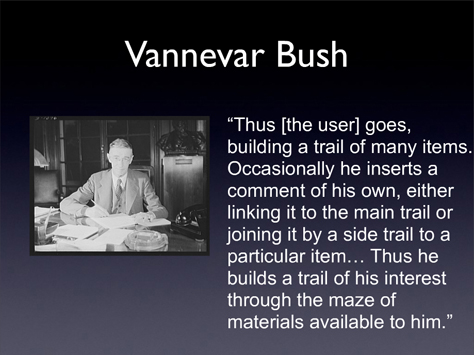
搜索引擎大事记

**As We May Think (1945):**

在享受过受二战影响的对科学发展的黄金时代后，超文本和内存扩展到概念在1945年7月真正走进来人们的生活，Vannever Bush的“As We May Think”在Atlantic月刊上发表了。



As We May Think.

他呼吁科学家们一起工作，建造一个造福全人类的一体的知识库。这里是摘录他主要观点的一些句子和段落。

专业化对取得进展来说越来越重要，相应的，在学科之间架起桥梁的努力也越来越肤浅。

面临的困难似乎较少因为我们不恰当发表的看法的程度和如今兴趣的种类，而更是因为我们现在的能力根本不足以使这些看法产生多大的影响。人类的获得的经验正在以巨大的速率增长，我们将这些迷宫一样的结果提取为暂时的重要条目的方式就像以前我们造横帆的大船一样（工程量大）。

如果一项纪录在科学上有价值，它必须一直扩展和储存。总之，它要能被查阅。

他不仅是数据存储坚定的支持者，他也相信如果数据源对人类有用，那么我们应该尽我们最大的努力使这些数据源好好运转。

我们在获取这些纪录时的愚昧主要是由索引系统的人工操作引起的……人不能希望人工的完全复制精神过程，但是他肯定能从这个过程中学到东西。至少他能进步，因为他得到的知识是相对持久的。

假如人能反思灰暗的过去并更加完整客观的分析他现在的问题，那么他的精神境界能得到提高。人类建造了一个如此复杂的社会，所以他需要更完整的机械化他的纪录以便将他的经验变成一个有逻辑的结果而不仅仅是被他有限的记忆能力遗忘掉大半。

随后他提出了一个无限制的、虚拟的、快速、可靠、可扩展、联合记忆存储的和可恢复的系统的想法。他将这个设备命名为memex。

**Gerard Salton (1960s - 1990s):**

卒于1995年8月28日的Gerard Salton是现代搜索技术之父。他在哈佛和康奈尔大团队开发出了SMART信息恢复系统。Salton’s Magic Automatic Retriever of Text包含了类似向量空间模型、逆文档频率（IDF），术语频率（TF），术语偏离值（term discrimination values）等概念和相关反馈机制。

他写了一本56页的书，叫做 [A Theory of Indexing](http://www.amazon.com/gp/product/0898710154/103-9884481-4626254?v=glance&n=283155)，这本书里很好的解释了很多他的实验，这些实验大部分是以搜索为基础的。Tom Evslin发表了（文章）[a blog entry](http://blog.tomevslin.com/2006/01/search_down_mem.html)介绍和Mr. Salton工作的情况。

**Ted Nelson:**

Ted Nelson在1960年创造龄Xanadu项目，并在1963年制造了超文本这个概念。他在Xanadu这个项目中的目标是建造一个有简单用户界面的计算机网络，用来解决类似归属感这样的社会问题。

当Ted着在反对复杂的标记代码，受损链接和其他一些在万维网中与传统HTML相关的问题时，他得到了很多有关建造一个万维网的启发。

直到现在还有很多关于Xanadu项目为什么没有腾飞的争论。

这是 [提供了有关Nelson先生的背景和许多资源的Wiki链接。](http://en.wikipedia.org/wiki/Ted_Nelson)

**Advanced Research Projects Agency Network:**

ARPANet是最终导致因特网出现的网络。这里[是一篇介绍ARPANet很好的文章](http://en.wikipedia.org/wiki/Arpanet)和一个[1972年有关ARPANet](http://video.google.com/videoplay?docid=-7426343190324622223)很有趣的Google视频。

**Archie (1990):**



最早的几百个网站出现于1993年，大部分在大学里，但是很久之后才有了Archie。最早出现的搜索引擎叫Archie，是1991年由Alan Emtage，一个蒙特利尔的McGill大学的学生创造的。开始打算叫它“Archives”，但被简写成了Archie。

Archie用结合一个基于脚本的数据采集器和正则表达式，将用户查询和文件名相匹配的方法帮助解决了数据分散的问题。

Bill Slaski 提供了[关于archie的更多背景知识](http://www.seobythesea.com/?p=106).

**Veronica & Jughead:**

由于口口相传，Archie流行起来，它开始变成计算机的代名词，Archie变得如此流行使得Nevada大学的计算服务系统组开发了Veronica。Veronica和Archie有相同的目标，但它基于纯文本。不久，另一个与Veronica作用相同的用户界面Jughead出现了，他们都用作从Gopher传文件，Gopher是Minnesota大学的Mark McCahill1991年开发作为Archie替补的系统。

**File Transfer Protocol:**

Tim Burners-Lee 在这个时候出现了，但这是还没有万维网。当时人们共享文件的主要方法就是通过File Transfer Protocol（FTP）。

如果你有一个想要共享的文件，你就会设置一个FTP服务器。如果有人想要得到这个文件，他就会用一个FTP客户端。这个程序在小的团队中很有效率，但数据在收集过程中也变得不完整。

**Tim Berners-Lee & the WWW (1991):**



来源：维基百科：

当1980年六月到十二月间，Berners-Lee是CERN的独立合同工时，他提议了一个基于超文本的项目，促进了研究人员之间分享和升级信息的能力。在Robert Cailliau的帮助下，他建立了一个叫Enquire的原型系统。

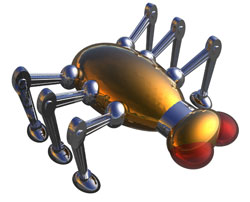
1980年之后，他离开CERN到John Poole's Image Computer Systems Ltd公司工作，但1984年他回到CERN并成为该公司的一位fellow。1989年，CERN是欧洲最大的因特网节点，Berners-Lee看到了将超文本加入因特网的机会。用他的话来说，“我必须将超文本的这个想法与TCP和DNS的想法结合--踏哒--万维网出现了”。他将Enquire系统中类似的重要想法用到万维网的创作当中,他设计并实现了第一个网络浏览器和编辑器（叫WorldWideWeb并在NeXTSTEP上开发）和第一个叫httpd的Web服务器（HyperText Transfer Protocol daemon的缩写）。

第一个网站在http://info.cern.ch/上建立并且在1991年8月6日上线。它解释了什么是万维网，人们怎样能拥有一个浏览器以及怎样建立一个网络服务器。他也是世界上第一个网络目录，因为Berners-Lee维护着他自己以及其他站点的条目。

1994年，Berners-Lee在麻省理工大学成立了[the World Wide Web Consortium (W3C)。](http://www.w3.org/)

Tim还创造了[the Virtual Library](http://vlib.org/)，世界上最古老的网络目录，Tim还写了一本书介绍开办这个网站的故事，叫做[Weaving the Web](http://www.w3.org/People/Berners-Lee/Weaving/).

**What is a Bot?（什么是Bot？）**



计算机机器人是一类用来执行人类无法达到的快速重复任务的一类程序。bot这个习语是网络上用来描述任何与用户交互或手机数据的东西的。

搜索引擎用“爬虫”来收集（或爬取）网络上的信息。他们就像普通浏览器一样请求网页信息，只是爬虫不仅读信息进行索引还要记录链接。

引用的链接可以被看作可信任网页的代称。链接的文本可以帮助描述一个网页的内容。多次引用的链接可以帮助确定一个网页或网站存在于哪个专题社区中。附加的链接可以帮助搜索引擎查找新的文件和便于以后搜索。

另一个机器人的例子时候Chatterbots，它是用来增加某个话题的分量的。这些机器人尝试想人一样表现并与人一起就特定话题交流。

**Parts of a Search Engine:（搜索引擎的一部分：）**

搜索引擎主要由三部分组成。搜索引擎**爬虫**通过跟踪链接来请求那些没有被索引过的页面，或那些上次索引之后就没有更新过的页面。这些爬来的页面被增加到搜索引擎的**索引**里（也叫目录）。当你用主搜索引擎搜索的时候，你不是真的在搜索网络，而是搜索稍微有一些过时的内容索引，这些索引可以大体上代表整个网络的内容。搜索引擎的第三部分是**搜索界面和相关软件**。对每一次查询请求，搜索引擎都要做以下这些事情中的大多数或全部：

保存用户输入的查询，检查（用户输入）并用高级语法匹配以及坚持用户有没有拼写错误，对拼写错误给出推荐的流行查询或纠正后的值。检查这个查询是否与其他垂直搜索的数据库相关（比如新闻搜索或产品搜索）并从这种查询经常搜到的结果里挑选几个提供相关链接。收集这个搜索结果相关网页的条目。这些结果根据网页内容、可用数据、链接引文数据进行排序。请求一下相关的广告放置在结果旁边。

搜索人通常只点击最上面的几个搜索结果，这在[Jakob Nielsen](http://www.useit.com/jakob/)的文章里被提到，并在[this search result eye tracking study](http://www.enquiro.com/eye-tracking-pr.asp).中做了备份。

**Want to learn more about how search engines work?（想知道更多关于搜索引擎如何工作的知识吗？）**

Google怎样收集和排序结果？Google的工程师Matt Cutts简单的讨论了Google是怎样工作的。Google的工程师Jeff Dean给华盛顿大学的同学做了一个演讲，谈到了在Google搜索的查询是怎样在搜索到视频的。芝加哥的Tribune运营了叫做Gunning for Google的任务，包括很多音频的采访，三个专栏，这些图形化的介绍了Google是怎么工作的。员工怎样为搜索引擎工作，以及因特网搜索引擎如何工作。

**Types of Search Queries:（查询的种类：）**

Andrei Broder写了一篇文章：[A Taxonomy](http://www.acm.org/sigs/sigir/forum/F2002/broder.pdf) of Web Search [PDF]，介绍了最经常的三种搜索条目：

信息--寻找关于一个题目的静态信息

事务--在哪里购物，在哪里下载，或与结果交互

导航--最后给我一个特定的URL

**Improve Your Searching Skills:（提高你的搜索技巧：）**

想更好的搜索吗？一般大规模搜索引擎提供：

高级搜索页面，帮助搜索人精炼他们的查询条件：请求新一点或老一点的文件，本地的或in nature？，在特定领域，用特定格式发布，或别的方式来精炼他的查询。比如说～字母说明与Google相关。垂直搜索数据库可以帮助结构化信息索引或限制搜索的索引以检索更结构化的源，文件以及信息。

Nancy Blachman的Google Guide提供了许多在Google中搜索到秘诀。Greg R.Notess的Search EngineShowdown提供了一个搜索特征的图表。

还有许多流行的小一些的垂直搜索服务，比如，Del.icio.us让你搜索有用户书签的URL,Technorati让你搜索博客。

**World Wide Web Wanderer:（万维网漫游者：）**

不久，网络上第一个机器人出现了。在1993年6月，Matthew Gray引入了万维网漫游者。他最初想要测量网络的增长，并且创建这个机器人来计算活动的网络服务器数量。不久他将这个机器人升级为计算实际URL数量。他的数据库叫Wandex。

这个Wanderer是一个解决方案也是一个大麻烦，因为它一天连接同一个页面数百次而导致了系统滞后。没多久之后，他修复了这个bug，但人们开始质疑机器人的价值。

**ALIWEB:**

1993年10月，[Martijn Koster](http://www.greenhills.co.uk/mak/mak.html)建造了一个类目录的网络索引器，或称ALIWEB，作为Wanderer的回答。ALIWEB爬取元数据并允许用户根据他们的网页描述提交他们希望索引的网页。这说明不需要机器人来收集数据，并且不多占带宽。ALIWEB不好的一面是很多人不知道如何提交他们的页面。

**Robots Exclusion Standard:**

Martjin Kojer还主持了[the web robots page](http://www.robotstxt.org/wc/robots.html)，这个网页为搜索引擎应该索引的内容和不应该索引的内容制定了标准。这个标准允许网管从整个网站的水平或基本页面阻止机器人。

默认情况下，如果信息存在于公共网络服务器，而且人们能链接得到，搜索引擎基本都会索引下来。

2005年Google领导了针对博客垃圾评论的改革，[创建了个人级的不能跟踪的属性](http://googleblog.blogspot.com/2005/01/preventing-comment-spam.html)。在Google的推动下，nofollow快速的改变了那些声称不是为商业目的及不能编辑的链接的目的。

**Primitive Web Search:（原始的网络搜索）**

直到1993年12月，三个成熟的搜索引擎机器人在网络上冲浪：JumpStation、World Wide Web Worm、和Repository-Based Software Engineering (RBSE)爬虫。JumpStation在网页上采集有关标题和题头的信息并用简单的线性搜索得到。随着网络的成长，JumpStation慢慢的停用了，WWW Worm索引标题和URL，JumpStation和World Wide Web Worm的问题是他们用他们发现结果的顺序排列这些结果，无差别（对待所有结果）。而RSBE爬虫执行了一个排序系统。

由于早期的搜索算法做的链接分析不够多，而且缓存太少的整页内容，如果你不知道你要查内容的确切名称或非常确定你要查的内容，那么你很难查到你想要的结果。

**Excite:**

Excite.

[Excite](http://www.excite.com/)是从项目Architext来到，Architext是1993年2月，有6个斯坦福的本科生发起的。他们的想法是通过统计分析词语的联系使得搜索的效率更高。他们不久以后，在1993年中，取得了资金支持并发布了他们的在网站上应用的软件。

1999年，一个叫@Home的广播提供商花65亿美元买下了Excite，并把它命名为Excite@Home。2001年10月<http://news.com.com/2100-1033-273689.html?legacy=cnet>。[InfoSpace](http://searchenginewatch.com/showPage.html?page=2158291)在破产程序下花1000万美元买下了Excite。

**Web Directories:**

VLib:

The Virtual Library.当Tim Berner-Lee开创了网络时他也创办了[the Virtual Library](http://vlib.org/)，一个松散的联合的专题专家维护的相关专题链接列表。

EINet Galaxy

Galaxy.

EINet Galaxy网络目录1994年1月诞生，它与现在的目录组织方式相似。它成功的最大原因是它不仅有web查找方式，还有Gopher何Telnet查找特征。1994年的网络容量其实并不要求一个网络目录，但其他的目录也跟随它接二连三的出现了。

Yahoo! Directory（雅虎目录）

1994年4月，David Filo和David Yang创办了[Yahoo! Directory](http://dir.yahoo.com/)，作为他们最喜欢网页的一部分。随着他们链接数量的增长，他们必须重新组织变成可搜索的条目。使这个产品超过Wanderer的是他们对每个URL提供了一个人力编制说明。随着时间的增长，由于Yahoo! Directory增长的雅虎变成了一个包含商业收费的站点。随着时间增长，列出商业站点的比率增加。目前的一年开销是299美元。很多提供信息的站点都免费加到雅虎目录下。

Open Directory Project

The Open Directory Project.

1998年，Rich Skrenta和他朋友组成的一个小组创办了[Open Directory Project](http://dmoz.org/)项目，是一个任何人都可以下载并部分或全部使用的一个目录。ODP（或称DMOZ）是因特网上最大的目录，几乎全部是志愿的编辑在维护。Open Directory项目后来受挫，它的网站管理员们等待着被雅虎目录并掉。[Netscape](http://news.com.com/2100-1023-218106.html)在1998年11月把Open Directory项目买下来了。同一个月的晚些时候，美国在线宣布将以45亿美元的价格收购Netscape公司的全部股份。

LII

LII.

Google提供图书管理员时事通讯，以帮助图书管理员和其他网站编辑，使信息更易于访问和更好的分类网站。第二个Google管理员时事通讯是由[Librarians' Internet Index](http://lii.org/)的执行官Karen G. Schneider发明的。LII是一个提供给图书管理员的高质量的目录。她的文章解释了当她和她的员工查找可信资源添加到LII时他们在找什么。大部分其他的目录，特别是那些有付费选项的项目，跟管理员选择有限目录查找相比质量更差。

[Internet Public Library](http://www.ipl.org/)是另一个维护的很好的网站目录。

Business.com

Business.com Directory.

由于维持一个目录是一件耗时的事情，而且一般而言，由于商业模型缺乏规模效应，使得目录的质量和大小会在得到了半打目录以后急剧下降。还有很多的小行业，垂直的，或面向本地的目录。例如[Business.com](http://www.business.com/)，就是一个商业网站的目录。

Looksmart



Looksmart是1995年成立的。他们凭借频繁增加他们引用率的和雅虎目录竞争。2002年过渡成一个每次点击支付的供应商，对列出站点的每一次点击，收取点击费。虽然这让他们提供起来那些需要付费的列表给MSN之类的门户网站而从中获利，但这使得他们经营起来的信誉和忠诚死亡了。[2003年，当微软宣布他们要丢弃Looksmart（这个合作伙伴）](http://www.theage.com.au/articles/2003/10/07/1065292591055.html)时，这也中断了他们的商业模式。

在2002年3月，Looksmart买了一个叫[WiseNut](http://wisenut.com/)的搜索引擎，但这没有带来牵引力。Looksmart同时还拥有一个垂直网站的文章列表，但由于相关性不大，LookSmart几乎失去了他们所有的动力。1998年，Looksamrt想通过花2千万美元收购非商业的Zeal目录来扩大它的目录，但在2006年3月28日，Looksmart关闭了Zeal目录，并希望用一个社区书签程序，[Furl](http://www.furl.net/),来控制拥堵问题。

**Search Engines vs Directories:（搜索引擎vs目录）**

大部分的搜索引擎只有有限的编辑检测程序，但主流搜索引擎都用自动搜索算法处理相关性，这些算法会判断网络上链接的能力。事实上，像[TrustRank](http://www.seobook.com/archives/000661.shtml)一类的算法，对信任站点的网络图给予更多信任，而不需要搜索引擎做编辑检测。因此，一些更优雅的搜索引擎允许那些链接到其他站点的网页在编辑检测的时候给予关键投票环节。

和高自动化的搜索引擎不同，目录是手工编制分类的网站。目录花费更大也更耗时，因为缺乏规模性而且需要手工加入条目并定期检查所列出网站的质量。

大体上，目录一般用在专业垂直目录，即时新闻网站（比如博客），和社区书签站点（比如[del.ici.ous](http://del.icio.us/)）。此外，我提到的这三种发布格式也帮助提高了主流搜索引擎的相关度，这些都进一步减少了搜索引擎对目录的依赖。

**WebCrawler:**

WebCrawler.

华盛顿大学的Brian Prinkerton在1994年4月20日发布了[WebCrawler](http://www.webcrawler.com/)。这是第一个索引整个页面的爬虫。随后它变得非常流行以致在白天都不能使用。最终，美国在线买下了WebCrawler并把它运行在他们的网络上。 到了1997年，Excite买下了WebCrawler，美国在线开始用Excite来维护他们的NetFinder。WebCrawler为随后的很多服务打开了一扇门。一年之内出现了Lycos，Infoseek和OpenText.

**Lycos:**

Lycos.

[Lycos](http://www.lycos.com/)是下一代搜索的主要发展，它是1994年7月在卡内基梅隆大学设计的。Michale Mauldin是这个搜索引擎的主要负责人并仍然是Lycos有限公司的首席科学家。在1994年7月20日，Lycos发布了一个拥有54000个文件的目录。除了提供相关的检索排名， Lycos公司还提供了前缀匹配和词语近似功能。但Lycos的最大区别是他目录的完全大小。到1994年8月，Lycos识别了394000个文件。到1995年1月，这个目录包含了150万文件。到1996年11月，Lycos搜索了超过6千万个文件--比其他搜索引擎都多。在1994年10月，在Netscape搜索单词“surf”的列出的搜索引擎里，Lycos排名第一。

**Infoseek:**

[Infoseek](http://web.archive.org/web/19960512212113/http:/www.infoseek.com/)也是从1994年开始的，据称他们是1月成立的。他们并没有带来多少创新，但他们提供了一些附加功能，在1995年12月他们宣称Netscape把他们作为默认搜索，这给了他们很大的曝光度。Infoseek的一个流行功能是运行网络管理员实时为搜索索引提供页面。这是一个搜索垃圾邮件者的天堂。

**AltaVista:**

[AltaVista](http://www.altavista.com/)在网络上的争论开始于同一个月。AltaVista为web场景买下了很多重要特征。他们有几乎无限的带宽（就当时而言），他们是第一个运行自然语言查询的，允许高级搜索技术并且允许用户在24小时之内添加或删除他们自己的URL。他们甚至允许入站链接检查。AltaVista还提供了许多搜索建议和高级搜索特性。

由于不好的管理，人们担心的结果开始出现，与门户相关的现象使得AltaVista在Inktomi和Google的越来越受欢迎中淡出了历史舞台。在2003年2月18日，[Overture](http://www.corporate-ir.net/ireye/ir_site.zhtml?ticker=OVER&script=410&layout=0&item_id=383636)签署一个协议表示希望以8千万美元股票和6千万美元现金收购AltaVista。在雅虎买下了Overture后，他们把一些AltaVista的技术加入到雅虎搜索中，并偶尔将AltaVista作为测试平台。

**Inktomi:**

Inktomi公司1996年5月20日成立，他们的搜索引擎是Hotbot。两个Cal伯克利分校的人用他们的研究中得到的技术改进创造了Inktomi。Hotwire列出了这个站点，很快地，这个站点火起来。

在2001年10月，Danny Sullivan写了一篇名为[Ink](http://searchenginewatch.com/showPage.html?page=2164241)tomi Spam Database Left Open To Public的文章，介绍了Inktomi是如何偶然的允许公众访问他们垃圾网站的数据库，当时那个数据库里有100万个URL。

虽然Inktomi开创了付费收录模式，但它离Overture的每次点击支付的拍卖模式的效率还差很多。给他们搜索结果许可的盈利仍然抵消不了他们的大规模开销。他们在创建一个可获利的商业模式的努力失败了，并在2003年12月把它以大约23.5千万美元或每股1.65美元的价格销售给雅虎。

**Ask.com (Formerly Ask Jeeves):**

在1997年4月，Ask Jeeves以一个自然语言搜索引擎的身份面世了。Ask Jeeves用人工编辑来处理每一个用户查询。Ask用了一段时间DirectHit，DirectHit以结果的流行度排序，但这种技术的核心算法很容易得到垃圾信息。2000年，Teoma搜索引擎问世了，它用Subject Specific Popularity的集群方式来组织站点，这是查找本地网络社区的另一种方式。在2001年，Ask Jeeves买下了Teoma用来代替他们所用的DirectHit搜索技术。

Jon Kleinberg的文章Authoritative sources in a hyperlinked environment [PDF]是一个介绍Teoma是怎么诞生的文章。Mike Grehan的Topic Distillation [PDF]解释了subject specific popularity是怎样工作的。

在2004年3月4日，Ask Jeeves同意以930万股基本股和期权以及15千万美元现金买下Interactive Search Holdings。在2005年3月21i日，Barry diller的IAC agreed to acquire Ask Jeeves文章中，IAC以18.5忆美元买下了很多著名站点，如Match.com, Ticketmaster.com, Citysearch.com. 并主推Ask。在2006年Ask Jeeves改名为Ask，同时他们把Teoma商标取消了。

**AllTheWeb**



[AllTheWeb](http://www.alltheweb.com/)是一个1999年发布的搜索技术平台，这个平台展示了Fast的搜索技术。他们有一个有高级搜索功能的整洁的用户界面。但2003年2月23日，AllTheWeb被[Overture](http://searchenginewatch.com/showPage.html?page=2165241)以7千万美元的价格收购了。在雅虎买了Overture之后，他们把AllTheWeb的一些技术应用与了雅虎搜索，并偶尔使用AllTheWeb作为测试平台。

**Meta Search Engines**

大部分源数据搜索引擎从其他多种搜索引擎的结果里搜索，然后结合并对结果重新排序。这在搜索引擎在网络中搜索并不高效且每个引擎有自己的索引方式的时候是一个很有用的特征。因为搜索改善了元搜索引擎减少的需求。

[Hotbot](http://www.hotbot.com/)由Wired所有，它有时髦的颜色，快速的结果，和一个很酷的名字，但它在Lycos买下它并且不予重视之后不就就销声匿迹了。之后它重新以一个元搜索引擎的角色出现，不同与大部分的元搜索引擎，Hotbot一次仅在一个搜索引擎的结果上增加条目，但它允许使用者在流行的搜索引擎中选择。目前[Infospace](http://www.infospace.com/)拥有的[Dogpile](http://www.dogpile.com/),可能是现在市场上最流行的元搜索引擎，但像其他元搜索引擎一样，它的市场占有份额也很小。

元搜索引擎的一个大问题就是大部分元搜索引擎会混合点击付费的广告在他们的搜索结果里，对一些商业查询，70％或更多的结果是付费的结果。我也创造过[Myriad Search](http://www.myriadsearch.com/),他是一个不含广告的免费的元搜索引擎。

**Vertical Search**

主流搜索引擎正在他们核心算法的搜索产品之外争取垂直搜索的内容和市场份额。例如，雅虎与MSN回答问题的服务，在上面，人们免费解答别人提出来的问题。Google有类似的产品，但要为他们的解答问题的工作付费。

Google，雅虎和MSN还在为谁是网络上默认的视频平台而争夺，这是一个垂直的市场，在这里，YouTube同样有着很重要的位置。

雅虎和微软在书籍搜索的领域联合起来，他们的产品叫Open Conten Alliance（开发内容联盟）。Google，在这个领域里独自奋斗，拥有一个Google Book search。

这三家主流搜索引擎都有新闻搜索的服务。雅虎有一些保费提供商让用户在新闻搜索结果里包含自己的内容。Google与AP及很多其他的新闻源合作将他们的新闻数据库扩展到超过200年以前。一家很有名的新闻服务网站Topic.net买了75％的所有权给三家最大的报业公司。成千的博客每日更新报道新闻，有一些还能与主流媒体竞争。如果没有新闻的足够选择，本地书签网站，例如[Del.icio.us](http://del.icio.us/)还经常更新流行条目。还有meme跟踪服务，例如[Techmeme](http://www.techmeme.com/)，跟踪在博客上流行的故事，还有类似[Digg](http://www.digg.com/)的站点，允许他们的读者为得到的信息投票增加其曝光度。

Google还有一个[Scholar search program](http://scholar.google.com/)用来帮助开展研究工作。

在一些垂直搜索领域，例如购物艘艘，一些第三方的厂商也有重要的市场份额，这来自于他们的线下分发和品牌（比如，黄页公司），或者在各大主流搜索引擎下做比较和仲裁。

2005年11月15日，Google发布了一个叫[Google Base](http://base.google.com/)的产品,它是一个你能想象到任何东西的数据库。用户可以上传物品和名称，描述，给它们合适的标签。从使用统计数据上，Google可以判断哪些垂直搜索产品他们应该开发和重视。他们相信拥有其他的垂直搜索产品能给他们的主要搜索产品带来市场。他们还认为，有针对性的衡量广告的相关搜索，可应用到其他场合，比如[Google bought dMarc](http://www.google.com/intl/en/press/pressrel/dmarc.html)就是一个广播广告公司。雅虎同样用收购流行产品的方式来扩大它的触角，比如，买了一个相片分享站点 [Flickr](http://www.flickr.com/),和一个社区书签站点[del.icio.us](http://del.icio.us/).

**Search Engine Marketing（搜索引擎营销）**

搜索引擎营销是通过搜索引擎进行的市场营销，通过有机搜索引擎优化，付费搜索引擎广告，付费收录程序来进行。

**Paid Inclusion（付费收录）**

就像之前提到的，很多网络目录都向商业站点收取一次性费用或年费。很多购物搜索引擎对放入他们数据库的站点按点击率收费。

如同大众搜索引擎一样，Inktomi使得付费收录模式流行起来。它们在2003年12月被雅虎收购。在雅虎放弃Google并开始自己的搜索技术研究之后，他们继续对列出的站点提供付费搜索程序。[Yahoo Search Submit](http://searchmarketing.yahoo.com/srchsb/choose.php)是唯一由大众搜索提供商服务的付费收录程序。Search Submit出售方式有按全年平均点击率出售和按每次点击出售两种方式。

**Pay Per Click（每次点击付费）**

每次点击付费广告允许搜索引擎以基于每次点击为单位的方式将目标买给广告商。典型的每次付费点击广告是以关键字为目标的，但在一些情况下，有些引擎也会以定位于本地，以行为定位，或运行厂商依据人气竞价。

在典型的每次点击付费广告拍卖会中，出价最高的竞价者会得到在关键字中排名第一的位置。一些搜索引擎，例如Google和微软，也会以网站连接速率作为价格的一项指标。这样能使他们的广告被点击的频率更高，而且使得广告更相关。创造有吸引力的广告并获得很高的点击率的商家将被允许少支付每次点击费用而获得流通。

**Overture (Formerly GoTo)**

Overture.

Overturn, 付费搜索引擎的先驱，最早是1998年Bill Gross以GoTo为名发布的。他的想法是在流通带宽中套利，并把它们出售。John Battlelle的文章The Search有一个有趣的章节是关于Bill Gross和overture的诞生的。John同时把这个章节发布在了[他的博客](http://battellemedia.com/archives/001775.php)里。

“我想到它越多，我越觉得因特网的价值在于它的责任。”Gross告诉我，“变现的保证是付钱给媒体的模式。”

Gross知道在过热和贪婪的市场提供几乎无风险的点击，保证了GoTo将腾飞。之所以很容易可以断定GoTo成功了，是因为因特网的泡沫导致像ouroboros的饥荒一样对带宽的需求，这家公司的管理使得萧条来的更慢了，之有一个简单的原因：它是成功的。

虽然Overture疯狂的成功了，但有它两个主要的缺点使得它没有取代Google的市场地位。

目标品牌定位：Google使它自己成长成为搜索的公司。Bill Gross决定不让Overture变成一个搜索公司，因为他害怕这会使他要出让合作伙伴。当美国在线选择Google当它的合作伙伴的时候，尽管Google也在扩大自己的品牌，这也已经几乎是在为首演的搜索广告平台Overture钉在棺材。广告联播网（Ad Network）的效率：Google AdWords 将广告点击率加入到广告价格中，这保证了更高的相关性和广告的有效性。直到2006年9月，Overture平台（这时候已经是雅虎搜索营销）仍然没有改正这个问题。

这两个缺陷意味着Overture严重依赖他的两个分销伙伴--雅虎和微软。Overture买下了AltaVista和AllTheWeb想要赢回一些筹码，但最终他们被[在2003年6月14日被雅虎以16.3亿美元收购了](http://news.com.com/2100-1030_3-1025394.html)。

**Google AdWords**

Google AdWords是2000年发布的。最早的版本由于使用CPM模式收取广告费用而失败了，有些关键字要价太高而没有人负担得起，而另外一些由于要价太低而损失。在2002年2月，Google用类似Overture的模式和重新发布了AdWords，但仍然加入了广告点击率作为排序的一个根据。

Affiliate 和其他的一些网络娱乐商很快的投入到AdWords的阵营，因为它很好的定位和使得最终客户能够舒服的坐在家里，穿着内衣而网络娱乐商仍能盈利。

随着时间的增长，AdWords越来越流行，主流大众开始纷纷接受它。Google开始关闭AdWords的一些漏洞。比如，为消除噪音并保持广告的相关，他们禁止了一个网站做多重广告。随后，他们开始考虑页面载入质量并建立基于质量的最小要价。这一举措挤压了许多套利者和联盟用户的利润。

Google打算采用AdWords的可跟踪的广告定位并扩展到其他媒体。Google已经在印刷品和报纸广告上做了测试。Google允许广告商从内容网站上买下图像和视频的广告。在2006年1月17日，Gppgle宣布他们买下了[dMarc](http://www.google.com/intl/en/press/pressrel/dmarc.html)广播公司，这家公司将帮助Google进行广播广告营销。

在2006年9月15日，Google与[Intuit](http://www.startupjournal.com/ecommerce/ecommerce/20060915-vara.html)结盟，允许小型买卖在QuickBooks上用QuickBooks买AdWords。它的目标是使得更多的小商家使用AdWords，使本地广告更相关。

在2007年3月20日，Google宣布他们已经对创建一个分布式每次动作付费联盟广告联播网进行了beta测试。在2007年4月13日，Google宣布他们花31亿美元买下了[DoubleClick](http://www.google.com/intl/en/press/pressrel/doubleclick.html)

**Google AdSense**

2003年3月4日，Google发布了他们的内容导向广告网络。在2003年4月，Google买下了Applied Semantics，这家公司有CIRCA技术可以帮助他们大幅度改善广告的针对程度。Google采用了[Tiffany co](http://www.tiffanysoutlet.com/)作为这个新的广告程序的名字。

[AdSense](http://www.google.com/adsense/)允许大小网络发布商在他们的内容上自动加入相关广告。Google最初允许以不同形式加入文本广告，最终允许加入图片广告和视频广告。广告商可以选择他们的目标关键词和广告形式的目标市场。

为了让这个网络成长和更有效率，Google加入了一个链接，允许广告商用自己内容网站的账号注册AdWords，Google还允许广告客户购买广告定位到特定的网站，网页，或统计信息。定位到网站的广告以每千次展示价格为基础（CPM），在广告拍卖会上与关键字定位和网站定位广告竞争。

本文来自：<http://www.yeeyan.com/articles/view/41036/12723>  
原文链接：<http://www.searchenginehistory.com/>