

2006年考试新之辅导搜索引擎的第三定律 PDF转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/62/2021_2022_2006_E5_B9_B4_E8_80_83_c40_62592.htm 搜索引擎走到今天，已经是一个结束过去，开辟未来的时候了。为了说清楚我所讲的第三定律，我们先来回顾一下第一和第二定律。第一定律相关性定律听起来象是一篇学术论文，的确，就连第一，第二定律的提法以前也没有过，但是第一，第二定律的内容确早已在业界和学术界得到了公认。其实这第一定律是早在互联网出现之前就被学术界广泛研究过的，那就是所谓的相关性定律。这个领域那时叫情报检索，或信息检索，也有叫全文检索的。那时的相关性都是基于词频统计的，也就是说，当用户输入检索词时，搜索引擎去找那些检索词在文章（网页）中出现频率较高的，位置较重要的，再加上一些对检索词本身常用程度的加权，最后排出一个结果来(检索结果页面)。早期的搜索引擎结果排序都是基于本文的第一定律的，如Infoseek，Excite，Lycos等，它们基本上是沿用了网络时代之前学术界的研究成果，工业界的主要精力放在处理大访问量和大数据量上，对相关性排序没有突破。词频统计其实根本没有利用任何跟网络有关的特性，是前网络时代的技术。然而，网络时代的主要文献是以网页的形式存在的，而几乎每个人都可以随心所欲地在网上发表各种内容，词频相同的两个网页，质量相差可以很远，可是按照搜索引擎的第一定律，对这两个网页的排序应该是一样的。为了能够排在某些检索结果的前几位，许多网页内容制作者绞尽脑汁，在其页面上堆砌关键词，搜索引擎对此防不胜防，苦不堪言。这种情况到

了1996年开始有了改变。第二定律 人气质量定律 1996年4月，我到赌城拉斯维加斯开一个有关信息检索方面的学术会议，会议的内容就象拉斯维加斯的天气一样，照例比较枯燥乏味。但远离公司的我，却难得有一个静下心来认真思考问题的机会。就在听一个毫不相干的论文演讲的时候，我突然把科学引文索引的机制跟Web上的超级链接联系起来 - 感谢北大，她在我上大三的时候就教授了我科学引文索引的机制，美国恐怕没有一所大学会在你本科的时候教这玩艺儿。科学引文索引的机制，说白了就是谁的论文被引用次数多，谁就被认为是权威，论文就是好论文。这个思路移植到网上就是谁的网页被链接次数多，那个网页就被认为是质量高，人气旺。在加上相应的链接文字分析，就可以用在搜索结果的排序上了。这就引出了搜索引擎的第二定律：人气质量定律。根据这一定律，搜索结果的相关性排序，并不完全依赖于词频统计，而是更多地依赖于超链分析。我意识到这是一个突破性的东西，回去以后就很快总结了思路，于96年6月申请了这一方面的美国专利。1999年7月6号，美国专利和商标局批准了专利号为5,920,859的，以我为唯一发明人的专利。大约在96年底，斯坦福大学计算机系的两位研究生也想到了同样的解决方法，他们后来创立了一个叫Google的搜索引擎，Google的网站上至今仍然说他们的这项技术是Patent-pending (专利申请中)，不知道美国专利局是不是还会再批这样的专利。Anyway, 超链分析的方法98年以后逐渐被各大搜索引擎所接受，由于链接是网络内容的一个根本特性，这时候的搜索引擎才开始真正利用网络时代的检索技术。世事难料，2000年起网络泡沫迅速破灭，各大搜索引擎要么

遭人收购，要么推迟上市，所有使用人气质量定律的搜索引擎公司都未能幸免。那么，搜索引擎的出路到底在哪儿？三定律 自信心定律 人气质量定律解决的还是一个技术层面的问题，然而搜索引擎从诞生的那一天起，从来就不是一个纯技术现象，它融合了技术，文化，市场等各个层面的因素。解决搜索引擎公司的生存和发展问题需要搜索引擎的第三定律--自信心定律。1998年的时候，没有太多的人拿一家远在硅谷500英里以外，刚刚成立的，叫作GoTo.com（现已更名为Overture）的公司当回事儿。它不过是买了一个搜索引擎的技术服务，然后再向那些网站的拥有者们拍卖他们网站在GoTo检索结果中的排名，谁付的钱多，谁的网站就排在前面，而且付费是根据网民点击该网站的情况来计算的，仅仅在搜索结果中出现并不需要付费。这就是自信心定律的最早实践者！根据这一定律，搜索结果的相关性排序，除了词频统计和超链分析之外，更注重的是竞价拍卖。谁对自己的网站有信心，谁就排在前面。有信心的表现就是愿意为这个排名付钱。需要声明的是，自信心定律也是我自己给这一模式起的名字，以前的文献中并没有人这样总结过。今天，在网络业一片萧条，那纳斯达克风声鹤唳的时候，GoTo却如日中天，市值高达13亿美金，收入高达雅虎总收入的35%。反观门户网站，有哪一个能从它们的搜索引擎服务中赚出总收入的三分之一呢？究其原因，就是因为GoTo最早实践了搜索引擎的自信心定律。以前的搜索引擎都是靠CPM来收费的，而CPM是从传统广告业借鉴过来的，没有考虑网络媒体即时性，交互性，易竞价的特点，而竞价排名，点击收费则是为网站拥有者直接提供销售线索，而不是传统意义上的广告宣传。自

信心定律一改过去搜索引擎靠CPM收钱的尴尬局面，开创了真正属于互联网的收费模式。在中国，百度一直致力于推动搜索引擎市场的培植和发展，现在，我们联合了各大门户网站的搜索引擎，一起通过百度的竞价排名系统来实践搜索引擎的第三定律，为数十万网站的拥有者提供一个展示自信心，吸引潜在客户，发现销售线索的平台。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com