# 缓存机制的理解

先来说说缓存。缓存广泛存在于计算机中，例如：作为内存和磁盘交换的媒介、浏览器缓存等。说白了，不过就是利用存储硬件将一些将要用或者被认为将要用的内容保存下来，以便后来可以用得更快。举一个我印象深刻的例子吧。LRU算法应该大家都不会陌生，其核心思想是“如果数据最近被访问过，那么将来被访问的几率也更高”。这个核心思想的由来是通过总结和数据分析得到的，大家可以回想一下，我们访问网站的时候，最常用的一个操作不就是后退吗？还有在生活上，大家会把经常用到的东西放在一个地方，也一样是一个缓存机制了。

在爬虫例子的两个缓存机制实现来看，基本思路是没有太大的不同的，它的区别不过是媒介（文件系统和数据库）、写入读出函数的区别。

这个缓存触发是很简单的，就是下载之后的备份和下载之前的检查是否有备份。对于缓存乃至其他的存储而言，关键点在于存储的组织形式，如果组织形式非常优秀，那么写入和读出就会非常地简洁和方便，所以文件系统实现方式就构建一个目录，目录及目录下的文件的命名都是严格的，便于查找；而数据库实现方式的话，就托管给了数据库系统。应该去选用数据库的实现方式，毕竟数据库是管理数据的利器，但是这又给大家提了一个要求，就是要去了解MongoDB。

这个实现方式更多地利用到了python类的知识，同时也重写了一些类中的特殊方法（双下划线），只要理清它们的调用关系，这些代码读起来也是不难。