**吴军-《数学之美》读书笔记**

1. **搜索引擎的“道”：自动下载尽可能多的网页；简历快速有效的索引；根据相关性对网页进行公平准确的排序。简化其根本是对索引进行布尔运算**
2. **图论与搜索引擎： 遍历算法有广度优先算法（BFS），深度优先算法（DFS）。网络爬虫使用离散表（哈希表）来记录网页是否下否下载过的信息。决定网页下载优先级排序的系统一般称为“调度系统”。**
3. **google的革命性“PageRank”网页排序算法：民主表决式-一个网页被很多其他网页所链接，说明他受的信赖和承认越高，排名就越靠前。网页排名计算原理为矩阵相乘，对零概率或小概率事件进行平滑处理。**
4. **TF-IDF ：信息检索关键词的加权科学方法。TF（Term Frequency）为词频，IDF（Inverse Document Frequency）为逆文本频率指数，log（D/Dw），D为全部网页数，Dw为关键词出现的网页数。**
5. **地址分析和有限状态机， 全球导航和动态规划。**

**6. AK-47：简单有效的哲学，简单有效的方案容易解释方法背后的原理，出错时容易debug，也便于找到今后的改进目标。**