

结束语 | 成长和进化，技术如此，我们亦如此

2020-05-18 陈东

检索技术核心20讲

[进入课程 >](#)



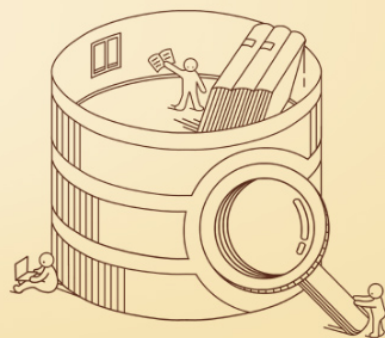
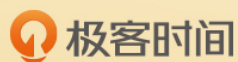
陈东

奇虎360商业产品事业部资深总监

你好，我是陈东。

检索技术是信息世界中很重要的基石。随着信息技术的发展，检索技术也在不断进化。

学习之路没有终点，保持探索和进化，才是我们成长的终极密码。



讲述：陈东

时长 07:54 大小 7.24M



你好，我是陈东。今天，我们走到了这个专栏的最后一个篇章。

一起学习的五十多个日夜，我们从熟悉的数组和链表开始，学到了搜索引擎、广告引擎以及推荐引擎。这就像一场长途跋涉，一路上景色不停切换，不变的，是我们探索这个信息世界的好奇心。

在我看来，信息是构成世界的一个重要维度。实际上，人类文明的开始，就是以拥有了语言和文字为标志的。因为语言和文字的出现，才使得信息可以被记录和传播，文明得以 发展。

而随着 IT 技术的发展，信息的产生速度也越来越快。有报告说，现在 18 个月产生的信息比过去 5000 年产生的信息总量还大。并且，随着互联网的普及和发展，最近每年的信息增长率达到了 50% 以上。我们可以预见 5G 普及之后的未来，信息生产的速度只会越来越快。那么，这么多的信息都有用吗？我觉得不见得。当信息变得庞杂了以后，相应地，信息的含金量就变低了。只有高效地将信息进行检索和提炼，才能取出我们需要的信息，这样的信息才是最有价值的。

在专栏中，我们提到的许多系统都为了“如何高效检索信息”这个基础却又终极的问题，做了精心的设计，从而帮助我们高效地处理信息。**系统如此，人亦是如此。**

在信息泛滥的时代，每个人都有大量的渠道可以接触到大量的信息。如果不进行筛选处理，你的 CPU 根本负荷不过来。比如说，如果你接触的信息都是各种八卦新闻、娱乐视频。那么，你能花在处理专业知识信息上的时间就会变少。甚至在遇到一些虚假信息的时候，如果你不加辨别就接收，这很可能会带来一些负面影响。

所以我们说**这是一个最好的时代，这也是一个最坏的时代**。好的地方在于，有了足够的信息供我们学习和娱乐；坏的地方在于，对于信息的选择和使用，会变成一个更困难的事情，利用不好，甚至会带来负面影响。**因此，我们也要构建起自己思维的检索引擎，学会将有价值的信息提取出来，并加以处理和使用。这样，我们才会不断地前进。**

那无论是系统还是人，我们的前进都必须是跟随着时代的脚步。如果你研究过某一个技术的发展史，你就会对这一点有更深的体会了。其实，在专栏中就隐藏着一条检索技术发展进化的时间线。下面，我来带你回顾一下。

在人类发明了文字以后，书的目录索引和图书馆的图书分类管理，其实就是检索技术在印刷时代的体现。

随着 IT 技术的兴起，信息数字化的变革开始催生出了像数据库这样的系统和 B+ 树这些检索技术。

进入互联网时代后，我们开始处理大数据。无论是基于 LSM 树的存储系统，还是以搜索引擎为代表的检索技术，都是互联网时代的结晶。

到了移动互联网时代，各种基于地理位置的服务不断出现，让我们开始将物理世界和信息世界更紧密地连接在一起。

现在到了 AI 时代，图像识别、人脸识别和智能推荐的发展，又催生出了和深度学习结合的检索系统。

检索技术越发展，涉及的知识越新，你就会感觉越陌生。实际上这很正常，毕竟新生事物要变得普及需要经过时间的发酵。但我们必须保证，我们的认知是随着时代的前进一起进化的。因为一旦我们放弃去学习、去进化，那我们不可避免地就会错过很多机会，甚至会做错许多决策。

因此，面对新生的事物，我们在保持好奇心的同时，还要去升级自己思维的检索引擎，使得我们的认知能跟得上时代的发展。甚至，如果学习的过程能够快人一步，那我们就会享受到时代更多的红利。

好了，说了这么多自我进化的好处。那不断地升级自己的检索引擎，是不是一件很困难的事情呢？完全重构系统的确是一个代价很大的事情，但是在已有系统上进行迭代升级，相对来说会容易许多。

因为，许多所谓的新技术和新系统，都是在一些基础技术上进行组合和微创新得到的。比如，我在加餐 1 中提到的 Roaring Bitmap。它在 2016 年才被发明出来，但是它用到的技术，其实就是很成熟的检索系统中的位图法、哈希表法、跳表法以及倒排索引。还有我们说过的向量检索，其实向量这个数学工具也不是新东西了，只是在近期随着 AI 的发展，才被更广泛地应用起来。

既然新的技术和系统并不难学，那我们到底该怎么学呢？其实，学习和成长就像是爬台阶。高处的目标虽然看上去遥不可及，但只要我们一个台阶一个台阶地往上走，等我们站到了一个足够的高度上再前进就会容易很多。就像让你直接学习 AVL 树和红黑树会很难，但当你了解了二叉检索树的原理和问题以后，再去学习就会简单很多。因此，当你了解了检索中的各种问题和相应技术以后，再去搭建自己的检索系统，或者学习使用开源的检索工具，也就更容易上手。

也正是因为如此，许多高手在面对新事物的时候，能使用自己思维的检索引擎，快速找到核心点去学习和消化，从而可以更上一层楼。

总结来说，成长和进化从来不是一件容易的事情，但只要你能构建起自己认知的检索思维，不断地从海量的信息中快速汲取新的知识，并且从打好地基做起，构建起自己的知识体系。我相信，迈上一个更高的台阶，是一个完全可以预期的事情。

虽然检索技术专栏结束了，但是我希望这个专栏对你的帮助并没有就此停止。学习之路没有终点，保持探索和进化，才是我们成长的终极密码。



陈东

奇虎360商业产品事业部资深总监



不知道在学习过程中，你有哪些体会和评价？
这里有一份专栏调查问卷，邀请你填写。

我们一起继续成长！

去提交

6月-7月课表抢先看

充 ¥500 得 ¥580

赠「¥ 118 月球主题 AR 笔记本」



【点击】图片, 立即查看>>>

© 版权归极客邦科技所有, 未经许可不得传播售卖。页面已增加防盗追踪, 如有侵权极客邦将依法追究其法律责任。

上一篇 20 | 推荐引擎: 没有搜索词, “头条”怎么找到你感兴趣的文章?

精选留言 (10)

写留言



一野 (wx:abley1874...)

2020-05-28

一路学完, 真的非常棒, 感谢老师细心指导。同时大量的例子、图解, 真正做到了深入浅出。期待老师下一次的专栏

展开 ∨

作者回复: 也谢谢你的一路的陪伴和坚持。深入浅出其实也是我写专栏的时候给自己定的目标之一, 你的这个评价让我备受鼓舞。如果文章中有不清楚的地方, 也欢迎继续留言讨论



1



paulhaoyi

2020-05-20

感谢老师精彩的课程，也期待后面更多更好的系列课程，结合AI，最新技术趋势。
哈哈，赶紧出新课吧，我迫不及待订阅了。😊

展开 ▾

作者回复: 哈哈，感谢你的支持。结合AI的检索引擎是一个蛮大和蛮新的话题，是最新的趋势。但这个领域也涉及了很多的知识基础，可能很多人不太好理解。我得想想要如何写才好。如果有新课了，我会第一时间发预告的。



1



牛牛

2020-05-19

首先感谢老师、我觉得这个专栏的质量非常高, 至少在我的心里、它与丁奇老师的MySQL、王争老师的数据结构、文浩老师的计算机原理一样深深影响了我, 用了四期`一周免费读`默默的跟完了课, 受益匪浅, 决定买下来思思品味, 值得反复阅读的一门课, 强烈推荐~~~~

展开 ▾

作者回复: 也非常感谢你的认可和鼓励! 我相信专栏中多少会有不完善的地方, 但整体来说, 我也希望这个专栏能从一个不一样的角度, 让我们能一起探索检索技术是什么和该怎么做。如果能给你带来更多的思考和收获, 那么我也会觉得非常开心和鼓舞。



1



一步

2020-05-18

感谢老师的付出，感谢老师认真回答了每一个评论。

展开 ▾

作者回复: 也谢谢你一直以来积极的提问。其实好的问题，不仅仅帮助自己更好地理解专栏内容，也帮助作者将专栏没讲清楚的内容进行补充说明。从这个角度讲，每个积极提问的人，都是专栏的共建者。谢谢你的支持!



1



青生先森

2020-05-19

首先谢谢老师的付出，经过《数据结构与算法之美》《MySQL实战45讲》以及陈老师的专栏，我越来越喜欢数据库方面的开发，希望老师多出一些专栏，一定支持，期待。

作者回复: 谢谢你的支持。现在我们对于数据库的使用也是混合的机制, 从关系数据库, kv数据库到时序数据库都有(MySQL+TiDB+Redis+aerospike+opentsdb)。你会看到, 数据库相关的技术也是一直在发展。



王坤祥

2020-05-18

作为一个转行的程序猿, 通过这门课, 我从整体上对计算机技术有了更深入的理解。

既然入行了, 需要学习很多的计算机的知识, 前段时间在看资料的时候突然想到在老师的课程中曾经涉及到了一些知识, 然后来到本课程中寻找关联, 知识点相互补充, 我顿时有种豁然开朗的感觉。因为老师的课程是与工程实践紧密联系的, 能够从实际应用的角度...
展开 ▾

作者回复: 很高兴看到这个专栏对你有帮助。我也一直觉得许多理论的知识 and 实际工作结合不够紧密, 导致很多人之前在学校学习教材的时候, 会疑惑这些知识, 比如说各种高级数据结构有什么用。但实际上, 你会发现, 在某些应用场景中, 还真必须使用这样的数据结构才合适。因此, 这个专栏, 我其实也希望能将知识点更好地串起来, 让你能理解这些技术的来龙去脉和相互之间的关系。当然, 由于篇幅限制, 还有许多知识点没有提到。但我相信, 你能掌握好合理的学习方法, 将更多的新知识吸收到自己的知识体系中。



峰

2020-05-18

舍不得技术思维比口音还厉害的陈东老师!!!!!! 我想再听老师讲课肿么办!!!!!!

展开 ▾

作者回复: 哈哈, 不介意口音的话可以多听几遍~
而且未来说不定我也可能在更多的场合和大家继续分享。



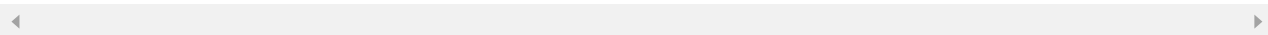
张和祥

2020-05-18

感谢! 虽然只学了一小部分, 但是了解到了很多搜索相关的技术信息和处理方式, 剩下的

就是实践技术了。

作者回复: 谢谢你的支持。专栏的内容其实覆盖面有一定的广度, 不过也有着一定的连续性。因此, 对于已经学习的部分, 希望你直接的工作有帮助; 而还没有学习的部分, 以后可以慢慢看, 希望可以对你的视野和思考有所帮助。



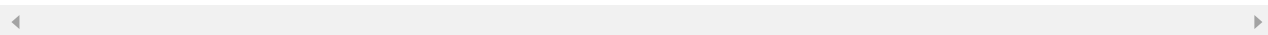
子一

2020-05-18

感谢老师的辛勤付出

展开 ▾

作者回复: 也谢谢你们的支持和鼓励, 这也是我更新的动力。希望专栏内容对你有帮助。



那一刻

2020-05-18

感谢老师的分享, 尤其是老师对于每篇文章的讨论问题和疑问都进行详细的回答, 是我们对于问题有了清晰的认识。再次感谢。

展开 ▾

作者回复: 也感谢你积极地提出各种问题, 帮助我将文章中没有讲到或没有讲清楚的知识点提出来, 让我能进行补充说明。

专栏结束了, 但是学习没有结束。有问题依然可以留言。我也会继续回答。

