



# 面向搜索引擎的 技术接受模型研究

苏君华 / 著

search >



# 面向搜索引擎的 技术接受模型研究

苏君华 / 著

search >

江西人民出版社

图书在版编目(CIP)数据

面向搜索引擎的技术接受模型研究 / 苏君华著. —南昌:

江西人民出版社, 2013. 11

ISBN 978 - 7 - 210 - 06250 - 9

I. ①面… II. ①苏… III. ①互联网络 - 情报检索

IV. ①G354. 4

中国版本图书馆 CIP 数据核字(2013)第 258268 号



面向搜索引擎的技术接受模型研究

苏君华 著

江西人民出版社出版、发行

新华书店经销

地址:江西省南昌市三经路 47 号附 1 号(邮编:330006)

编辑部电话:0791 - 86898980

发行部电话:0791 - 86898815

网址:www.jxp-ph.com

E-mail:jxp-ph@tom.com web@jxp-ph.com

2013 年 12 月第 1 版 2013 年 12 月第 1 次印刷

开本:787 毫米 × 1092 毫米 1/16

印张:15

字数:235 千

ISBN 978 - 7 - 210 - 06250 - 9

赣版权登字—01—2013—418

版权所有 侵权必究

定价:28.00 元

承印厂:南昌市红星印刷有限公司

赣人版图书凡属印刷、装订错误,请随时向承印厂调换

本成果受到南昌大学社会科学学术著作出版基金资助

本成果是江西省高校人文社会科学研究项目TQ1204的最终成果

## 前 言

在 Web2.0 时代,搜索引擎俨然已经成了人们生活、工作中必不可少的大百科全书,越来越多的人开始依赖搜索引擎获取信息,从而带动着搜索引擎市场的竞争越来越激烈。多元化的用户需求、激烈的市场竞争推动着搜索引擎的发展。搜索引擎作为网络信息检索系统,用户对它的接受和使用即代表了用户接受了搜索引擎的技术及其创新。因此,可以采用技术接受模型来分析用户对搜索引擎的接受程度,解释影响用户接受的各种因素,为搜索引擎的进一步发展指明方向,为用户选择搜索引擎提供参考。

本研究在技术接受模型的基础上,通过文献回顾、调研、访谈等方式,总结了搜索引擎的特性并将其作为影响变量引入技术接受模型,从系统特征、个体特征、社会影响等方面构建了面向搜索引擎的技术接受模型,开发了搜索引擎用户接受的调查量表,并进行问卷调查。利用回收的 472 份有效样本数据,采用探索性因子分析和验证性因子分析方法验证了问卷的有效性和可靠性;采用结构方程和多元线性回归分析方法,分析了各潜变量之间的路径关系。研究发现面向搜索引擎的技术接受模型能够很好地预测用户的行为意图,解释了用户行为意图 79% 的变异性,具有一定的理论价值。

(1) 技术接受模型可以用于解释搜索引擎用户接受行为,主观规范、有用性认知是用户接受搜索引擎的重要解释变量。由于搜索引擎本身的简单易用,易用性认知对用户的行为意图没有显著影响。

(2) 用户对搜索引擎的任务关联、系统质量认知与有用性认知正相关,说明搜索引擎的系统质量以及能够帮助用户解决信息问题的功能显著影响用户对搜索引擎的认知,进而影响用户的行为意图。

(3) 搜索引擎的系统质量涉及搜索功能、界面、公正性认知、服务多样性、内容质量、排序质量、安全过滤和结果显示方式等方面,其中公正性是搜索引擎区别于一般学术信息检索系统所特有的,搜索服务提供商获取高额利益的同时还必须保证搜索引擎的公正性,良好的用户体验是搜索引擎获益的前提。

(4) 用户的信息素质认知与有用性认知正相关,信息技术通晓认知与易用性认知正相关。信息素质反映的是用户的信息能力,信息技术通晓反映的是用户的技术能力,这充分说明了个人能力在用户接受搜索引擎的过程中发挥着重要作用。

(5) 用户的个体差异是技术接受模型中的重要调节变量。在用户接受搜索引擎的过程中,易用性认知对有用性认知的影响受到经验的调节;主观规范对行为意图的影响受到性别、年龄、经验的调节;有用性认知对行为意图的影响受到性别的调节。

因此,各个搜索引擎要想在激烈的市场竞争中取胜,首要的是提高其系统质量,考虑用户未来的、变动的、个性化的需求,让用户有更好的使用体验;同时搜索引擎还应该提高用户的认知度,一方面是认知广度,再一方面就是认知深度,即用户对搜索引擎提供的各项功能的认识、理解与应用。另一方面,对用户而言,搜索引擎虽然简单易用,但要充分应用其提供的各项功能,用户还需要具备一定的信息素质和信息技术通晓水平,学会利用搜索引擎、学会分析信息需求、分析信息搜索结果。

## 目 录

前 言	1
第 1 章 绪论	1
1.1 问题的提出	1
1.2 搜索引擎概述	6
第 2 章 理论基础	10
2.1 技术接受模型及其扩展	10
2.1.1 技术接受模型(TAM)	10
2.1.2 扩展技术接受模型(TAM2)	12
2.1.3 技术接受与利用整合理论(UTAUT)	14
2.1.4 技术接受整合模型(TAM3)	16
2.1.5 TAM 的扩展	18
2.1.6 技术接受理论在搜索引擎领域的应用研究	20
2.2 信息素质与信息技术通晓	26
2.2.1 信息素质	26
2.2.2 信息技术通晓	29
2.3 搜索引擎评价及用户满意研究	33
2.3.1 搜索引擎评价研究	33
2.3.2 搜索引擎用户满意度研究	49
第 3 章 研究假设与研究方法	53
3.1 用户接受搜索引擎的影响因素分析	53

- 3.1.1 CNNIC 与正望咨询的调查分析 53
- 3.1.2 开放式问卷的内容分析 55
- 3.2 理论模型与研究假设 60
  - 3.2.1 理论模型 61
  - 3.2.2 变量定义及研究假设 62
- 3.3 问卷设计与调查 90
  - 3.3.1 问卷设计 90
  - 3.3.2 问卷调查与样本分析 100
- 3.4 数据分析方法 104
  - 3.4.1 信度分析 104
  - 3.4.2 效度分析 104
  - 3.4.3 结构方程模型 107

## 第4章 研究结果 111

- 4.1 信度与效度检验 111
  - 4.1.1 信度检验 111
  - 4.1.2 效度检验 116
- 4.2 结构方程模型的分析验证 132
  - 4.2.1 结构方程模型的数学表示 132
  - 4.2.2 结构方程模型的识别 134
  - 4.2.3 结构方程模型的估计 135
  - 4.2.4 结构方程模型的评价 138
  - 4.2.5 结构方程模型的修正 142
  - 4.2.6 实证结果解释 146
- 4.3 调节变量的效应分析 151
  - 4.3.1 性别的调节效应 152
  - 4.3.2 年龄的调节效应 155
  - 4.3.3 经验的调节效应 159



---

4.4	前置变量的回归分析	163
4.4.1	系统质量认知的回归分析	163
4.4.2	信息素质认知的回归分析	172
4.4.3	信息技术通晓认知的回归分析	176
<b>第5章</b>	<b>研究结论</b>	<b>180</b>
5.1	研究结论	180
5.1.1	结构方程模型分析	180
5.1.2	调节变量的效应分析	185
5.1.3	前置变量的回归模型分析	187
5.2	研究结论的启示	189
5.2.1	搜索引擎系统质量	189
5.2.2	提高搜索引擎的用户认知度	194
5.2.3	提高用户信息素质和信息技术通晓水平	197
	<b>参考文献</b>	<b>201</b>
	<b>附录 A 用户影响因素调查问卷</b>	<b>214</b>
	<b>附录 B 搜索引擎用户接受调查问卷</b>	<b>216</b>
	<b>附录 C 各修正模型的路径图</b>	<b>226</b>
	<b>后 记</b>	<b>230</b>

## 1.1 问题的提出

根据中国互联网络发展中心(CNNIC)2010年1月发布的《中国互联网络发展状况统计报告》,截至2009年底,我国IPv4地址已经达到2.3亿,数量仅次于美国,是全球第二大IPv4地址拥有国,年增长率达到28.2%;网络域名总数为1682万,其中80%为.CN域名;网站数量继续平稳增长,达到323万个,年增长率为12.3%。而自2003年开始,我国的网页规模基本保持每年翻番增长,2008年的网页数量为161亿个,2009年达到336亿个,年增长率超过100%,网页的规模反映了互联网的内容丰富程度。而截至2009年底,中国网民规模达到3.84亿人,较2008年增长28.9%,在总人口中的比重从22.6%提升到28.9%,互联网普及率在稳步上升。<sup>①</sup>

飞速发展的互联网络把全世界丰富的信息资源带到了人们的身边,用户所需要的大部分信息(工作、生活、学习、娱乐等等)在互联网上都可以找到,而且大部分是免费获取。而根据CNNIC统计报告中的个

<sup>①</sup> 第25次中国互联网络发展状况统计报告[EB/OL][2010-03-15]. <http://www.cnnic.net.cn>

人互联网应用指数<sup>①</sup>的构成情况,我们也可以发现我国网民的互联网应用目前还是以信息获取为主。互联网成为人们信息获取的重要来源。

网络信息资源在数量上具有海量化的特征,同时它种类繁多,除了文本信息,还包括大量的图像、视频、音频、软件、数据库等形式的信息。而这些存储在网络计算机上的各种形式的信息分布分散、开放,呈无序化的状态,给人们带来丰富信息的同时,也给人们获取信息带来了极大的困难。同时,迅速产生的大量信息中很多未经筛选,信息质量差,而且大量失效信息不能及时清理,成为垃圾信息,无形中又成为人们有效地获取、利用信息的障碍。网络信息资源相对于传统信息资源的“量”与“质”的巨大变化使得人们面对互联网时不会再有信息贫瘠的感觉,而是处于一种信息丰富但自身知识贫乏的尴尬局面。

为了解决人们在海量信息中搜寻和获取信息困难的问题,除了对网络信息组织方式、方法进行不断地优化外,人们一直在探索快速、有效地获取网络信息的检索工具。搜索引擎可以从各种网络资源中浏览和检索用户所需要的信息。随着网络科技和搜索技术的日新月异,搜索引擎已经成为影响人们生活的重要网络应用工具,并呈快速发展趋势。美国皮尤研究中心的“皮尤互联网与美国人生活项目”(Pew Internet & American Life Project)于2008年8月发布的一项研究报告显示,有近半数的网民(49%)几乎每天都使用搜索引擎<sup>②</sup>。据互联网流量监测机构 com-Score 公布的统计数据显示,2009年7月份全球用户搜索数已达1137亿次,同比增长41%<sup>③</sup>。截至2009年12月底,中国搜索引擎用户规模达2.8亿人,年增长7834万人,年增长率38.6%,搜索引擎在网民中的普及率达73.3%,搜索引擎已成为排名前三甲的网络应用之一。

根据2009年中国搜索引擎用户行为研究报告,2008年6月到2009年6月,搜索引擎用户年新增的5949万人中约有4340万人来自上网时间不到一年的新网民,

---

① “个人互联网应用指数”反映个人互联网应用水平,对应于互联网的四种应用,即信息获取类、休闲娱乐类、消费类、互动参与类,个人互联网应用指数包括四类二级指标:网络信息指数、网络娱乐指数、网络消费指数、互动参与指数。第25次中国互联网络发展状况统计报告[EB/OL][2010-03-15]. <http://www.cnnic.net.cn>

② Search Engine Watch. SEM News[EB/OL][2010-04-05]. <http://www.searchenginewatch.com>

③ 2009年中国搜索引擎用户行为研究报告[EB/OL][2010-03-15]. <http://www.cnnic.net.cn>

即新网民带来的新增搜索引擎用户人数约占搜索引擎年增长用户人数70%以上,成为搜索引擎用户增长的主要动力。<sup>①</sup> 这一系列的数字表明,在国内搜索引擎仍然存在着巨大的发展空间。

面对应用的快速增长以及未来广阔的市场空间,搜索引擎市场竞争更加激烈。2009年,中国搜索引擎首选市场形成了以百度为首,谷歌为第二梯队,搜搜、搜狗、雅虎为第三梯队的用户首选品牌格局。<sup>②</sup> 与此同时,不断有新的搜索引擎进入市场,微软推出必应搜索,腾讯放弃谷歌启用自己独立的搜索技术,等等。

飞速增长的网络信息、日趋多元化的用户信息需求和行为以及日益激烈的市场竞争,一方面推动了搜索引擎的多元化发展;另一方面也对搜索引擎提出了更新、更高的要求,促使人们更积极地展开对搜索引擎的研发。目前对于搜索引擎的研究,一方面集中于搜索引擎功能的增加、性能的提高;另一方面则集中于搜索引擎的评价。搜索引擎的评价将会有力地支持搜索引擎的发展,为用户提供一个了解、选择搜索引擎的依据。

目前,搜索引擎的评价研究主要是基于传统信息检索工具的评价标准。1973年,美国学者 Lancaster & Fayen 提出衡量信息检索系统的6项评价指标:覆盖范围、查全率、查准率、响应时间、用户负担和检索结果输出格式<sup>③</sup>。该指标体系从技术和用户需求方面给信息检索系统评价建立了一个基本框架。搜索引擎评价体系也基本按照这个脉络建立,主要从搜索引擎索引库、检索功能、检索效果、用户负担四个方面考虑,并根据网络信息资源的特征做了相应的调整。如 Leighton H. V. & Srivastava J. 提出了“相关性范畴”的概念,通过对不同范畴的检索结果赋予不同的权值计算查准率,并在此基础上提出“前X命中记录查准率” $P(x)$ 的概念,用来反映检索工具在前X个检索结果中向用户提供相关信息的能力。<sup>④</sup>

国外搜索引擎评价研究最主要的一个特点是强调“人性化”。在确定评价指标

① 2009年中国搜索引擎用户行为研究报告[EB/OL][2010-03-15]. <http://www.cnnic.net.cn>

② 2009年中国搜索引擎用户行为研究报告[EB/OL][2010-03-15]. <http://www.cnnic.net.cn>

③ 王静疆. 搜索引擎评价指标体系比较研究[J]. 图书情报工作, 2008(10): 136-138, 116

④ Leighton H. V. & Srivastava J.. First twenty precision among World Wide Web search services (search engines): AltaVista, Excite, Hotbot, Infoseek, Lycos[J]. *Journal of American Society for Information Science*, 1999, 50(10): 870-881

的过程中,充分强调检索效率重要性的同时,常常把用户负担放在第一位。但搜索引擎成功的标准是其用户量,技术的优势如果不能被用户认可,其价值将难以体现。<sup>①</sup>因此,随着用户服务意识的提高,关注用户负担已经远远不能满足用户使用搜索引擎的需求,还需要增加用户体验。<sup>②</sup>

信息技术自诞生以来就得到了企业界的青睐,很多企业不惜重金引进信息技术、研发信息系统。在投入的初始阶段,信息技术确实以其高附加值、高渗透力为企业带来了生产率的高速提升。但随着投入的进一步加大,信息技术对生产率的影响似乎并不像人们所预期的那么乐观,巨额的投入并没有带来预期的回报,导致了信息技术“生产率悖论”问题。而问题的原因就在于信息技术推广使用过程中遭到了用户的不接受甚至抵制。因此,要提高生产和工作效率,要发挥信息技术的价值,就不能只是简单地获取信息技术,还应该明白是什么原因导致用户不接受和抵制信息技术的使用,进而采取有效措施促进用户对信息技术的接受和使用。Doll & Torkzadeh 的描述很好地归纳了用户接受在信息系统研究中的重要地位:“用户接受是系统价值链上的重要变量,它将上游的有关系统成功原因的研究与下游的信息技术组织影响的研究连接起来。”<sup>③</sup>因此,理解用户的信息技术使用行为,探索影响用户接受信息技术的因素,预测和解释用户对信息技术的接受和使用,进而改变系统特性和实施过程来促进用户对系统的接受和使用具有重要的理论和实践价值。

已有的理性行为理论(Theory of Reasoned Action, TRA)、计划行为理论(Theory of Planned Behavior, TPB)模型具有很好地预测和解释人类社会行为的能力,具有广泛适用性,但我们需要的是特别针对解释人类对于信息技术的行为的理论。针对这个问题,学者们从不同的角度提出了各种观点,其中一部分学者从社会心理学、行为科学等角度提出,企业用户面对信息技术、信息系统并不是被动地全面接受,而是在环境、个体等各种因素的影响下,选择接受或拒绝新技术或新系统。通过研究用户对新技术的接受度和使用行为,学者们提出了众多的理论与模型,其中最

---

① 宋迪,吕英杰,李一军.基于用户偏好的搜索引擎指标分析与评价[J].图书情报工作,2007(1):122-125

② 王静疆.搜索引擎评价指标体系比较研究[J].图书情报工作,2008(10):136-138,116

③ Doll W. J. & Torkzadeh G. . Developing a multidimensional measure of system - use in an organizational context[J]. Information & Management,1998,33(4):171-185

代表性、应用最广泛的是技术接受模型( Technology Acceptance Model, TAM)<sup>①</sup>。TAM 专门用来描述和解释用户对信息技术的接受程度,并能够解释技术使用者的行为。

TAM 广泛应用于各种信息技术使用行为的预测和解释,近年来在互联网方面的应用特别多,研究对象包括电子商务、电子邮件、即时通信、在线游戏、在线学习、在线拍卖、电子政务、数字图书馆等等。如:Lederer 等验证了 TAM 在网络新闻组使用中的适用性,认为新闻组网站信息的质量、网站设计的易理解与易发现等外部变量通过有用性认知和易用性认知影响用户对网络新闻组的使用。搜索引擎是互联网上的导航工具,它通过采集、标引众多网络站点来提供全局性网络资源控制与检索机制,将全球互联网中所有的信息资源作一完整的集合、整理、分类,方便用户查找所需信息。<sup>②</sup> 搜索引擎是现代信息技术在信息检索领域的应用,是传统信息检索系统在互联网上的另一种再现(一种是网络数据库检索系统),当用户接受和使用搜索引擎进行网络信息检索时,就说明他们接受和使用搜索引擎采纳的技术创新活动。因此,可以采用技术接受模型来分析用户对搜索引擎的接受程度,解释影响用户接受的各种因素,为搜索引擎的进一步发展指明方向,为用户选择搜索引擎提供参考。

(1) 构建面向搜索引擎的技术接受模型,有利于丰富网络环境下信息检索的基本理论。信息检索理论的基本原理是用户需求表达与信息表述之间的匹配,这也是信息检索理论研究的核心问题,如果技术的先进性再结合用户的认知、情感,信息检索理论将更加完善。面向搜索引擎的技术接受模型中既包括了系统因素,也包括了用户个人认知、情感,为信息检索理论发展指明了方向。

(2) 构建面向搜索引擎的技术接受模型,有利于搜索引擎的发展。以技术接受理论为基础,结合信息素质与信息技术通晓理论,分析影响用户接受搜索引擎的各种因素并进行实证分析,为搜索引擎设计开发理念、技术的发展指明方向。

(3) 构建面向搜索引擎的技术接受模型,有利于搜索引擎的评价。搜索引擎是

---

① Davis F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology[J]. *MIS Quarterly*, 1989, 13 (3): 319 - 339

② 焦玉英. 信息检索进展[M]. 北京: 科学出版社, 2003: 152

网络环境下为用户提供的查找信息的工具,其性能的好坏、质量的优劣应该以用户的体验为基础,由用户进行评价。以用户接受理论为依据,从技术和用户体验的角度评价搜索引擎应该是应用性工具评价的基本出发点。

(4) 构建面向搜索引擎的技术接受模型,有利于用户选择适合的搜索引擎。互联网信息资源飞速增长,搜索引擎的数量也迅速增加,各搜索引擎的搜索范围大不相同,所使用的检索方式也日益复杂。以用户接受理论为基础,从技术和用户认知的角度同时评价搜索引擎,可以为用户选择合适的搜索引擎提供依据。

如何将技术接受研究的理论成果应用到搜索引擎领域,如何利用理论研究所揭示的重要的行为影响因素来引导用户接受搜索引擎,这是本研究所关注的问题。

## 1.2 搜索引擎概述

飞速发展的互联网将异常丰富的信息资源带到了每个人的身边,从来源上看,有政府的信息、科研机构的信息、大学的信息、企业的信息、个人的信息等等;从资源类型来看,有馆藏目录、电子书刊、参考工具书、数据库、电子邮件、电子公告等等;从内容上看,有政治信息、经济信息、科研信息、教育信息、生活信息、娱乐信息等等,用户在工作、生活、学习所需要的大部分信息在互联网上都可以找到,互联网成为人们获取信息的主要来源。这些信息资源数量巨大、内容广泛、类型多样、分散无序、动态变化。网络信息资源具有很多和传统信息资源不同的特点,从而导致对网络信息资源的加工处理、检索利用的难度显著增大。

搜索引擎在一定程度上解决了这个难题,它是人们在海量信息中搜寻和获取网络信息资源的重要工具,它根据一定的策略、运用特定的计算机程序搜集分布在网络上的信息,在对信息进行组织和处理后,为用户提供检索服务。搜索引擎对互联网的站点资源和其他网络资源进行标引,提供信息导航、信息检索服务的服务器或网站,主要由四部分构成,分别是信息采集模块、信息标引模块、信息检索模块和用户接口模块。其工作原理:首先,通过网络蜘蛛、机器人等信息采集工具或人工浏览的方式,根据网页链接,定期在互联网上采集相关的新网页,要尽可能多、快地采集各种类型的新信息;其次,标引模块对采集的网页中所包含的信息进行分析,提取出标引项,建立索引数据库;信息采集模块和信息标引模块共同完成建立、维护、

更新索引数据库的任务;信息检索模块则根据用户的检索提问式,提取出检索词并与索引数据库中的标引词匹配,检索出包含检索词的网页,并计算网页与检索提问式之间的相似度,进行相关性排序,然后输出结果;用户接口是用户输入检索提问式、显示查询结果并提供相关性反馈的界面。

搜索引擎的自动采集软件通常从已有的访问记录中提取初始的 URL,或从一些比较常用的主页开始,读取网页的内容,发现其中的链接,然后追随这些链接,从一个主页跳到另一个主页,寻找尚未访问过的文件;同时由于网页有时效性并且在不断地被更新,因此还要回访已经收集过的网页,确保搜集到的信息是最新的,而且尽可能地避免死链接和无效链接。信息采集通常有三种采集策略,分别是深度优先策略、广度优先策略、空间优先策略。

搜索引擎的这种信息采集策略成为影响采集结果的一个重要因素。初始的 URL 决定了链接遍历的方向,自动采集软件无法到达的网页、链接中断的网页以及未被链接的网页,都不能被搜索引擎所收录,这导致搜索引擎无法索引互联网上的所有信息。各搜索引擎由于选择的初始 URL 不同,且采取的采集策略不同,各搜索引擎所采集到的信息也不完全相同。由于网络信息的海量化,导致大量的网络信息湮没在网络的深处,这就是目前学术界所讨论的深网资源或网络隐性信息资源中的一类。同时,由于技术的原因,搜索引擎的采集软件只能发现 HTML 格式的页面,而对于非 ASCII 文本格式的内容则无法索引,如图像文件、视频文件、音频文件、可执行程序、压缩文件等,除非这些文件带有文本;互联网上存在着大量的网络数据库,数据库中所包含的信息内容都经过严格的质量控制,是应用价值较高的网络信息资源,但由于数据库信息都是结构化的信息,必须通过提问式检索才能获取相关的信息内容,而搜索引擎的信息采集软件无法理解数据库的信息结构,也无法实现信息检索,因而网络数据库所包含的信息内容也成为深网资源。因此,用户通过搜索引擎所能够获取的信息只是网络信息中极小的一部分,而且在这一小部分的信息中,其质量天差地别。

互联网是一个民主的世界,充分的民主导致网络信息充满了自由性、随意性和分散性,到目前为止网络信息依旧缺乏严格的管理和审查,缺乏有效的工具对信息进行选择和筛选。搜索引擎的自动采集软件在信息采集的过程中也没有严格的审



查与筛选机制,它能够采集到政府部分、教育部门、科研机构等发布的权威信息,也会采集到大量的垃圾信息、虚假信息、不良信息、违法信息、冗余信息等等,从而使网络信息的可靠性、精确性、权威性及其使用价值大打折扣。

搜索引擎的信息标引工作分为人工标引和自动标引,但由于网络信息资源数据量的巨大以及数据采集的高度自动化,当今的搜索引擎都采用了自动标引的方式。自动标引软件扫描各网页中的每一个词,依据词频、词位置以及超文本链接从网页中抽取能够表达网页主题的词作为标引词来建立网页标引记录,根据各个搜索引擎定义的相关算法计算每一个网页针对每一个标引词的相关度。由于汉语以字为基本的书写单位,而作为能够表达明确含义的最小语言成分却为词,并且汉语不同于拉丁语系语言中的词与词之间有空格为界,即汉语没有明显的区分标记,因此对汉语言的分词就显得格外重要。但汉语虚词众多,不可能像西文那样构造非用词表进行分词,而且,绝大多数汉字当与不同的汉字组词时,其词可能为关键词,也可能为非用词。因此,搜索引擎的自动标引软件对汉语信息标引更加困难。而且上述的这种字面标引方式往往不能很好地反映原文的中心内容,所选择的标引词也不一定能够得到众多用户的认可,尤其是复杂的汉语结构没有明确的词与词之间的界限,导致网页标引质量不高。除了搜索引擎标引技术方面的问题之外,搜索引擎优化的作弊行为也在一定程度上降低了标引的质量。搜索引擎优化本是针对搜索引擎对网页的检索特点,对网页中的相关元素进行优化,从而更能适应搜索引擎的检索规则,提高搜索引擎排名,使网页内容更容易地被相关用户获取,提高网站访问量。但部分网站却借助搜索引擎优化的名义作弊,在页面上堆放大量与页面主题不相关的误导性的关键词或相关的重复性的关键词,提高其在搜索引擎中的排名进而提高其被用户点击的几率。

搜索引擎一般对网页内容进行全文标引,在标引的基础上采用不同的方式对网页编制文摘。大多数搜索引擎直接摘取网页中的部分内容作为文摘,如 Lycos 的摘要包括网页标题、正文的第一句; Altavista 则利用网页内容的前 250 个字符作为文摘。搜索引擎的这种文摘编制方法并不能概括网页信息的中心内容,同时也使搜索引擎作弊更容易。

搜索引擎的信息检索软件是将用户输入的检索请求通过系统编码,与索引数