THE FUTURE OF MEDIA SERIES

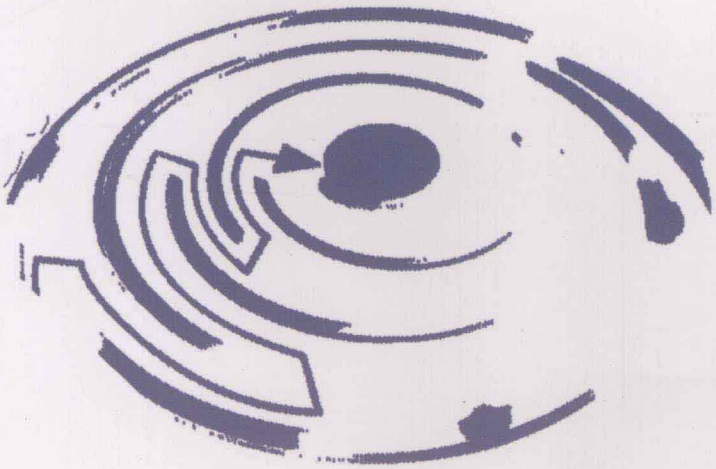
未来媒体丛书

高等院校新媒体系列教材

搜索

SEARCH ENGINE

梁晓涛 汪文斌 主编



WUHAN UNIVERSITY PRESS
武汉大学出版社

THE FUTURE OF MEDIA SERIES

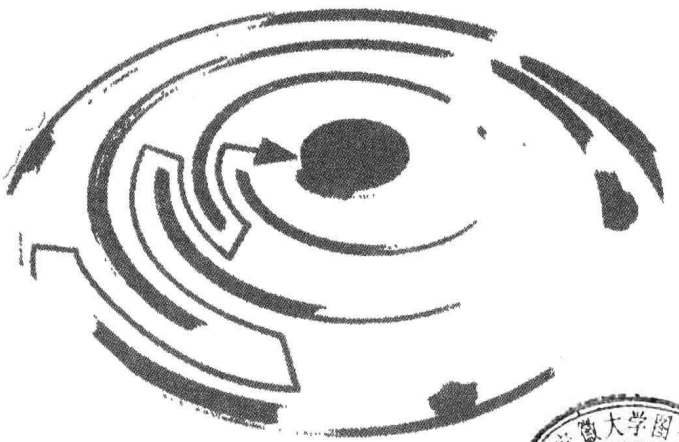
未来媒体丛书

高等院校新媒体系列教材

搜索

SEARCH ENGINE

梁晓涛　汪文斌　主编



WUHAN UNIVERSITY PRESS
武汉大学出版社

图书在版编目(CIP)数据

搜索/梁晓涛,汪文斌主编. —武汉:武汉大学出版社,2013.6
未来媒体丛书
高等院校新媒体系列教材
ISBN 978-7-307-11035-9

Ⅰ.搜… Ⅱ.①梁… ②汪… Ⅲ.互联网络—情报检索—高等学校—教材 Ⅳ.G354.4

中国版本图书馆 CIP 数据核字(2013)第 125401 号

责任编辑:张 欣 责任校对:黄添生 版式设计:马 佳

出版发行:**武汉大学出版社** (430072 武昌 珞珈山)
(电子邮件:cbs22@whu.edu.cn 网址:www.wdp.com.cn)
印刷:湖北新华印务有限公司
开本:787×1092 1/16 印张:16.25 字数:379 千字 插页:2
版次:2013 年 6 月第 1 版 2013 年 6 月第 1 次印刷
ISBN 978-7-307-11035-9 定价:38.00 元

THE FUTURE OF MEDIA SERIES

未 来 媒 体 丛 书

高 等 院 校 新 媒 体 系 列 教 材

丛书编委会

序言

网络新媒体的蓬勃发展和日新月异无疑是全球传媒界最引人瞩目的变革。广大受众以高度的热情欢迎技术创新所带来的各种全新体验。作为专业媒体从业者，透过现象探求其发展规律，科学地应对这场变革，把握传媒业的未来，是当前亟待破解的一个新课题。

这套《未来媒体》丛书试图对互联网上衍生的新业态进行一次全景式的扫描，选取其中具有代表性的移动互联网、社交网络服务、微博、搜索和网络视频等五个形态进行深度剖析和研究。

《移动互联网》对 iOS、安卓、Windows 等三大国际智能终端操作系统，在技术对比的基础上，通过典型案例对各操作系统上的应用商城、特别是移动应用进行了研究。

《社交网络服务》选取著名社交网站为对象，从业务、技术、应用、界面、安全、运营等方面进行了全面对比和分析。

《搜索》解剖了各具特色的搜索引擎服务商，提出专业化、个性化、智能化是搜索引擎未来的三个发展方向。

《网络视频》选取了国际主流视频网站和 CNTV、Youku 等国内视频网站为样本，从技术架构、业务特征、运营模式等角度进行了深入分析。

《微博》以国内、外主流微博平台为对象，从信息管理、技术特征、设计风格等方面进行了比较研究。

这套丛书基本涵盖了新兴媒体领域内具有典型性的运营机构及其应用和服务。其中既包括大型互联网公司，也包括仅有 10 名员工的微小企业；既有依靠技术创新，借助资本市场的力量迅速崛起新媒体服务机构，也有历史悠久、依托资金、人才和资源优势的传统媒体巨头。事实上，在由高科技推动的大众媒介日益走向融合的今天，传统媒体与新媒体的边界已难以区分，因此丛书有意回避了

“新媒体”这一概念，统一称之为“未来媒体”。

系统性地针对新兴媒体进行全面分析和研究是一件很有意义的工作。它有助于我们把握媒体的未来脉搏。近年来，中国网络电视台在网络视频、IP 电视、互联网电视、移动电视等方面进行了诸多探索，也积累了一些经验，我希望和广大媒体从业者以及关注新媒体发展的读者共同分享这些体会与认识。

鉴于网络媒体是一个日新月异、飞速发展的领域，过去的经验不断受到挑战，既有的规律也不断地被打破，这方面的研究工作还待继续，中央电视台也将与大家一起继续探索和思考。

2012. 12

目 录
CONTENTS

THE FUTURE OF MEDIA SERIES

1 搜索概述

1.1 搜索相关概念

随着计算机和互联网技术的飞速发展，网络上的信息量急剧增长。网络已经成为人类有史以来资源数量最多、资源种类最全、资源规模最大的一个综合信息库。其信息来源丰富、分布广泛，各种类型的信息资源异构地分布在网络空间中，但如果不能使庞杂的信息有序化，就很难有效获取。如何准确有效地从互联网上获取信息就成为一项艰巨的任务，目前解决这一问题的最佳方案是利用搜索引擎。搜索引擎正是为了解决“信息丰富，知识贫乏”这一奇怪现象的问题而出现的技术。

搜索引擎是一个信息处理系统，以一定的策略在互联网中搜集、发现信息，对信息进行理解、提取、组织和处理，并为用户提供检索服务，从而起到信息导航的作用。搜索引擎一般包括信息搜集、信息整理和用户查询三部分。从用户的角度来看，它是一个帮助人们进行信息检索的工具。搜索引擎已经成为信息领域的产业之一。它要用到信息检索、人工智能、数据库、数据挖掘、自然语言理解等领域的理论和技术，具有综合性和挑战性。又由于搜索引擎有大量的用户，由此衍生出许多商机，具有很好的经济价值。

本书以搜索产品为主要内容，介绍搜索产业的整体状况以及搜索平台与现代传媒的关系。图1-1所描述的是目前广为所用的商业搜索引擎产品图。尽管图1-1不能涵盖业界所

图 1-1　搜索引擎产品图

使用的所有的搜索产品，但从此图可以看出，搜索产品已经发展成为 IT 领域不可或缺的核心产品。

1.2 搜索与媒体以及搜索平台对传媒发展的意义

诞生于 20 世纪 90 年代、起源于为信息检索提供服务的 IT 技术应用的搜索引擎，至今不过二十年的发展历史，已发展成为在数字网络时代大背景下构建全球信息化社会的强大工具；同时随着相关技术的不断深入开发与利用，搜索引擎所蕴藏的巨大商业价值逐渐被挖掘，也极大成就了专业运作搜索引擎的企业组织，典型代表如全球最大的搜索引擎 Google、全球最大的中文搜索引擎百度等，在整个世界经济、政治、文化发展中越来越发挥着举足轻重的作用。

“搜索引擎”已不再是原有的仅指某项信息技术的专有名词，而是被凸显出更多政治、经济、社会、文化等价值的复合概念，它正成为改变着整个人类社会未来生存形态的推动力量。“搜索引擎已经不仅仅是一项网络技术和一个提供信息的普通平台，它已发展成为一种新型的、有影响力的媒介，能够控制信息流动，起到舆论导向作用，直接影响人们认知世界的方式。

正因如此，现实中发展的搜索引擎，在成为推动构建人类信息化生存形态力量的同时，也发挥出超越其本性的功能，即媒体的功能，有评论言：“搜索引擎可能正由技术权力的合理追求转向经济权力的贪婪追求，继而转向社会控制力越界追求。”

1.2.1 搜索引擎由技术工具演化成为一种新的网络媒体

搜索引擎源起一个网络信息技术应用，从功能实现上讲它可以根据一定的策略、运用特定的计算机程序从互联网上搜集信息，在对信息进行组织和处理后，为用户提供检索服务。从搜索引擎的发展演进中看，互联网为搜索引擎插上了放飞的翅膀，搜索引擎基于互联网迅速发展成为一种交互式信息平台。互联网一直本着不受任何控制的精神在发展，而且在其发展初期，互联网被视为“公共传输模式”的媒介，在这样的数字化网络环境下，搜索引擎作为一个互联网信息技术应用工具的意义在于，出于信息方面的原因而寻求特定的媒体内容，它使用户接触更多的信息内容，并将这些内容视为真实的。工具性的使用是积极的、有目的的，它意味着媒体使用的功利性、目的性、选择性和参与度；工具性的倾向导致态度和行为方面给产生更强的效果，它包含着对资讯和信息的更为强烈的使用和参与动机，这里的参与意味着一种乐于对资讯和信息加以选择、诠释和回应的状态。搜索引擎由技术工具演化成为一种新的网络媒体。

麦克卢汉的传播技术决定论认为“媒介是人体的延伸”。技术的任何进步，都会使人类更有效的生活和劳动，媒体的任何发展，都能延伸人类五官的功能，所有媒体均可以同人体器官发生某种联系。所有的这些媒体，都影响并改变着人们的生活方式、工作方式和思维方式。人体的任何延伸，都会对社会的发展造成影响，促成变革，使一部分人行进在

时代前列，而使另一部分人落伍。

在如今盛产内容的时代，人们通过搜索引擎这一网络媒体，从海量信息中搜寻并甄别出真正所需的内容，从而满足节约信息成本的本能需求，并缓解过量信息带来的信息焦虑。根据媒介效果研究中的“使用与依赖模式”所描述的个人的传播需求与传播动机、信息寻求策略、媒介使用情况、功能性替代品和媒介依赖之间的关系，如果某些需求和动机限制了人们对信息寻求策略的选择，它们就有可能导致人们对某些传播渠道的依赖，相对地，人们对特定传播渠道的依赖也会产生其他一些效果，如态度和行为的改变等，反作用于其他的社会关系，使之发生变化。搜索引擎技术的功能特点使其媒介功能性替代品极其有限，搜索引擎成为受众获得有意图、有目的和有动机的选择内容满足的唯一策略。另外，德瑞德在媒介经济研究中提出的集中的三个区分层次中的关于受众集中，指的是受众市场份额集中的境况，在搜索引擎发展历程中所反映的按照搜索引擎用户比例数定义的市场份额数据证实了这一点。以上两点说明受众对搜索引擎的媒介依赖，预示着搜索引擎作为媒体所具有的巨大影响力。

搜索引擎已不再是专有技术的代名词，时代赋予它新的媒体内涵，正如以上所述人们对特定传播渠道的依赖也会产生其他一些效果，这反映的就是搜索引擎的媒体化趋势的结果。

1.2.2 搜索引擎对媒体业发展的意义

理解搜索引擎对媒体发展的意义，需要界定“媒体”的概念。在很多情况下将“媒体”和“媒介”两者的概念是混用的，即彼此意义相同或指代一致。但笔者在本文中，将媒介的概念更多表达的是一种宽泛的传播学概念，正如麦克卢汉所说，媒介即讯息；而媒体的概念，则更多意指媒体组织。丹尼斯·麦奎尔《大众传播理论》中说，将媒体组织同时看做“社会机构”和“产业”来进行阐述，并且指出“媒体（组织）不同于其他任何企业组织形式”。由此可以看出“媒体”作为一种特殊的组织形式所具有的特征属性：既是市场运作下的经济实体组织，又是致力于保持“中立”承担公共责任的社会机构组织。这两个相互矛盾的属性也造就了媒体组织内在的两难冲突：赢利和社会目的之间的潜在冲突的问题。

搜索引擎作为网络媒体对媒体的发展的意义如何体现出来？一般人们都将搜索引擎视为一种新媒体进行看待。相比其他传统媒体自然性的垄断发展，搜索引擎是通过完全的市场竞争发展起来的，而且市场垄断竞争格局也是严格根据市场规律而形成。所以基于信息技术发展起来的搜索引擎对整个媒体产业经济发展的意义是显而易见的。

在中国，网络媒体被受众预设为依靠海量性、自由性、客观性而获得公正名声的媒体平台。而搜索引擎作为一种纯技术色彩的网络服务工具，其科学和规范性更是被寄予厚望，搜索引擎的公信力日渐增强。有机构曾做过调查，搜索引擎的相关性和可信度都大大超过了传统媒体。而且网络搜索也被称为“可信赖的、使用度很高的信息来源”。几乎所有的研究都表明，受众对搜索的依赖已经远远高于对任何一个门户网站的依赖。这其中，除了用户的深度需求之外，搜索引擎基于科学和规范的客观性声望，也起了最大的作用，

所谓“数据不会骗人”，其召唤力可想而知。于是，搜索作为媒体似乎是要并且确实能够担当起一种社会的公器的角色了。正是因为搜索引擎公信力的日渐增强，主流媒体的注意力经济效应逐渐在搜索引擎领域显现出来。

技术的优势推动市场的垄断，根据皮卡特在 Media Economics 中所述，通常可接受限度的市场集中门槛法则是：排名前四位的公司对某产业控制超过 50%，或前八家公司对某产业控制超过 70%，就被认为是高度集中。基于搜索引擎发展历程中的数据，无论是全球搜索还是中文搜索的市场结构都是高度集中：全球搜索市场 Google 一直保持在 60% 以上的市场份额，而其他的搜索引擎均在 10% 以下；中文搜索市场百度和 Google 一直占据 80% 左右的市场份额。所以，可以说 Google 和百度的发展趋势引领着整个搜索引擎的发展趋势，搜索引擎的媒体化趋势分析，以百度和 Google 的相关数据为准具有合理性。

当搜索引擎对信息的控制力足够大时，其媒体特性逐渐显现，相应所须承担的媒体责任亟待明确，正如麦克奎尔所说：媒体不同于其他企业，要挑起公共责任的重担，无论它们喜不喜欢。一旦搜索引擎严重损害了信息甄选机制的独立性和公正性，也损害了搜索引擎这一新型传媒的公信力。这也是搜索引擎媒体化趋势分析的意义所在。

1.3 搜索的发展历程

1.3.1 搜索的产生

在 1990 年之前，一般情况下人们进行信息检索是到图书馆查阅大量书籍来获取自身需要的信息。

1990 年，蒙特利尔的麦吉尔大学发明了 Archie。Archie 是第一个自动索引互联网上匿名 FTP 网站文件的程序，但它还不是真正的搜索引擎。Archie 可以帮助用户在互联网的任意一个匿名 FTP 服务器上查找文章和目录。

1994 年，第一个既可搜索又可浏览的分类目录 EINet Galaxy 上线，除了网站搜索，它还支持 Gopher 和 Telnet 搜索。同年，美籍华人杨致远等创办了 Yahoo，随着访问量和收录链接数的增长，Yahoo 目录开始支持简单的数据库搜索。在这一年中，华盛顿大学开始了 WebCrawler 项目的研究。WebCrawler 是互联网上第一个支持搜索文件内全部文字的搜索引擎，在它之前，用户只能通过 URL 和摘要搜索（摘要一般来自人工评论或程序自动摘取的正文前 100 个字）。随后的 Infoseek 是另一个重要的搜索引擎，它沿袭 Yahoo 的概念，并没有什么独特的革新。1995 年，它与 Netscape 的战略性合作，使它成为一个强势搜索引擎。

1995 年，一种新的搜索引擎形式出现了——多元搜索引擎。用户只需提交一次搜索请求，由多元搜索引擎负责转换处理后提交给多个预先选定的独立搜索引擎，并将从各独立搜索引擎返回的所有查询结果集中起来，处理后再返回给用户。第一个多元搜索引擎是华盛顿大学硕士生开发出的 Metacrawler。

1.3.2 搜索的发展

1990 年以前，没有任何人能搜索互联网。1990 年，加拿大麦吉尔大学计算机学院的师生开发出 Archie。当时，万维网还没有出现，人们通过 FTP 来共享交流资源。Archie 能定期搜集并分析 FTP 服务器上的文件名信息，提供查找分布在各个 FTP 主机中的文件。用户必须输入精确的文件名进行搜索。Archie 告诉用户在哪个 FTP 服务器能下载该文件。虽然 Archie 搜集的信息资源不是网页（HTML 文件），但和搜索引擎的基本工作方式是一样的，自动搜集信息资源，建立索引，提供检索服务。所以，Archie 被公认为现代搜索引擎的鼻祖。Robot（机器人）一词对编程者有特殊的意义。Computer Robot 是指某个能以人类无法达到的速度不断重复执行某项任务的自动程序。由于专门用于检索信息的 Robot 程序像蜘蛛一样在网络间爬来爬去，因此，搜索引擎的 Robot 程序被称为 Spider 程序。

1993 年，Matthew Gray 开发出 World Wide Web Wanderer，这是第一个利用 HTML 网页之间的链接关系来检测万维网规模的机器人程序。开始，它仅仅用来统计互联网上的服务器数量，后来也能够捕获网址。

1994 年 4 月，斯坦福大学的两名博士生，美籍华人杨致远和 David Filo 共同创办了 Yahoo。随着访问量和收录链接数的增长，Yahoo 目录开始支持简单的数据库搜索。Yahoo 的数据是手工输入的，所以不能真正被归为搜索引擎。事实上它只是一个可搜索的目录。雅虎于 2002 年 12 月 23 日收购 Inktomi，2003 年 7 月 14 日收购包括 Fast 和 AltaVista 在内的 Over-ture，2003 年 11 月，Yahoo 全资收购 3721 公司。1994 年 7 月，卡内基·梅隆大学的 Michael Mauldin 将 John Leavitt 的 Spider 程序接入到其索引程序中，创建了 Lycos。除了相关性排序外，Lycos 还提供了前缀匹配和字符相近限制，Lycos 第一个在搜索结果中使用了网页自动摘要，而最大的优势还是它远胜过其他搜索引擎的数据量。

1995 年，一种新的搜索引擎形式出现了——元搜索引擎（A Meta Search Engine Roundup）。用户只需提交一次搜索请求，由元搜索引擎负责转换处理，提交给多个预先选定的独立搜索引擎，并将从各独立搜索引擎返回的所有查询结果，集中起来处理后再返回给用户。第一个元搜索引擎是 Washington 大学硕士生 Eric Selberg 和 Oren Etzioni 开发的 Metacrawler。1995 年 12 月，DEC 正式发布 AltaVista。AltaVista 是第一个支持自然语言搜索的搜索引擎，也是第一个实现高级搜索语法的搜索引擎（如 AND，OR，NOT 等）。用户可以用 AltaVista 搜索新闻组的内容并从互联网上获得文章，还可以搜索图片名称中的文字，搜索 Titles，搜索 Java applets，搜索 ActiveX objects。AltaVista 也声称是第一个支持用户自己向网页索引库提交或删除 URL 的搜索引擎，并能在 24 小时内上线。AltaVista 最有趣的新功能之一是搜索链接指向某个 URL 的所有网站，同时在面向用户的界面上，AltaVista 也作了大量革新，它在搜索区域放了“tips”以帮助用户更好地表达搜索模式。这些小 tip 经常更新。这样，在搜索过几次以后，用户会看到很多他们可能从来不知道的有趣功能。这些系列功能，逐渐被其他搜索引擎广泛采用。1997 年，AltaVista 发布了一个图形演示系统 Live Topics，帮助用户从成千上万的搜索结果中找到想要的。

1995 年 9 月 26 日，加州伯克利分校助教 EricBrewer，博士生 Paul Gauthier 创立了