略。

方向已经指明了，路线也清晰了，接下来，我们就可以开始学习了。那最后呢，我还有一句话想要送给你：道阻且长，行则将至，我们一起努力！

课堂讨论

在我今天给到的检索技术的两张图中，你对其中哪部分知识最感兴趣呢？另外，除了我今天和你分享的学习方法，你也可以在留言区说一说，对于这个专栏，你准备怎么学？

上一篇　开篇词 | 学会检索，快人一步！

下一篇　01 | 线性结构检索：从数组和链表的原理初窥检索本质

精选留言 (8)

写留言

Kǎfkā²⁰²⁰ 置顶
2020-03-23

本节可以看作是给出了检索技术核心这门课的检索

展开

> 作者回复: 没错。其实我们常见的目录，就是一种索引。可见，即便在互联网之前，人们已经对于如何高效进行信息检索进行了许多研究。只要人类社会中存在信息，就会有信息检索技术的用武之地。

铭毅天下（公众号）
2020-03-24

通过本专栏，结合老师的方法，梳理出自己的检索知识体系！

知识盲点要扩散阅读和实践，由不知道变知道，由知道变掌握，由掌握变熟练！

作者回复: 加油！多读多交流多实践。

盘胧
2020-03-24

这可不容易，/(/// ω ///)/加油

展开

伊娃橙
2020-03-24

这个专栏后续会有各大电商网站检索技术架构的对比分析吗

作者回复: 电商平台中，其实既有搜索，又有广告，又有推荐。这些检索技术我在专栏中都有提及。相信你都了解了以后，会对电商平台中的检索技术有更全面的认知。
至于各大电商网站的架构对比，核心框架不会相差太多，但各家都会根据自己的特点做相应的调整。

aoe
2020-03-24

第一次听说：NoSQL 中的 LSM 树！学到新知识了！

展开

ヾ(◍°∇°◍)ﾉﾞ
2020-03-23

在真实的搜索系统涉及的技术应该是非常复杂的吧？有一个真实情况的总结吗？

作者回复: 真实系统的确是非常复杂的。所以系统架构做的事情就是把复杂的大系统拆解成多个简单小系统。而每个小系统就容易实现得多。
由于篇幅有限，这个专栏不会详细介绍一个系统的每个细节，但是会把整体框架和核心知识点和你分享。

每天晒白牙

2020-03-23

基础知识+实践+总结自己的知识库

展开

作者回复: 欢迎一起探讨

pedro

2020-03-23

希望专栏多些实战干货，现在书资料太多偏理论，但实战起来步步艰辛。

作者回复: 从第三篇开始，都是以解决实际问题出发进行介绍。这个专栏的一个目标就是希望结合实践，让技术能落地。