



Internet 信息检索系列

武汉大学信息管理学院  
武汉大学信息资源研究中心

审定并推荐

# Internet 通用搜索引擎 检索指南

R·霍克 著  
金丽华 译

第二版

 辽宁科学技术出版社  
LIAONING SCIENCE AND TECHNOLOGY PUBLISHING HOUSE

武汉大学信息管理学院 审定并推荐  
武汉大学信息资源研究中心

Internet 信息检索系列

# Internet 通用搜索引擎 检索指南

第二版

R·霍克 著 金丽华 译

辽宁科学技术出版社

***The Extreme Searcher's Guide to Web Search Engines :  
A Handbook for the Serious Searcher. Second Edition***

Copyright © 2001 by Randolph E. Hock.

All rights reserved. No part of this book may be reproduced in any form or by any electronic or mechanical means including information storage and retrieval systems without permission in writing from the publisher, except by a reviewer, who may quote brief passages in a review. Published by CyberAge Books, an imprint of Information Today, Inc., 143 Old Marlton Pike, Medford, New Jersey 08055.

本书中文版由 Information Today Inc. 授权  
著作权合同登记号为：06-2002 年第 097 号  
© 2003 中文版权归辽宁科学技术出版社所有

**图书在版编目 (CIP) 数据**

Internet 通用搜索引擎检索指南 (第二版) / (美) R·霍克著; 金丽华译. - 沈阳: 辽宁科学技术出版社, 2003.2

ISBN 7-5381-3849-8

I . I... II . ①R... ②金... III . 因特网 - 情报检索 - 指南 IV . ①G252.7②TP393.4

中国版本图书馆 CIP 数据核字 (2002) 第 096823 号

---

出 版 者: 辽宁科学技术出版社

(地址: 沈阳市和平区十一纬路 25 号 邮编: 110003)

印 刷 者: 沈阳新华印刷厂

发 行 者: 各地新华书店

开 本: 787mm × 1092mm 1/18

字 数: 200 千字

印 张: 13

印 数: 1 ~ 5 000

出版时间: 2003 年 2 月第 1 版

印刷时间: 2003 年 2 月第 1 次印刷

责任编辑: 杨廷郊

封面设计: 庄庆芳

版式设计: 于 浪

责任校对: 刘 庶

---

定 价: 28.00 元

联系电话: 024-23284360

邮购咨询电话: 024-23284502

E-mail: lkzsb@mail.lnpgc.com.cn

http: // www.lnkj.com.cn

# 序

## FOREWORD

专业研究人员对网络搜索引擎都有一定的看法。凭直觉我们就知道它们的局限性：命中的次数过多，检准率不足，对检索过程和输出缺乏控制。从我们将 Internet 放进我们的检索工具箱的那一刻开始，就取而代之，不再使用主题目录、指示丰富的元网站，以及其他由人参与管理的检索工具了，我们寄希望于能够改善检全率、检准率和提高检索结果的质量。我们认为，搜索引擎的使用对象是那些对问题了解不深入的人。它们主要被用在查询大海捞针那样的问题，查询稀奇古怪的用语，以及查询用别的方法找不到的信息上。但我们反对——现在仍然反对——它们关于标引“全部”或大多数网页的说法。我们不喜欢电子商务的诱惑，反对由它构建我们需要的简单的检索方式。我们也不相信享有专有权的检索程序“黑匣子”，反对由它们来决定我们应该使用或不使用哪些文献。

我们正在遭受掠夺。但并不是说在网络出现之前的联机检索年代已长大成人的那代人，他们过得不好。那时，我们总抱怨当时的标引质量不好；数据被漏检、词序被改变或者被伪造；抱怨覆盖面有缺漏，内容受到许可的限制，技术上有问题，以及操作员有误。但当时我们却是检索工具的主人。我们使用了靠词检索运算符，中间截词和后截词，检索集，及几十种系统专用的命令。我们学会了怎样来识别损害和怎样绕过它们的方法。我们使用了布尔逻辑式来检索数据库，无论它们有什么不足，检出的信息确实是我们需要的。

说我们检索的商业数据库仍不完善，恰恰说明已经有了评价它的标准。而网络则没有这种标准。不同形式的相关性排序，更适于网络的良莠不齐、有时甚至是稀奇古怪的本性。但是它的运作却是不透明的，而且，不透明也不适于高级检索人员。我们对检索结果被命中或漏检、来路不明、常常令人生疑表示不满。我们想知道为什么能够得到想要得到的。对上网偶尔为之的人会找到答案和在网上来去自由而兴奋不已；而对于高级检索人员，则必须知道结果是有效的或者至少是有代表性的，才能令他们满意。他们想知道结果出现了错误是怎么造成的，是还没有掌握检索方法呢，抑或是受



到他们所使用的检索工具没有收录文献的局限呢。

在我第一次遇到 **Ran Hock** 时,他是一名先进的商业联机服务方面的培训顾问,在商业、学术研究,自然科学和技术方面享有很高的威望。像我们中的许多人一样,他掌握了布尔逻辑的要素,知道如何使用靠词检索运算符,范畴修饰语和截词符号确保得到最相关的结果。还知道使用鲜为人知的系统手段,保证检索成功和有效。

**Hock** 撰写本书的目的,不仅是出于专业精深,而且,对搜索引擎发展目标提出了殷切的希望。他从一般到系统,一个搜索引擎一个搜索引擎地把我们带进现代搜索引擎世界。比较表、通用的检索方法和大量的例子构成了第一章和第二章的内容,是熟习检索网络的独立部分。

**Hock** 采用严谨的检索手段进而考察了每种搜索引擎。它们各自的优点和缺点是什么?网站覆盖的规模有多大?其缺省方式是怎样工作的——也就是黑匣子里面装了些什么?怎样使用功能强大的检索方法,例如布尔逻辑检索、词组检索、截词检索和范畴检索?怎样使用标点符号和大写的专有名词?输出选择是什么?还能够提供什么——可直接检索主页数据库,如地图、股票行情或软件库吗?缩小检索结果的范围有哪些方法?主页上有哪些入口?网站的个性化功能怎样?他还分析了在网络上获取信息的另外一种方式——主题目录的结构、内容和起源,它已成为今天多数搜索引擎的一个组成部分。每章中都给出了网页的实例,网页上翻和下翻能力,最后还都有方便而精练的结束语。**Hock** 从网络和桌面两个方面介绍了元搜索引擎及其特点。在讲述当中给出了有用的检索技巧,以及自己评价搜索引擎的方法。

是不是把“顶级”这个词用在搜索引擎检索技术上有点太极端了呢?完全不是的。“顶级检索人员”,我称为“超级检索人员”,这两种称呼是相似的——可能毫无二致。它是信息专业中的专门术语。顶级检索人员就像体育迷一样,他们无所畏惧、精力投入,富有信仰和组织纪律性。在实际上,顶级检索人员还的确把检索看成是一种体育活动;他们同样要冲刺、同样在有最后有成功的喜悦——或者“不行就再来一次”。追求顶级的本意是,要知道怎样吃干榨净设备,让设备成为滑雪板,成为冲浪板,成为山地车——或搜索引擎。**Ran Hock** 的这本书将帮助你登上制高点。

**Reva Basch**

谨以本书献给 Pamela, Matthew, Stephen 和 Elizabeth

## 致 谢

## ACKNOWLEDGMENTS

如同本书的第一版，衷心感谢 Information Today 公司的品格高尚的工作人员，感谢他们的大力支持、鼓励、卓绝的工作和耐心。特别感谢 John Bryans 在处理“大图片”上不懈的努力；感谢 John Eichorn 的编辑技能和耐心；感谢 Janet Spavlik 热情解决版式等相关问题。还感谢策划编辑 Deborah Poulson，生产经理 Heide Dengler 以及设计师 Jacqueline Walter 和 Kara Mia Jalkowski。特别感谢在成功地销售第一版中起到重大作用的大量销售人员。更要再次感谢 Information Today 公司的老板 Tom Hogan 先生，我非常佩服他对联机工业发挥的作用，再次感谢他对本书的支持。

对新英格兰联机用户集团 (NENON) 的我的朋友们表示谢忱，是他们建议我使用“顶级检索人员”这个术语。我还要感谢世界上的顶级检索人员，是他们希望和热衷于将搜索引擎用到“顶级”，尽可能地充分利用网络中的惊人的研究资源。

# 前言

## INTRODUCTION

如果你正好处于从未成年到成年的成长时期，那么，会有助于你理解在本书的第一版在 1999 年出版后搜索引擎的发展情况。搜索引擎正处在成长中的青少年后期。个头在长，有了成熟的迹象，并向自己提问“我是谁（是做什么的）？”他们好奇心强烈，仍然备受成长的煎熬，偶尔还会得皮肤病。排除微小的细节不算，最大的变化是长大了，成熟了。到 2001 年为止，大多数的搜索引擎在规模上都翻了两倍或三倍。成熟的过程也甚为明显。每个搜索引擎都一直观察着别人，而在界面和检索方法上采用了更好的方式。有些搜索引擎正在验明正身，但尚不能肯定是否能成为主要的入口或搜索引擎。未成年时期的好奇心仍有增无减。（如果一个搜索引擎屏幕上采用了彩色标记，另一个准会说“我们也用吧。”）

好啦！让我们从类比中走出来，走到突破点，还是来说搜索引擎吧。还会有基本的比喻，但搜索引擎已经长大和成熟了。

在介绍搜索引擎时，本书不是着重讨论已经发生的变化，而是把重点放在什么是上。如果你以前就熟悉搜索引擎，你会明白这种区别的。如果你对搜索引擎的详细情况是个新手，就更无所谓以前它们是什么样子了。本书的目的，是帮助你了解搜索引擎必须为你，特别是为“顶级检索人员”提供哪些东西。

本书的书名叫做《网络搜索引擎顶级检索人员指南：顶级检索人员手册》，希望这个书名能够反映本书的概念、目的和内容。

“顶级检索人员”这个词含有这样的意思：有不少的检索人员对最小公分母的方法和结果不满意。有很多人需要延伸这个极限。要列名于顶级检索人员的行列，并且不需要特殊的技能或培训。但却要求人们了解有关的知识和设备。

“网络搜索引擎”这个术语有几种定义。我们认为，搜索引擎乃是由万维网提供的一种服务，用户可以利用它输入提问词，并检索包含网站上绝大部分内容的数据库。更为特殊地说，用户可以在搜索引擎上输入一个或多个词，还可以加上修饰语，以查找感兴趣的网页。该词可以与“网络检索服务”相互替代，在本

书中,后者更多情况是特指整个网站,并从多种选择中选中其一依次提供给搜索引擎。搜索引擎不过是诸多提供物中的一种,目的是向用户提供一个进入网站的起始点,或者入口。

书名中的“指南”和“手册”,进一步规定了本书的内容。本书重在条理性和可读性,取材对有效的检索工作都是必要的。本书不想包罗万象,没有谈及网络搜索引擎相关的理论问题,重点放在实用上——这些知识对于快捷地找到答案既实用,又必不可少。

至于书名中的“顶级检索人员”,因对“顶级检索人员”的不同看法而异,主要是为了强调目的。使用这个词语有助于定义可能的用户,特指经常利用网络,用于商业、科研和其他研究目的的那批用户。有那么一些人需要服务有较强的检索功能,这种功能要高出使用搜索引擎服务提供的简单模式和方法能获得的功能。但这类服务中的大多数,主要瞄准曾经被 Infoseek 称作“水平平平的非技术方面的消费者”这批人群。尽管这些用户也可以极为有效地利用最简单的方法进行检索,但是,只有多做出努力,多学习知识,才能具备较强的检索能力。

更为特殊地说,本书的目的是为了更好地利用网络搜索引擎,帮助读者:

- 了解主要搜索引擎的基本结构
- 熟悉搜索引擎通用的属性(方法,益处、选择、内容等方面)及区别
- 了解主要搜索引擎的主要优缺点
- 掌握检索技巧与方法,提高检索的效率和便利

这种目的在第一章“搜索引擎概述”和第二章“常用的检索方法”中做了概述。第一章简单介绍了网络搜索引擎的发展历史,较为详细地介绍了它们的工作原理。第二章介绍了不同搜索引擎的检索方法和选择。

第三章到第十章专论了8个先进的搜索引擎的情况。之所以选择这8个搜索引擎,主要基于搜索引擎的规模、主导位置和检索功能方面的考虑。其中包括了 Yahoo!,尽管它从结构上与其说是搜索引擎倒不如说是网站目录。它将 Google 的规模庞大的数据库整合,所以将其划入搜索引擎行列也是说得过去的。在上述各章中,更多地强调了服务的“检索”特性,而把大量的,形式相当不同的入口放在了次要位置,这是因为它们常常与检索功能没有密切的关系。在许多搜索引擎上都备有网站目录,比其他各类入口给予了更多的关



照,这主要是因为目录内容被自动并入检索结果的程度越来越明显的缘故。专论各章在编排上做了充分的考虑,以使检索人员能够容易地掌握那个搜索引擎必须提供什么,怎么样来使用它。

在本书的第一版中,曾经包括了 WebCrawler 的内容,主要来自历史角度它是第一个被广泛使用的网络搜索引擎的考虑。因为目前它的内容微弱,不能对高级检索人员提供有用的东西,所以在再版时把它去掉了。同样还去掉了 Go. com。在 2000 年年中,在提出其新版时,其生产商曾经宣布过,新版的 Go. com 主要放在娱乐和休闲方面,将其置于本书范围之外了。新的版本还取消了被认为是 7 个“强项”中的 4 项。此外,它的数据库的规模也不足 1 亿项记录。简直不能与本书收录的搜索引擎同日而语。

本书再版中还增加了两个搜索引擎领域中的新生力量 Google 和 Fast Search。是它们的规模、惟一性和检索功能把它们放进“必须知道”的范畴的。此外,我们还认真考虑了另外两个新的搜索引擎 MSN Search 和 NBCi. com(以前叫 SNAP)。然而,它们对高级检索人员的思想库并没有提供特殊或重大的贡献,因而也决定不收录它们。

第十一章评述并用例子说明了许多元搜索引擎,探讨了它们必须提供什么,介绍了在它们没有提供这些东西时须注意的问题。

第十二章介绍了本文没有谈到的搜索引擎,在哪里能找到它们。还论述了“跟踪”的问题,并提供了跟踪时建议使用的资源。

有关跟踪的问题,作者提供了一个网页,通过这个网页可以了解不可避免发生的变化,这种变化关系到本书的内容(见“前言”后面的“关于顶级检索人员网页”)。

从本书的第一版以来,搜索引擎不断完善,消除了某些缺憾。但这并不是说,已经找不到什么纰漏。如果事情并不像你想象的那么好,并不是因为它们把事情弄糟了。原因可能是搜索引擎没有达到其生产商的许诺目标,它们也不应该总是按经验丰富的高级检索人员的要求去做事情——这方面的问题,在有关它们的文件中都做出了说明,只有面对这种现实,才能成为一个健康、快乐的检索人员。但绝大多数的许诺的确都可以兑现,能够利用它们在万维网空间中迅速、方便地查询成千上万的有用网页。

在前言的最后,还要在“通用”的水平上提出一个评语:将数字革命比喻为工业革命并不是夸张。网络对世界贡献的财富无与伦比。无可争议,网络搜索引擎是攫取这些财富的最为有效的工具。让我们积极投入这个振奋人心的时代,去享用、去攫取吧!

## 关于“顶级检索人员网页”

[www.extremesearcher.com](http://www.extremesearcher.com)

万维网和搜索引擎正在不断地发生变化。而搜索引擎的变化有两个方面：完善中的变化和根本性的变化。后一种变化经常是原理的变化，并不是在怎样利用现有的原理去做上发生的变化。正是由于这种原因，本书所介绍的大部分内容会适用一个相当长的时期。然而，为了跟踪变化，我们向读者提供了“顶级检索人员网页”。该资源可以提供有关什么是新出现的和发生了什么变化的信息。重点是放在对我们有效检索搜索引擎方式有影响的变化上。这个网站还提供有许多到搜索引擎和在第十二章中提到的联机资源的链接。

检索“顶级检索人员网页”，请到 [www.extremesearcher.com](http://www.extremesearcher.com)。

希望你喜欢利用它，并将意见和建议用电子邮件发送给 [ran@extremesearcher.com](mailto:ran@extremesearcher.com)。

# 目 录

## CONTENTS

### 序

### 致谢

### 前言

### 关于“顶级检索人员网页”

|                                    |    |
|------------------------------------|----|
| <b>第一章 搜索引擎概述</b> .....            | 1  |
| 发展简史.....                          | 1  |
| 搜索引擎的构成.....                       | 4  |
| 数据库后面的数据库.....                     | 7  |
| 各类入口.....                          | 8  |
| 典型搜索引擎主页的构成要素.....                 | 10 |
| 对网站服务的实际期待.....                    | 13 |
| 搜索引擎覆盖的主题领域.....                   | 16 |
| 更新的频次.....                         | 16 |
| 典型的检索和排序因素.....                    | 17 |
| 搜索引擎的性能评价.....                     | 19 |
| 搜索引擎检索的重复情况.....                   | 20 |
| <b>第二章 常用的检索方法</b> .....           | 22 |
| 检索对象的选择.....                       | 22 |
| 布尔逻辑检索.....                        | 23 |
| 截词检索.....                          | 25 |
| 词组和靠词检索.....                       | 26 |
| 名称检索.....                          | 27 |
| 范畴检索——在记录的特定部分中检索.....             | 28 |
| “More Like This One”（更多相似的内容）..... | 36 |
| 输出的选择.....                         | 37 |
| 国际版本.....                          | 37 |
| 各类入口.....                          | 37 |
| <b>第三章 AltaVista</b> .....         | 45 |
| 概述.....                            | 45 |
| AltaVista 的主页.....                 | 47 |
| AltaVista 的工作原理.....               | 47 |

|                                          |    |
|------------------------------------------|----|
| AltaVista 的 Power Search (强效检索) .....    | 48 |
| AltaVista 的 Advanced Search (高级检索) ..... | 50 |
| AltaVista 的 Ranging Search (排序检索) .....  | 51 |
| 检索方法 .....                               | 52 |
| 结果页面 .....                               | 57 |
| AltaVista 的网站目录 .....                    | 61 |
| 其他可检索的数据库 .....                          | 61 |
| 其他各类入口 .....                             | 64 |
| 其他功能 .....                               | 65 |
| 小结 .....                                 | 66 |
| <b>第四章 Excite</b> .....                  | 68 |
| 概述 .....                                 | 68 |
| Excite 的主页 .....                         | 69 |
| Excite 的工作原理 .....                       | 70 |
| Excite 的高级检索 .....                       | 70 |
| 检索方法 .....                               | 71 |
| 结果页面 .....                               | 73 |
| Excite 的目录 .....                         | 76 |
| 其他可检索的数据库 .....                          | 76 |
| 入口与个性化特征 .....                           | 77 |
| 其他各类入口 .....                             | 80 |
| “PRECISION SEARCH” (精确检索) 页面 .....       | 81 |
| 小结 .....                                 | 82 |
| <b>第五章 Fast Search</b> .....             | 84 |
| 概述 .....                                 | 84 |
| Fast Search 的主页 (简单检索) .....             | 85 |
| Fast Search 的工作原理 .....                  | 86 |
| Fast Search 的高级检索 .....                  | 86 |
| 检索方法 .....                               | 88 |
| 结果页面 .....                               | 91 |
| 目录 .....                                 | 92 |
| 其他可检索的数据库 .....                          | 92 |
| 其他各类入口 .....                             | 92 |
| 小结 .....                                 | 92 |
| <b>第六章 Google</b> .....                  | 94 |
| 概述 .....                                 | 94 |

|                                 |     |
|---------------------------------|-----|
| Google 的主页·····                 | 95  |
| Google 的工作原理·····               | 96  |
| Google 的高级检索·····               | 96  |
| 检索方法·····                       | 97  |
| 结果页面·····                       | 100 |
| 目录·····                         | 102 |
| 其他可检索的数据库·····                  | 103 |
| 特殊的各类入口·····                    | 103 |
| 小结·····                         | 104 |
| <b>第七章 HotBot</b> ·····         | 105 |
| 概述·····                         | 105 |
| HotBot 的主页·····                 | 106 |
| HotBot 的工作原理·····               | 107 |
| HotBot 的高级检索·····               | 108 |
| 检索方法·····                       | 110 |
| 结果页面·····                       | 114 |
| Hotbot 的目录·····                 | 116 |
| 特殊的选择和各类入口·····                 | 116 |
| 小结·····                         | 118 |
| <b>第八章 Lycos</b> ·····          | 119 |
| 概述·····                         | 119 |
| Lycos 的主页·····                  | 120 |
| Lycos 的工作原理·····                | 121 |
| Lycos 的高级检索·····                | 121 |
| 检索方法·····                       | 122 |
| 结果页面·····                       | 125 |
| Lycos 的目录·····                  | 126 |
| 其他可检索的数据库·····                  | 127 |
| 个性化和各类入口·····                   | 128 |
| 其他各类入口·····                     | 129 |
| 小结·····                         | 131 |
| <b>第九章 Northern Light</b> ····· | 132 |
| 概述·····                         | 132 |
| Northern Light 的主页·····         | 133 |
| Northern Light 的工作原理·····       | 134 |
| Northern Light 的高级检索·····       | 134 |

|                            |            |
|----------------------------|------------|
| 检索方法                       | 136        |
| 结果页面                       | 141        |
| 目录                         | 142        |
| 特殊的选择和“Search Forms”(检索方式) | 143        |
| 小结                         | 149        |
| <b>第十章 Yahoo!</b>          | <b>150</b> |
| 概述                         | 150        |
| Yahoo! 的主页                 | 151        |
| Yahoo! 的工作原理               | 152        |
| Yahoo! 的高级检索页面             | 152        |
| Yahoo! 的检索方法               | 154        |
| 结果页面                       | 155        |
| Yahoo! 的主题目录               | 156        |
| 个性化和各类入口                   | 157        |
| 其他各类入口                     | 159        |
| 小结                         | 160        |
| <b>第十一章 元搜索引擎</b>          | <b>161</b> |
| 元搜索网站                      | 161        |
| DOGPIL                     | 163        |
| IXQUICK                    | 166        |
| MetaCrawler                | 169        |
| ProFusion                  | 172        |
| Search.com(前称 SavvySearch) | 174        |
| 其他元搜索网站                    | 177        |
| “客户”元搜索程序                  | 178        |
| 小结                         | 179        |
| <b>第十二章 其他搜索引擎与跟踪</b>      | <b>180</b> |
| 其他搜索引擎                     | 180        |
| 跟踪                         | 181        |
| 搜索引擎的新闻网站                  | 181        |
| 会议                         | 183        |
| 期刊                         | 183        |
| <b>第十三章 结束语</b>            | <b>185</b> |
| 名词解释                       | 187        |
| 出版者的话                      | 192        |

# 搜索引擎概述

## 发展简史

搜索引擎的发展历史十分短暂，还不到十年的时间。本节概述它的发展简史。

在有网络搜索引擎之前，是一派混乱景象。要想在 Internet 上查找某些信息，必须知道信息的确切地址。“gopher”的研制迈出了走出混沌、走向开始以某种程度组织 Internet 内容的极为重要的第一步，开始用菜单形式排列服务器的 Internet 地址。

（“gopher”这个词本来是 Minnesota 州立大学的吉祥物的名称，从此诞生了第一个“gopher”。）“gopher”不能处理 HTML 文件，主要是处理标引文件的标题或极为简单的描述。进入 gopher，可以下载所选择的文件。从 gopher 产生 Archie（检索 gopher）、从 Archie 产生 Veronica（检索全部“gopher 空间”）、从 Veronica 产生 Jughead，此时，它们已经与卡通文字没有太大的关系。在此之后，它们的名称被专业化。甚至很少有人知道 Jughead 曾经是什么了。

Gopher 出现刚刚过了两年，就被万维网的迅速发展而掩盖了。用万维网可以检索超文本，可以全文检索，可以使用图像浏览器等一些使用方便、相互交互的高技术——而且研制出了网络搜索引擎。

第一个成功的搜索引擎是 WebCrawler，由华盛顿大学研制，1994 年 4 月亮相。在一年之内，有三家搜索引擎相继出台，包括 Lycos、Infosseek 和 Open Text。在 1995 年的晚些时候，AltaVista 和 Excite 问世。极为有趣的是，今天我们检索人员使用的检索技术已

经在某种程度上出现在这些早期的搜索引擎中了，这包括布尔逻辑检索、截词检索等方法。遗憾的是，并且是随着这种情况对目前的不断影响，竟没有搜索引擎使用在联机检索例如 DIALOG 和 LEXIS-NEXIS 中已经使用的复杂的检索技术和方法。此外，既没有搜索引擎，也没有它们的同类——网站目录，使用已经有上百年历史的、完整的分类法的理论和实践。这种情况，以一种实用的方式告诉我们，大多数的搜索引擎都是为偶尔为之的上网人员研制的，并没有针对那些急于使用更为复杂的方法和技术的人们。

HotBot 出现于 1996 年，Northern Light 出现于 1997 年。HotBot 使用了较为复杂但用起来很方便的界面，配备了规模庞大的数据库（是 1997 年末之前最大的搜索引擎）。Northern Light 将网站检索与独家专有的信息资源紧密结合。Google 出现于 1998 年。它按“被引用次数”对结果排序，并与一个超简单的界面有效结合，成为一种广为各类检索人员使用的搜索引擎。与此同时，争做最大的搜索引擎的竞争也有所缓解，直到 1999 年 Fast Search 出现。Fast Search 宣称，它的数据库含有 2 亿项记录。由于这一影响，外加其他方面的竞争因素，对规模的竞争又异军突起。到 2000 年 1 月，有 4 个搜索引擎突破 2 亿项记录。

在“早期的”搜索引擎中，Open Text 首先退出了历史舞台。到 1998 年的早些时候，它已不复存在。在两三年内，更多的搜索引擎销声匿迹。与此同时，已有的搜索引擎不断争秀，其中，有些过于肤浅、有些则在完整的“检索”面之外，增加了“入口”特征。我们可以预期，这些搜索引擎的建造者会不遗余力地提高检索能力。有迹象表明，竞争会沿此底线常盛不衰。在某些情况，确实是沿着正确发展道路迈步，如果搜索引擎开始实现这一目标的话。

在企业界的这块剩余的风水宝地，搜索引擎公司对流行风尚持有极大的怀疑态度。在 1996 年和 1997 年，这种风尚令人认为，搜索引擎已经有了一种“高级”的方式，无论这种高级方式在实际上是否比复杂方式多做了什么，也不管相同的东西是否被安在主页上。

出于效益上的重要考虑，在 1998 年，提出了“个性化”和“入口化”。个性化的入口或“网站频道”的思想被充分体现在在主页上由用户确定和选择的新闻领域、当地天气和电视节目、个人股票行情跟踪、个人效率手册等方面。（当然还有占星术。）搜索引擎的建造商都希望跟踪别的领先的搜索引擎，并认为，这种方法



不过是吸引广告收益罢了。受这种思想的推动，上述两种紧密相关的模型迅速成为主要搜索引擎的几乎是通用的企业模型。虽然还有许多的用户还对它没有足够的认识，但是，从把网站提高到家庭和案头的“用具”层面这种角度来看，这种入口化/个性化的方法确实向前迈进了一大步——人们可以信手拈来、又不复杂、可以频繁使用，而且，最为重要的是具有具体和明显的好处。

在 1999 年和 2000 年，出现了更为精巧的、提示较少的，但却是极为强大的入口，从而导致有关入口的概念的争议。在使用入口的头一年多里，附加的工具（例如网站目录等）主要被放在主页上，希望人们使用它们。在 1999 年，发生了重要变化，这些另加的内容被自动地放入结果页面中——人们在检索搜索引擎数据库的同时，还可以检索主题目录，公司名录等，由此产生的结果也被放到常规结果一块。资源的整合大大提高了结果的质量。检索人员可以得到高度相关的结果，从而也再不必单独使用几个检索工具了。用户不仅可以检索网站索引，而且不用花太大的气力，就可以检索网站目录、公司名录和工具书等——宛如检索老式的联机商业数据库中的“参见文档”。

再下一步发展是，赋予用户就像赋予搜索引擎建造商一样。引起用户关注的工具被保存、强化、复制和评价。从网络搜索引擎出现一开始，那些想阅读本书的人（顶级检索人员）和需要本书所讲述的入口和工具的人，根本不是搜索引擎的主要用户。主要用户并不关心复杂的、面向研究的检索方法。我们来看看有代表性的检索，其情况便可一目了然。Lycos 的提供商提供了一个有趣的，但有时是经过压缩的受喜爱的检索的列表。在有代表性的一周内，在排序在前面的 50 次检索中，有 46 次是检索娱乐、体育或游戏领域。这种情况与信息检索毫无瓜葛。但是却需要面对这样的实际，多数搜索引擎想挣钱的主要地方，不再将搜索引擎用在专业目的的研究人员身上。好在使用搜索引擎的人数不断增加。而将搜索引擎用于专业目的、用于投资、用于提高有关科学、人文、商业和医学方面的知识水平的人员，数量增长得更快。检索“世界摔跤协会”的次数不可能下降。然而，在提高知识水平上的、有价值的检索的次数正在增长。有多种理由可以让搜索引擎的生产商注意到顶级检索人员。但顶级检索人员也需要使用更为高级的检索方法，只有这样，这些检索方法才能长命百岁，并得到改进。