通过面试。

你可能会好奇Google的索引大约有多大，我只能告诉你基本的索引，即上百万的索引每一份使用了上万台服务器，而在全世界它有很多份。此外，在亚洲的索引，中日韩文的内容会多些，而在欧洲的法、意、德、西（FIGS）语言的内容会多些。

最后总结一下今天的内容：

1. 通过网络爬虫的例子，你可能对计算机科学和工程的区别有了更多的理解。事实上，任何事情，原理要尽可能简单，而实现起来要尽可能周到。

2. 一个好的工程师，不能简简单单地把那些有明确答案的问题做好，而要能做好没有答案的开放式问题。事实上，我对网络爬虫的理解从进Google到离开Google十多年来也是不断在加深的。我昨天用10分钟就把网络爬虫的原理讲给你听了，但是你真要能写一个好的程序，没有几年的工作经验是办不到的。

3. 大家对今天信息的量级要有体会。在图论出现后的很长时间里，现实世界中图的大小都是在几千个节点以下的规模（比如公路图、铁路图等）。那时候，图的遍历是一件很简单的事情，因此在工业界没有多少人专门研究这个问题。但是等到了互联网出现后，图的大小就从几千增加到上万亿了。

4. 很多数学方法，早期看上去没有什么实际用途，但是随着时间的推移会一下子派上大用场，这就是数学的妙处。

祝近安

吴军

2018年7月25日