

第7章 搜索与设计

许多用户发现，使用传统的浏览器搜索所需信息是一种低效率的方法。通常，用户知道网站中已有的内容以及网站的位置。超级用户、频繁访问者以及其他一些人经常使用搜索工具，他们期望迅速得到搜索结果。网络站点的一个主要功能是提供一个搜索工具，这个工具可以使用户有控制网站内容的能力，并允许他们过滤掉不需要的内容而方便地浏览自己所关心的问题。比较大的网站，尤其是那些数据复杂的网站必须提供搜索工具，并应考虑使它成为中心导航方式。当然，搜索设备的设计必须考虑用户的因素，应仔细考虑以下内容：用户期望，搜索需要的类型，搜索页面设计，帮助系统和进行新搜索以前的搜索结果显示。

7.1 网际搜索

设计者在设计网站时，一定要注意不要陷入大量模仿网上现有搜索工具的误区。网际搜索的需求与单一站点或一组控制站点的搜索需求相比，有较大的区别。遗憾的是，用户通常期望站点搜索设备应与公共搜索设备相似，例如：Lycos。公共搜索设备不得不处理那些几乎是不可能出现的任务，比如搜集和索引那些由令人误解的信息填充的巨大而又不断变化的 Web。但是用户又设想通过快速简单的搜索快速地从这些繁杂的信息中得到有用信息。总之，网络搜索比在草堆上找一根针还困难。尽管有这些困难，用户仍然极大地依赖公共搜索工具。由于用户将逐渐了解这些搜索工具的功能，所以设计者应该考虑用户在使用网际搜索工具时的经验。下面将介绍有关 Web 搜索的一些组件以及其中涉及到的问题。

7.2 Web 搜索总论

Web 搜索的需求是令人沮丧的。用户期望在一个全球性的搜索引擎例如 Alta Vista (<http://www.altavista.com>) 快速地输入一个简单的搜索短语就可以得到现实的结果。想一下在图书馆里，你期望在几秒内在一本书中找到特定的部分，你就能理解曾经令人满意的搜索工具的不足。在搜索过程中用户将遇到太多的信息，或不相关的信息，或根本得不到信息。尽管有这些挫折，用户仍然希望搜索引擎可以在几分钟内给他们提供正确的结果。

搜索引擎应用中的许多问题导致用户搜索结果的不正确。搜索事实上仅仅应用于已知项目或非常专业的主题的搜索。在查找一个一般的短语时，以“hamburgers”为例，搜索引擎不必连接有关 hamburgers 的所有站点，或者比较大的 hamburgers 销售连锁店。事实上，在许多搜索引擎中测试这个搜索的结果将产生大量的链接，这些链接中的一部分与 hamburgers 没有太大的关系，比如它们可以和 Hamburg 和 Germany 有关，或与一些家庭网站的食谱相关。这个问题的出现是由于搜索太不专业化所致。当你搜索“White Castle Slders（美国一种著名的区域性汉堡包）”时，可能会得到一系列有用的信息。

当搜索一个主题的一般性信息时，用户希望使用目录而不是搜索引擎。目录与搜索引擎的不同在于目录是人为编辑的，并通常包括有限数量的链接。Yahoo可能是最著名的全球性的搜索目录，但是它现在也提供搜索引擎。实际上，许多搜索引擎已经开始同时提供搜索引擎和目录。一些流行的目录例如 www.about.com 或 www.dmoz.org 是由对某些特定的内容负责的个体组织的。目录的好处在于它仅限于对“优秀站点”的链接，并可以提供对站点的评论。

定义：Web目录是人为编辑和组织的收藏集，它包括网站链接及相关信息的描述和回顾。

与目录比较，搜索引擎更类似于电话号码簿，它只能用于搜索。这类似于使用信息服务并搜索电话号码，除此之外，你可以寻求与特定主题相关的信息。比如在美国使用 411 电话信息服务搜索“中国饭店”的号码，而不是搜索特定的某一个中国饭店。当搜索某一个特定的中国饭店时，用户将得到一个满意的结果。然而当搜索一般的信息时，如果比较幸运，系统将花费一段时间来搜索，然后给用户一个特定的中国饭店的电话号码，甚至于会给用户一个在系统目录中存在的有代表性的号码。在许多实例中，目录会给出列表中的第一个或任意一个项目。搜索引擎以同样的方式工作。它们擅长于返回专业的答案，但是结果会有变化。搜索引擎总是尝试成为综合性的并可以列出大量站点。尽管由于搜索引擎欺骗情况的增加而引入了人为编辑的方式，搜索引擎主要还是自动收集和组织链接。

定义：搜索引擎是网站信息的自动收集者和组织者，用户可以依赖它来运行搜索命令。

7.3 用户搜索方式

在了解搜索引擎的工作原理和如何利用内部和外部搜索引擎来提高站点设计水平之前，首先要了解用户怎样使用搜索设备。人们之所以使用搜索是有一些原因的。使用搜索的最主要原因是，他们在搜索存在的一些信息。例如用户搜索像 `RBA_4456` 这样一个特殊的部分。在这个实例中，假设搜索设备已经接触过相关的项目，尤其是如果该项目是唯一的时，用户搜索会相当简单。然而一般情况下，用户不知道当前待搜索项目是否存在，事实上他们正在搜索当前项目是否存在。诸如“Robot shops”这样的搜索是更一般的，在此搜索中，用户希望找到一个可以修理机器人的商店。其他情况下，用户可以简单地进行探索性的搜索，而得到对既有事务的更深入的理解。例如：“Robot Butler（机器人管家）”短语不仅可以搜索此类设备，而且可以搜索到提供机器人仆人的相关站点。人们希望用户进行已知项的搜索但奇怪的是，人们经常采用探索性的搜索。无论搜索的目的如何，用户都必须经过以下四个步骤。

1. 查询的表达

基于使用搜索设备的不同，用户创建的查询多少有变化。简单的查询可以包括关键字如：“Robot Butler”。复杂的查询可以包括逻辑运算如“Robot AND Butler”。许多搜索引擎利用带有符号的查询语句如“+Robot+Butler-Jeeves”。搜索工具甚至可以支持自然语言界面，这使得用户可以问像“Where can I buy a robot butler?”之类的问题。查询公式不仅可以包含选择的关键字，而且可以包括已经提炼的搜索标准，比如搜索区域、数据范围和数据类型等。此时用户也可以

定义返回结果的形式，比如每次返回十条记录，按最后更新时间排序等。然而除关键字以外的高级规则通常是高级搜索的一部分，它经常是被有经验的用户使用的。

2. 执行搜索和等待结果

搜索的第二步通常是等待一个短暂时间之后简单地单击按钮，此时间包括搜索引擎的运行和结果的显示。尽管在此期间没有太多的交互，但千万不要忽略。用户有时认为这是一个不连续的过程而不愿等待结果的显示。

3. 浏览结果

一旦结果在屏幕上列出以后，用户将仔细阅读它们，看其中是否包含有趣的信息。在此期间，用户很大程度上依赖于系统提供的信息，例如：相关信息的列表和结果的描述，包括综述、修改日期和文件大小。在此阶段为了决定下一步的操作，用户还可以对搜索结果进行过滤和排序。当然根据结果所做的决定还依赖于实际查询的返回结果。结果可能变化很大，这就是所谓的不包含任何匹配信息的负面结果。另一个极端的情况是有大量的返回结果，其中每个文件都在选择之中。大多数情况都在没有匹配文档和大量匹配文档之间。

4. 确定如何处理结果

基于搜索结果，用户将决定如何操作。例如，如果没有返回结果，用户可以使用新的查询语句重新搜索，或者放弃。如果结果没有正确地反映搜索要求，用户可以重新搜索。当然用户可以选择一部分搜索结果进行检验。当有大量结果时，用户可以探索其中的一部分或重新搜索或退出。

这种一般的想法在设计搜索设备时应予以考虑。在本章的后面部分将介绍一些用户在搜索过程可能涉及的理论和实际的设计建议。在此之前将综述搜索引擎的功能。

7.4 搜索引擎如何工作

搜索引擎如何工作呢？首先，通过一个称之为 Spidering的过程，大量的网页从各个网站聚集起来。然后被收集的网页按照它们的内容建立索引。最后，搜索网页被创建，以便用户可以输入查询表达式，看什么网页与他的查询相关。这个过程的最好情况是搜索引擎应创建得尽可能的大，然后用某种方法来组织它，最后由用户通过输入查询表达式，从大量的信息中找到自己所需要的内容。

7.4.1 收集网页

Web每时每刻都在增大，网络的实际大小是不可知的。在每一时刻都有大量的文档添加到Web中，与此同时，也有大量的文档被删除。收集所有的网页并保持它们的最新状态是至关重要的。用户通常想知道哪一个搜索引擎可以覆盖网络的大部分，事实上，即使是世界上最大的搜索引擎也只能覆盖整个网络在线文档的三分之一，一些引擎可能更少。在未来，这一情况可能会有所变化，但令人欣慰的是，并不是Web的所有内容都要检索。克服搜索结果信息的混乱是更加困难的一个问题。

大多数搜索引擎使用一种叫做 spiders、robots或gathers 的程序来搜集网页。我们使用“spiders”来称呼搜集网页的程序。它从一些URL开始，由在相应的列表中查找的用户提交或由

域名注册区中列出的 URL 组成。当 spiders 访问各种不同的地址时，它将保存整个网页或网页的一部分以便分析，同时要查找网页的链接并跟踪搜索。例如，如果 spiders 在访问 <http://www.democompany.com>，可能根据本网页的链接跟踪搜索。并非所有的搜索引擎都深层次地索引站点本身，但大多数引擎还是要跟踪搜索链接，尤其是当其他网页与当前页紧密相关或者包含大量细节信息时。

7.4.2 索引页

下一步搜索引擎将试图确定当前网页的主要内容。这一步通常称之为“索引”。每一个搜索引擎使用的方法有所不同，但基本上都包括以下内容：<TITLE>、<META>标签的内容、注释文本、链接标题、题头和正文。根据这些内容，搜索引擎将试图提取网页的主要内容。每一张网页都有不同的相关内容，在判断它的主要内容时，正文文本位置以及某些词语的出现频率都应考虑。然而并不是网页的所有细节对搜索引擎都有相同的反应。例如，“终止词”，它们经常被搜索引擎忽略，因为它们通常被认为是与有用信息不相关的内容。终止词一般包括 the、an、a 等。大多数引擎都包括有终止词，有一些搜索引擎比如 Alta Vista 会把 the 当作终止词。

尽管终止词可能通过限制检索文件的大小和把检索集中在更加细致的关键词上而提高效率，但是，这有可能导致与用户所想像的查询不匹配。新手可能会认为使用“ The Best Butler Robot ”作为查询语句比使用“ Best Butler Robot ”好。有时终止词对搜索有重要的意义。比如搜索“ Rock the Town ”这样一支歌曲，其中 the 是这个短语的必不可少的一部分，没有它其他的歌曲可能出现在查询结果中。然而如果是搜索“ Rock the Casbah ”，有干扰意义的 the 可以被删除，因为 Rock 和 Casbah 很少连续出现，除非在歌曲的名字中。考虑到 Web 无限制的主题，确定使用什么样的终止词会有很多疑问。

一旦使用许多关键字对网页进行了分析，它与其他有相同关键字的网页被归为一类并存储在数据库之中。归类和相关内容排序是搜索引擎的一个非常精细的操作过程。如何确定某一网页的级别高于其他网页是搜索引擎设计专家经常考虑的问题。

提供搜索机制

搜索的最终目的是搜索网页本身。搜索网页是用户实现查询的界面，它通常包含一个主查询文本编辑框，当然也包括其他一些由高级用户使用的关键字编辑框。例如，图 7-1 所示的 Northern Light (www.northernlight.com) 搜索界面。

在公共搜索引擎中，搜索页面有很大的变化。比较图 7-2 所示的 HotBot 的高级搜索界面和 Goto.com 的基本界面，可以很容易得出这一点。

像 www.ask.com 这样的站点支持从“ Why is the sky blue? ”这样简单的自然语言到布尔型表达式和其他过滤器等的检索。一旦执行查询，搜索引擎将提取相关信息，并把它们显示在结果页面上。图 7-3 展示了 Google (www.google.com) 搜索引擎的一个结果页面。

从结果网页中用户可以选择一部分进行分析，也可以重新定义查询进行二次检索，或者放弃当前操作，运用新的方法以达到目的。搜索引擎的一般功能示于图 7-4 中。

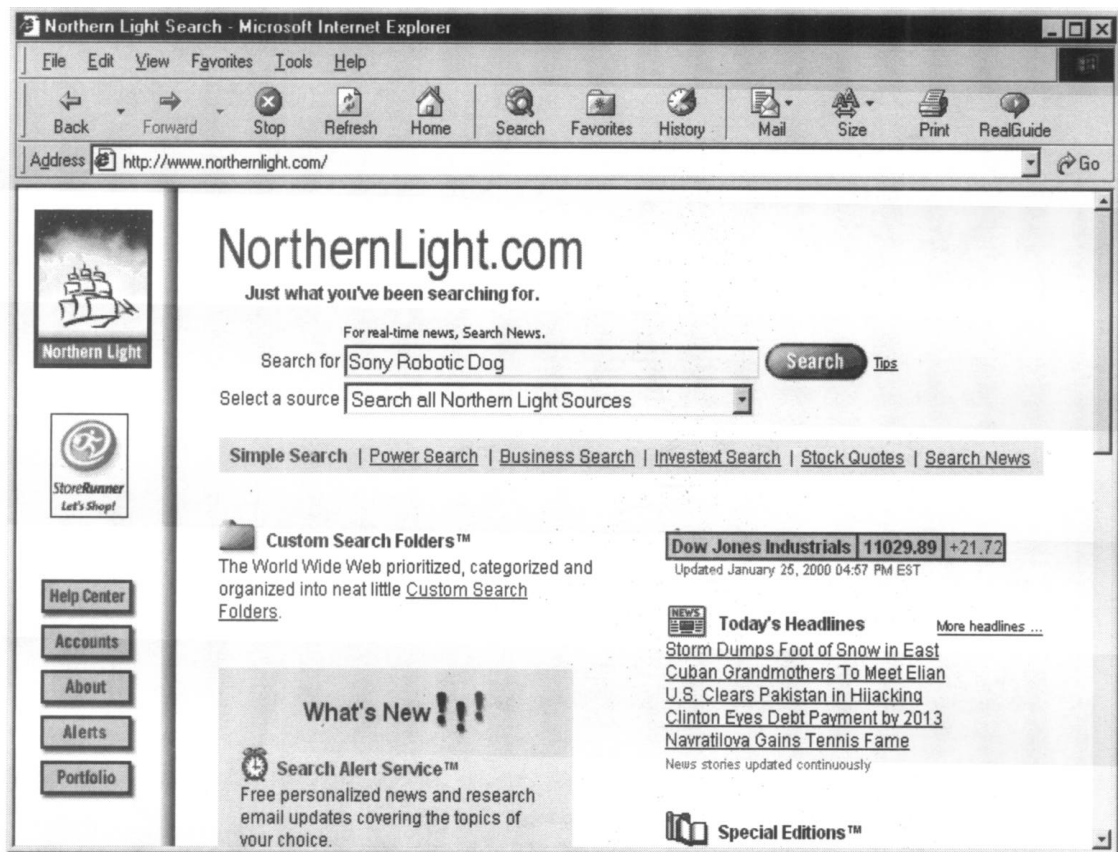


图7-1 North Light的基本搜索界面

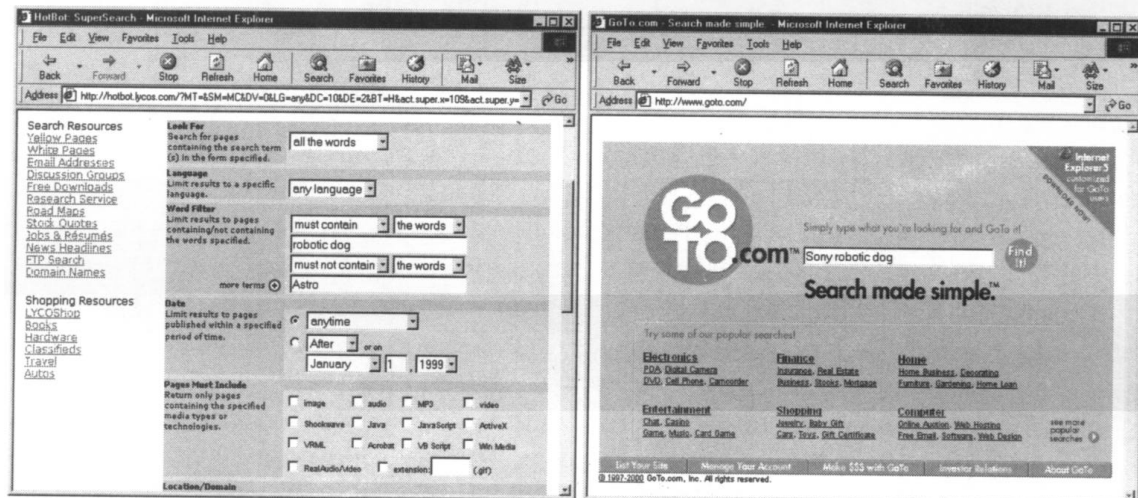


图7-2 搜索界面可以根据访问者类型变化

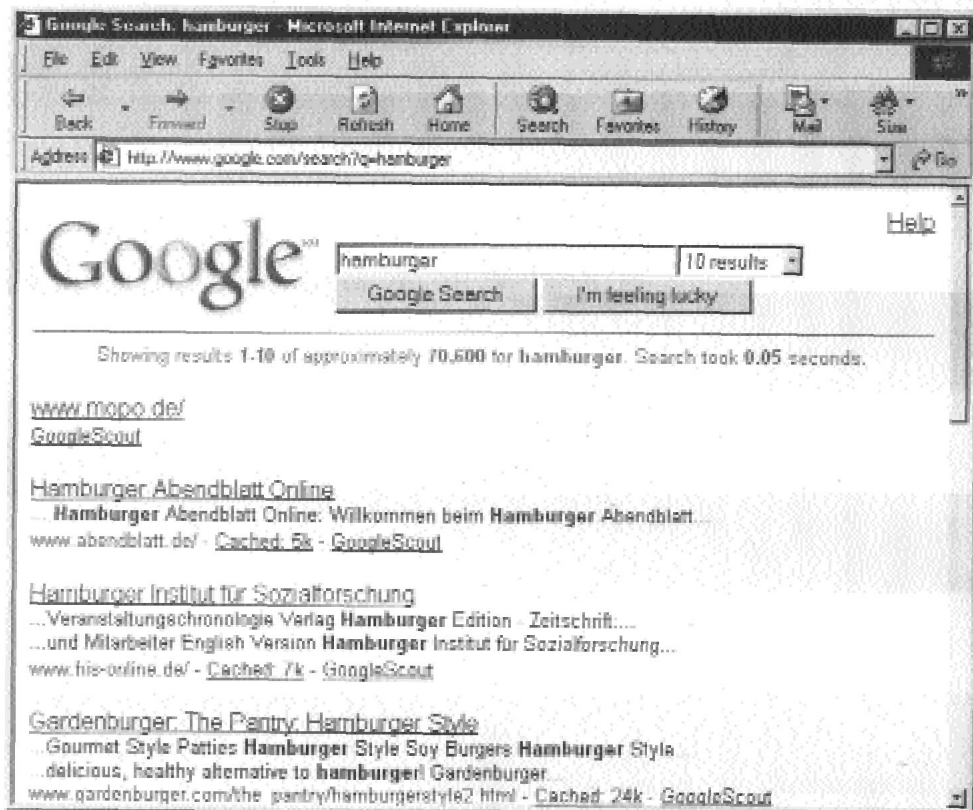


图7-3 Google的结果页面非常简洁而清晰

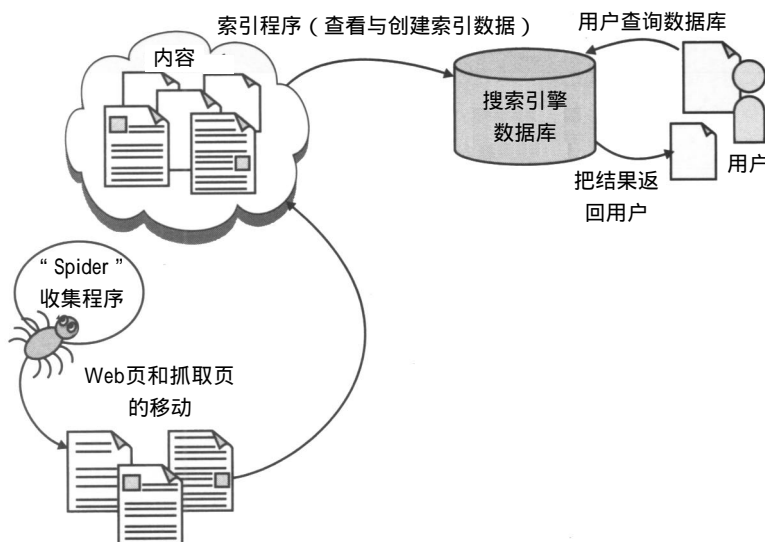


图7-4 搜索引擎的概况

理解用户对搜索引擎的期望是很重要的，因为用户可能运用他们自己的检索经验而抵制你设计的局部站点搜索引擎。标签、格式布局 and 结果页应当像用户期望的公共网页那样。当然并不是完全模仿公共搜索引擎。公共搜索引擎并非完全准确，它们经常是适应广告客户的需要，或大量地欺骗用户访问该站点以提高它们本身的知名度。值得庆幸的是，在本地检索范围内的组织中没有任何一个会借助于字面上的欺骗去提高等级。本地搜索引擎应被认为是准确和值得相信的。

规则：结合用户使用搜索引擎的经验，并使用类似的布局和标签设计本地搜索引擎，但要避免模仿那些不可控制的公共搜索引擎的设计方法。

在讨论增加本地搜索以前，先简单介绍一下用户如何使用搜索引擎和其他设备来提高站点的流量。由于它每个星期都会发生变化，这并不是对一个主题的完整讨论。读者为了查找最新的信息，都直接求助于网上现有的搜索引擎，尤其是Search Engine Watch (www.searchenginewatch.com)。

7.5 搜索引擎改进

站点所有者总是希望被搜索引擎列在第一位。假如你是一个不知名的旅游经纪人，但是你的站点被搜索引擎列为第一位，这样，客户可以通过搜索引擎首先光顾你的站点，你一定会欣喜若狂。但很遗憾，有很多人都想成为第一，如果被标记为第 4036 位的话，将不会有很大意义。事实上，列位为 20 位以后的站点根本不会有人访问。基于以上原因，网页设计者总是想方设法获知搜索引擎如何对网页进行分类，以及如何通过用某种方式建立关键字以提高自己的网页在搜索引擎中的排名。在某些方面，这一点类似于人们喜欢把公司的名字命名为“AAATravel”而使自己的公司排列于电话号码本的第一位。但是，如果所有的旅游经纪人都想把自己列于搜索引擎的前十位，那将会有有一个潜在的问题。设想如果整个国家只有一个电话号码本，那么将会有很多页都填满了以 AAA 开头的公司名称。网络也存在这个问题，这就是为什么当用户在运行一个查询时会出现很多结果的原因。

搜索引擎上的第一次大战以根据在搜索引擎上所处位置收费的兴起而结束。使自己位于搜索引擎第一位的窍门广为流传后，由于位于同一搜索引擎的其他站点也使用了同样的技术，那么，任一站点都不可能永远位于搜索引擎的首位。诸如 Goto (www.goto.com) 之类的引擎已经开始根据站点所处的位置进行收费。优先位置权被标题广告使用以适应于特定的搜索短语。正如把公司名字命名为 AAA 可以使之位于电话号码的首位，但用户更加喜欢巨大的彩幅广告。搜索引擎将最终采纳同样的方式。而且，当用户成为网络高手时，他们将更多地依赖于目录来查询一般信息，除非在查询非常专业和复杂的信息时才使用搜索引擎。搜索引擎大战的最终结果必定是返回传统的信息提取模式，即像在其他广告中那样，你将为自己的位置付费。自此，设计者如果还考虑寻找窍门来使自己位于搜索引擎的靠前位置，就很愚蠢了。

7.5.1 添加到搜索引擎

让搜索引擎搜集站点的网页是使站点可以在网站上被发现的的第一步。实现这一步的最简单方法是告诉搜索引擎自己网站的存在。大多数引擎允许用户添加一个可被索引的 URL。例如，

Lycos通过一个简单的窗体允许用户添加 (<http://www.lycos.com/addasite.html>) 一个可以搜集的站点。当然, 添加站点到每一个搜索引擎是一项乏味的工作, 因此许多供应商 (<http://www.submit-it.com>) 渴望给开发人员提供一种批量的方法来帮助用户向搜索引擎提交站点。大多数网站提交软件如 WebPosition Gold (<http://www.Webposition.com>) 都包含自动提交功能。

一个比较大的问题是, 用户应该把自己的站点提交给多少个搜索引擎。一些用户仅喜欢把链接提供给十个比较大的搜索引擎, 尤其是 Yahoo!。大量研究以及作者的经验表明, 大的搜索网站尤其是 Yahoo对搜索引擎涉及的信息阻塞负有责任。但是一些网站改进专家认为这是不正确的, 他们认为最好是尽可能为站点创建更多的链接。事实上, 一类被称为 “ Free For All ” 的站点已经崛起, 开始为人们服务, 它认为 “ 所有用户都应与我链接 ”。事实是这些链接中的大多数恰恰没有用处, 并且经常产生大量没有价值的信息流和信息垃圾。再者, 如果你要恢复链接和 e_mail, 那么几乎是所有的人都在做相同的工作, 即恢复链接。

7.5.2 自动排斥

在把网站加到每一个搜索引擎之前, 应先考虑到自动搜索整个站点并不是一个好的注意, 除非这个引擎是内部引擎或公共搜索引擎。首先, 在 cgi-bin目录中的源程序是不能被检索的。其次, 许多网页可能是过渡的, 如果让用户通过引擎检索使用它们, 可能会发生错误。最后, 你可能不希望用户浏览所有网页, 尤其是网站底层的一些页面。所谓 “ 深层的链接 ” 可能会使从公共搜索引擎进入的用户产生混淆, 这是由于用户是从站点的底层开始的, 他们没有获得用来给访问者定向的主页或导入页面的信息。

搜索引擎和自动站点收集工具如离线浏览器所遇到的最大麻烦是, 他们可能被用来发动拒绝服务攻击。大多数搜索引擎的基本想法是先读网页然后尽可能快地跟踪网页。如果让一个搜索引擎尽可能快地浏览一个网站, 这样, 所有的请求都将很快, 它可能导致站点不能执行请求, 从而拒绝用户的合法访问。值得庆幸的是, 大多数用户是没有恶意的, 但是要知道当搜索引擎不断索引同一个动态生成的网页时, 这种情况可能会不经意地出现。

7.5.3 Robots.txt

为了采取有限制的自动访问, 采用了 “ 自动排斥 ” (Robot Exclusion) 协议。基本的想法是采用一个特殊的文件即 robot.txt, 此文件存储在网站的 root目录下。例如, 如果 spider正在索引 <http://www.democompany.com>, 它将首先在 <http://www.democompany.com>/找到文件 robots.txt, 在执行索引之前首先要分析这个文件。

注意 如果遇到类似于 <http://www.bigfakehostingvendor.com/~customer> 之类的站点, 你会发现许多 Spider将忽略 <http://www.bigfakehostingvendor.com/~customer/>下的 robots.txt 文件, 很遗憾, 这时必须请求供应商在 robots.txt中为你设置一个项。

Robots.txt的基本格式是特定 Spider或你想限制的用户代理的列表, 以及哪些目录路径不被接受的声明。例如:

用户代理：*

不接受：/cgi-bin/

不接受：/temp/

不接受：/archive/

在这个例子中，拒绝了对一些目录的自动读取操作，如 cgi-bin目录，temp目录，以及 archive目录，在这个目录里，我们可能要删除一些旧的但还必须在线的文件。必须认真检查你所编辑的robots.txt文件的内容。例如：

用户代理：*

不接受：/cgi-bin/

不接受：/images/

不接受：/subscribers-only/

不接受：/resellers.html

在这个文件中，subscribers-only和resellers文件不能被索引。然而，让人知道了它是敏感信息。例如，如果把某个目录隐藏起来不允许检索，除非用户通过 e_mail付款获得一个URL，你不应该把它写在robots.txt文件中。使用户知道隐藏文件或目录的存在是一个问题。不怀好意的访问者事实上很可能仔细阅读 robots.txt文件来分析不想让用户看见的那些内容。这很容易做到，仅仅在URL中输入http://www.companytolookat.com/robots.txt即可。

一定要意识到自动排斥标准，假定所有 spider程序将遵守它。当然，不怀好意的 spider可能忽略这个文件，当某人决定攻击你的网站时，你可能被迫在服务器上去堵塞某些特定的 IP地址和用户代理。

7.5.4 由<META>控制自动排斥

除robots.txt外，对那些不能访问他们域的根目录的用户来说，还可以使用 <META>标签来控制索引。不允许对某一页的索引，可以在 HTML文件的<HEAD>部分使用<META>标签：

```
<META NAME="robots" CONTENT="noindex">
```

同时你可以通知搜索引擎不要跟踪本页的链接：

```
<META NAME="robots" CONTENT="noindex, nofollow">
```

当使用这种类型的排斥时，记住不要把自动排斥与下面这样的抵触信息混淆，如：

```
<META NAME="robots" CONTENT="index, noindex">
```

或

```
<META NAME="robots" CONTENT="index, nofollow, follow ">
```

搜索引擎既可以忽略整个信息又可以任意索引。公共搜索引擎更多地支持 robots.txt而不是 <META>。

7.6 优化搜索引擎

优化搜索引擎并不是一件困难的工作。首先就是像创建搜索引擎一样思考问题，换句话说，

根本上就不思考。搜索引擎几乎就是观察网页本身，根据以往的经验并根据一定的规则判断网页的内容。这些规则包括单词的出现频率，<META>标签和许多其他内容。但是，搜索引擎不能准确地区别 Miami Dolphins 足球队网页和 Miami 网页中展示的 dolphin，因为它不清楚网页的具体意义。之所以这样是因为搜索引擎主要依赖于关键词匹配和一些有关单词在页面中的位置和站点链接数的直观推断。如果设计师知道搜索引擎正在查询哪些内容，那么他可以很容易地优化网页，使其在搜索引擎中处于较高的级别。以下部分将简要介绍搜索引擎需要查找的内容以及人们为了提高站点搜索级别而摸索的窍门。

7.6.1 <META>标签

许多搜索引擎都查看 <META>标签页中的关键字和描述。例如，如下 <META>标签可用于 robot butler 网页中：

```
<META NAME="Keywords" CONTENT=" Butler-1000, Robot butler, Robot butler specifications, where to buy a robot butler, Metallic Man Servant, Demo Company, robot, butler">
```

注意内容怎样首先从具体的关键词和短语开始，然后以一般的关键词结束。

这一点应该用于对搜索引擎的使用。

如果搜索引擎碰到 <META>标签，它将以 CONTENT 属性中关键词出现的频率来确定网站的优先权。基于此，一些设计者在 <META>标签页中频繁地重复使用关键词。

```
<META NAME="Keywords" CONTENT=" Robot butler, Robot butler, Robot butler, Robot butler, Robot butler, Robot butler, Robot butler, Robot butler, Robot butler, Robot butler">
```

然而，有些搜索引擎会认为此页仅仅是在导入关键词，而有可能从它们的索引中删除本页。如果关键词的重复不是那么明显，就变化它的组合，如下所示：

```
<META NAME="Keywords" CONTENT=" Robot butler, Butler-1000, Metallic Man Servant, Robot butler, Butler-1000, Metallic Man Servant, Robot butler, Butler-1000, Metallic Man Servant, Robot butler, Butler-1000, Metallic Man Servant ">
```

搜索引擎不会认为它是不合适的。一个更好的方法是使关键词的重复模式不像是仅仅改变其次序。

```
<META NAME="Keywords" CONTENT=" Butler-1000, Robot butler, Metallic Man Servant, Robot butler, Butler-1000, robot, Robot butler, Democompany, Metallic Man Servant, Butler-1000, robot, butler, Robot butler, Butler-1000">
```

但是，设计者还应该注意频繁地使用某一些单词和短语可能会降低网页的级别，甚至于被搜索引擎完全从索引中删除。

搜索引擎同时也查看 <META>标签的描述值。例如：

```
<META NAME="Description" CONTENT="The DemoCompany Robot Butler is
```

the most outstanding metallic man servant on the market. The Butler-1000 comes complete with multiple personalities and voice modules including the ever-popular faux-British accent.">

它将被包含在 robot butler 网页，并被引擎检索到，而且将作为搜索引擎的返回页。因为其中有很多用户可用的信息，通过这些信息用户可以决定是否访问这个网站。比较合适的情况是保留1到2句描述信息，至多不能超过3到4句。

7.6.2 标题和文件命名

搜索引擎确定网站级别的一个重要方面是标题。例如：

```
<TITLE>Robot Butler</TITLE>
```

就搜索引擎的排名来说，这是一个不好的标题的例子。相对来说比较好的标题是：

```
<TITLE>Butler-1000: Specification of Demo Company's Robot Butler,  
the leading metallic man servant on the market</TITLE>
```

请牢记用户通常会仔细看标题，他们把标题当做书签，比较长的标题是搜索引擎使用的，而不是用户使用的。

文件名对于搜索引擎同样重要。使用“ butler1000_robot_butler.htm ”作为文件名比使用“ butler.htm ”作为文件名更合适，因为前者可以体现你有一个好的域名和目录结构，并且可以创建一个同样有意义的 URL。例如，如果把 democompany.com 命名为 www. democompany.com。我们将有如下一个 URL：

```
http://democompany.com/products/robots/butler1000_robot_butler.htm
```

注意到它几乎包含了与标题相同的信息。这样可以使用户知道它们位于何处，而不是求助于如下 URL：

```
http://democompany.com/products.exe?prod=robots&mod=butler1000
```

7.6.3 相关文本内容

得到索引的有效途径是在网页内容中包括关键词和短语。许多搜索引擎浏览整个网页中的文本，尤其是处于网页顶部或标题标签中的文本（像 <H1>或<H2>）更受青睐。搜索引擎也格外重视超级链接中的文本。因此

```
<A HREF="specifications.htm">Specifications</A>
```

就不如

```
<A HREF="specifications.htm">Robot Butler Specification</A>
```

便于搜索引擎搜索。

搜索引擎在把注意力集中于网页内容时遇到的一个问题是，设计者经常在主页中包括大量的图形。如果搜索引擎只浏览 <META>标签和网页标题，将降低网页优先级。这个问题可以通

过首先使用标签的ALT属性提供的额外信息来解决。例如：

```
<IMG SRC="robot.gif" ALT="Butler-1000: Demo Company's industry  
leading robot butler">
```

当然，把实际的文本置于网页之中更好。一些设计者把文本写得非常小，并且把它的颜色设为与底色相似，这样用户将看不见它们，但是搜索引擎却可以看到，可谓一举两得。例如：

```
<FONT SIZE="1" COLOR="white">The Demo Company Butler1000 is the  
best robot butler. The Democompany Butler1000 is the best robot  
butler. The Democompany Butler1000 is the best robot butler.</FONT>
```

注意 当使用极小或不可见的文本时，一些搜索引擎可能会把这些网页从搜索引擎中清除。

7.6.4 链接和切入点

搜索引擎确定优先权的另一个根据是链接的数量（包括进入页和退出页的所有链接）。路标页面，比如主页一般都有许多进入和退出链接。搜索引擎趋向于把路标网页的级别定得比较高，因此站点中的主要网页必须有到其他网页的链接。一些搜索引擎比较看中那些与许多站点相关的站点。因此，人们开始创建站点的目的仅仅是为了与其他站点链接。

另一个提高级别的方法是向网站甚至向引擎提交大量的网页。所有这些切入点都可以被称为门口页（doorway pages），它们指向网站的主要内容。遗憾的是，对于许多用户，入口网页更像诱骗页面，由于它们使用了错误的信息来吸引访问者，而最终导致用户抛弃他们实在不想看到的页面。

7.6.5 花招

搜索引擎专家使用的花招是多种多样的，并且总在不时改变。许多想法是非常简单的，仅仅是传统网页设计技巧的一点点缀。例如，许多设计者利用不可见的像素背景达到强制布局的目的。搜索引擎改善者会想为什么不在这些图像元素上添加 ALT属性以改善效果。

想像在网页中有如下代码，对于那些停留在这些不可见像素上的可怜的用户，它仅仅是弹出一个工具提示。

```
<IMG SRC="pixel.gif" ALT="robot butler robot butler robot butler">
```

网页中采用不可见的文字、小文本和多个图像，或仅仅加载<META>或<TITLE>标记，这并不是最复杂的花招，但却常常奏效。

其他花招包括声名狼藉的“诱饵和转移”（bait and switch）方法，它生成一个特殊的搜索引擎页面，然后提交到一个搜索引擎。一旦等级非常高时，诱饵页面将会被一个真正的为用户建立的页面替换。这种方法的一个更加复杂的版本被称为“喂狗”法（feeding the dogs），在“喂狗”的场景中，你写一个程序，当程序识别到搜索引擎点击到站点时，你可以把一个它希望看的网页“喂”给它。像一条贪婪的狗，它吞食着食物而丝毫没有想到它只是在吃信息猪鼻子的仿制品。当实际的用户点击站点时，他们没有被“喂食”，而是获得实际的站点。

区分搜索引擎和实际的用户并不是非常困难的事情，因为搜索引擎识别自己且来自固定的 IP 地址。实际上，“喂狗”的花招只是检测浏览器的一种修改版本。搜索引擎通常对这种方法无能为力，除非它们不考虑动态产生的网页；而考虑到它不断增长的重要性，那是不可能的；或者采用另外一种方法，不通知站点它们是搜索引擎。一些搜索引擎已经开始提供一个有关索引的链接页面，用户可以决定它们是否提供了不同于搜索引擎的东西。

所有的搜索引擎改善面临的问题是，它促使设计者停止为用户设计网页，而转向搜索引擎。这是另一种形式的“更多的为自己而不是为用户的需要设计”。

规则：不要一味地为吸引搜索引擎设计网页，毕竟网页最终是为用户设计的。

另一个有关搜索引擎的有趣的方面是，许多大的组织并不依靠它们增加流量。实际上，对许多公司，如果没有直接键入公司的名称，你就不得不用搜索引擎去发现它们。然而，尽管他们自己有些疏忽，这些站点继续获得大量的商机。通过诸如 Gvu 因特网调查组织的研究，人们经常直接在 URL 中输入。他们是怎样发现这些站点的？搜索引擎并不是唯一的增加流量的方式。有许多方法可以让用户访问你的站点。一个逐渐流行的方法是依靠因特网以外的手段吸引访问者，如电视、广播、印刷。布告板、直接投寄和广泛的其他集会正被用来传播最新的网站地址。这可能有些偏题，下面将非常简要地讨论一些用于增加流量而不同于搜索引擎的在线技术。

7.6.6 标题广告

在因特网上同样存在广告。标题广告被用在从个人网页到大流量网站的各种站点上。标题广告只是专门为一个特殊的站点或产品服务的小图像。标题广告的大小可以从一个微型按钮到一个完全大小的标题。因特网广告局 (The Internet Advertising Bureau) (www.iab.net) 提供了有关标题广告大小的规范说明。

利用标题广告进行推销涉及到两个主要的方面。第一，必须在合适的人面前展示合适的标题广告。然后，必须让用户点击他们。如果你运行自己的网站，你可以自己把握标题广告。然而，在大多数情况下，你希望从其他的地方转向你的站点。一个方法是同其他站点交换标题广告。有许多在线相互交换标题广告的程序像 LinkExchange (www.linkexchange.com)。但是，你会发现必须在特殊的站点上订购广告。可以直接和站点联系，但是同样有大的为标题广告付费的网络如 DoubleClick (www.doubleclick.com)，他们有能力在许多站点上进行广告业务，甚至可以将广告做到指定的领域如大学校园。

使用标题广告的第二方面是让人们点击标题广告，这是一件非常困难的事情。研究表明许多用户看上去患有“标题广告失明症”，他们对具有标题广告特性的事物视而不见。我们在第 6 章已经讨论过了，标题广告本身不过是一个图像链接。下面将快速介绍如何改进标题广告的一些基本知识。通过对点击率的研究，表明色彩和动画可以帮助吸引某些用户。文字同样也可以吸引用户，尤其是当文字看上去具有点击吸引力时。一些人发现在标题广告下放一个文字链接对提高点击率很有用 (见图 7-5)。



图7-5 在标题广告下旋转文字链接

同搜索引擎推销类似，一些人求助于将标题广告设计成具有 GUI界面的小件的方法吸引用户点击（见图7-6）。

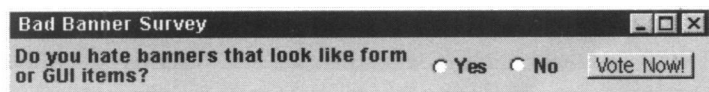


图7-6 具有GUI界面的标题广告

一些不道德的标题广告设计者甚至将标题广告设计成看上去非常像站点导航，如图7-7所示。对一个新手而言，要区分它到底是不是站点的一部分是异常困难的。



图7-7 与站点导航类似的标题广告

注意，许多正统的标题广告网络将不允许这样的标题广告，用户会因为受到欺骗而愤怒。

也许标题广告最大的问题是，在一个如此小形状的图形的标题广告中，能说多少东西呢？一个允许用户搜索、购物、甚至玩游戏的丰富的标题广告具有吸引力；但像搜索引擎一样，标题广告将不是唯一吸引访问者的途径。

其他在线推销技术

除了标题广告，有许多其他在线推销技术可以用于扩大知名度并增加流量。以下列举一些。这些不是全部的技术，只是用来启发设计者思考一些其他方法，而不是局限于使用户用搜索引擎发现你的站点。

1) 邮件（EMAIL）。有目标的邮件被认为是高效联系客户的方法。不幸的是，人们经常漫无目标地发送邮件。在没有得到收信人同意的情况下发送邮件被称为“骚扰”（spamming），骚扰信当然是令人愤怒和不合法的。采用选择（opt-in）信的方式可能更好些，信被发送到那些需要的人手中。但是必须确定受信人已经确认要求收到邮件。不幸的是，一些填了表的用户没有意识到他们已经被加入到要求收信的列表中。这将会导致以后的麻烦。

2) 新闻组提交（NEWSGROUP POSTING）。一些人发现将小的广告提交到新闻组是有帮助的。一般而言，采用非侵犯性形式如简单地将站点与新闻组相关联，是在新闻组网络系统中推销站点的更有效的方法。

3) 竞争（CONTESTS）。站点已经开始采用彩票或竞赛的形式争夺知名度。用户只有访问站点一定次数以后才有资格参加竞争。使用竞争方式的缺点是常常只能吸引一群渴望赢得某物的人。另一个麻烦的方面是它们常常受到法律约束。一些地区和国家不允许某些竞争，因此在采

用竞争方式前咨询一下在线竞争机构是一个好主意。

4) 免费赠送 (FREEBIES)。像竞争一样, 各种各样的赠送当然也能够帮助吸引用户。如果给用户一个机会, 当累加到一定点击数可以用作商品或服务的折扣时, 你会发现用户很愿意跳进这个圈套。想像一下吧, 在现实的世界中, 飞机常飞线路的免费赠送推销活动经常发挥作用。当然, 给用户一些折扣或小价钱物品如 T恤是不能和给用户实实在在的硬钞相比的。

5) 点击支付和会员制 (PAY FOR CLICKS AND AFFILIATE PROGRAMS)。一些站点已经尝试点击支付的模式。通过直接点击站点或通过点击其他站点间接访问你的站点, 从而支付用户一定的金额。一些站点甚至向那些指向它的站点支付, 而不管这些是怎么发生的。

会员制进一步延伸了点击支付的模型——不仅仅是支付结果, 甚至可以同其他人共享利润。理解会员制的最好方法是看一个例子。我们的例子是, DemoCompany公司考虑为其他站点销售他们的机器人建立一个会员制。成为会员的站点将提供网页和链接以推销 DemoCompany的机器人。然而, 实际的交易本身将在DemoCompany的服务器上进行。在会员制中, 成为会员的站点可以获得交易的一部分赢利。会员制已经在许多大的公司如 Amazon得到应用, 同时一些会员制网络像 Be Free(www.befree.com)也积极推进会员制的发展和服务。也许会员制最显著的一个缺陷是, 一些用户认为会员制和多级市场模式有着相同的感受, 尤其是当会员制以非专业方式运行时。

6) 好的域名 (GOOD DOMAIN NAMES)。设计者常犯的一个大错误是当站点投入市场时, 低估了他们的站点名字。不要低估一个简单而难忘或众所周知域名的威力。有多少出名的站点的名字像 123-travel-for-less.com, www.thisisridiculouslylongdomain.com, 或更加糟糕的是, 站点不属于通常的域空间如 .com; 注意, 绝大多数著名的网站的域名不超过 10 个字母, 有的站点仅有 3 到 6 个字母。域名要容易记忆和拼写, 如 Amazon, 要非常普通, 不要采用难于发音和拼写的域名。采用一个特殊产品或服务的最普通域名是十分有用的。例如, 用户在寻找书的时候只是简单地猜测域名是 book.com 的形式。

7) 以前的知识 (PREVIOUS KNOWLEDGE)。事实表明, 许多用户经常访问以前知道的网站。当你问一些用户关于因特网上有关书的站点时, 他们很可能提到 Amazon, 但更多的人在日常生活中听说过 Barnes 和 Noble。注意网站上许多流行的东西和现实世界对应物有着紧密的联系, 这些对应物通常或者同网站的属性相同, 或者促进 Web 属性的推广。主要的电视网络系统和印刷集团已经开拓了许多网站渠道。因为因特网是开放的, 不要认为它一下子就可以改变传统的市场。

通常不能判断用户是如何找到一个站点的。许多因素影响用户到达一个指定的网站。本书仅仅涉及到了这些主题的一些皮毛, 介绍的全局搜索部分也只是站点交易的一小部分。有兴趣的读者可以访问 <http://www.wilsonWeb.com/>

以获得站点推销的更多信息。

7.7 对本地搜索的需要

一旦用户访问到你的网站, 他们将发现本地搜索对浏览本地内容十分有帮助。然而, 并不是所有的网站都需要一个内部的搜索引擎。网站数据量的大小、数据的类型都对是否需要一个搜索引擎有影响。如果一个网站只有一打左右的页面, 它将发现搜索引擎的帮助微乎其微, 还

不如使用严格的浏览方式和导航帮助如站点映像（将在第 8 章讨论）。但是，如果一个站点的页面超过 100，就几乎总是需要一个搜索引擎。不考虑站点的大小，如果站点具有复杂的数据类型，尤其当数据结构十分规则时，搜索引擎将十分有用。像编号、产品分类和其他结构化数据如住址列表等都可以很好地用于搜索。

规则：如果一个网站有规则的结构化数据、非常难以理解的数据或超过 100 个页面，就应该包含一个本地搜索引擎。

除了数据类型，访问者类型和对站点访问的目的同样影响对搜索引擎的需求。如果一个网站为了迎合老的用户，就应该包含一个搜索引擎。老用户通常使用搜索引擎以避免启发式的浏览模式。如果一个站点有着显著规模的回头客或一定数量的经常访问者，那么建立一个搜索引擎将有意义。一个老用户已经知道某件产品在什么地方，为什么还要强制他们浏览整个网站。老用户已经知道寻找的产品的编号，他们只希望在搜索区中输入编号，然后直接到响应的页面。

规则：如果一个站点要迎合超级用户或经常的回头客，那么就应该提供一个搜索机制。

7.8 添加搜索工具的过程

以下的八个步骤总结了将搜索工具添加到网站的过程：

第一步：决定什么将被索引。

你是希望整个文档还是希望其中的一部分被索引？在绝大多数情况下，需要索引的是文档的一部分和技术上支持的数据库，或者是用户希望被搜索的内容。千万不要索引你能够索引的任何东西。

第二步：决定怎样索引信息。

一旦决定了你将索引的内容，必须决定如何索引他们。是让搜索引擎为整个文档的所有单词建立索引集合，还是为网站的某些特殊网页建立一个特殊搜索词汇字典和关联词汇字典。

第三步：选择搜索引擎。

一旦你想出信息的规模和类型以及他们的索引方式，再选择合适的搜索引擎是非常重要的。有许多搜索引擎可供利用，有免费的和商业的。搜索引擎可以安置在本地，也可以利用第三方的搜索引擎服务。有关搜索引擎和提供的服务，请参见 <http://www.searchtools.com>。

第四步：设计搜索界面。

设计搜索界面是为了解释用户搜索的类型。通常，搜索被分为基本的和高级的搜索形式。搜索界面应该同站点一致，适合用户搜索的需要及搜索数据的类型。

第五步：设计结果页面。

确定一旦搜索成功返回，页面要处理正确的结果；同时，当搜索失败时，要对负面结果进行一些处理。

第六步：索引数据。

在这一步中，搜索引擎将遍历站点的所有部分，同时建立一个索引。可以通过手工强制优

化索引结果。

第七步：集成搜索引擎和搜索界面。

这一步使搜索界面可以访问索引。一般地，只要设置搜索窗体的 <FORM>标记的 ACTION 属性即可。对搜索结果页面的处理要困难些，但可以获得搜索结果页面，并将搜索结果页面转换成搜索引擎可以识别的特殊模板形式。

第八步：测试和监视。

实现一个搜索引擎的关键点是要确保对重要的查询返回正确的结果。搜索引擎应该被监视，并且通常的查询应能被识别。用户应能够被允许对个人搜索结果分级，以利于进一步提炼。

下面的篇幅不涉及到如何实际地建立索引（建立索引的方法依赖于被索引的数据以及被使用的搜索引擎），而是如何设计搜索界面。

7.9 设计搜索界面

一旦确定要建立一个搜索设施，设计者先要考虑的是用户要搜索什么？通常，搜索引擎被添加到一个站点，用户使用完全的文本搜索对所有的内容建立索引。就全球范围的搜索而言，当用户为了搜索编号为 KF-456 的文档，而得到从会议纪要到技术报告几乎所有含有相应编号的文档时，用户会不停的摆脑袋；对用户来说，文档的排序也几乎是任意的，重要的文档不在列表的首位。令人感兴趣的是，为什么要使用这种搜索方式。设计者假定，既然公共搜索引擎尚且如此，那么他们的本地搜索引擎也应该这样。这看上去是个好主意：既然用户对公共站点的格式化搜索字符串熟悉，那么他同样可以将这些知识用到站点上来。但是，由于各种原因，搜索引擎并不是十分的精确，包括一些极力想获得前位的网站。公共搜索引擎的结果也不是总有意义的，有时在某些情况下完全靠运气而不是靠精确。考虑你自己的网站，当用户输入“Robot Butler”时，如果希望显示一个特殊的网页，你就可以做到。记住，先前也提到过，当建立一个本地搜索引擎时，可以模仿公共搜索引擎的风格、语法和界面，但要避免模仿它们不精确的功能。

使用本地搜索的主要优势是可以利用受控制的词汇表处理用户可能希望的搜索。除了使用更加精确的方法把关键字同网页相关外，甚至还可以建议用户运行普通的查询。记住，本地搜索引擎给设计者提供了比公共搜索引擎更加强大的控制。

7.9.1 访问搜索

一个大的问题是用户怎样使用搜索工具。一些站点建立一个标记为“搜索”的特殊按钮，当用户点击按钮时，可以引导用户到一个特殊的搜索页面上去。其他站点则在所有的页面上放置一个搜索区。两种方法的比较见图 7-8。

尽管将搜索直接放在页面上可以避免用户的一次点击，在一个具有内容的页面上放置一个搜索区是非常基本的。当需要复杂的查询时，仍然有必要建立一个特殊的搜索页面。在每一个页面上放置一个高级搜索工具是不可能的，因为使搜索工更为突出并使它同提交内容的网页相分离。因此存在的问题是，是把简单的搜索工具放在内容页面上，还是放在一个特殊的搜索页面上。不论采用哪种方法，都应该能从网站的每一个页面上找到搜索。

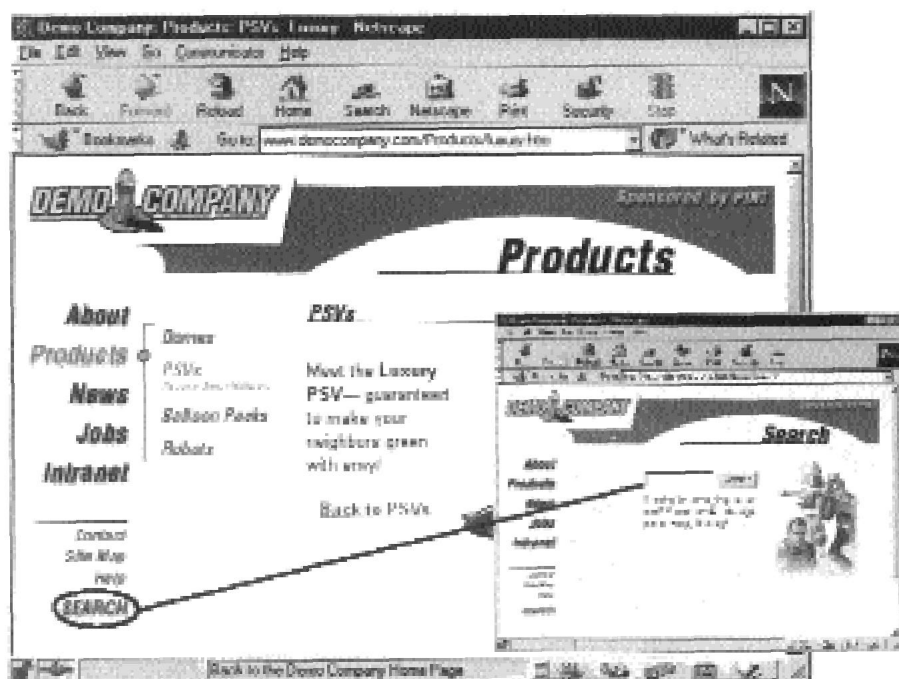
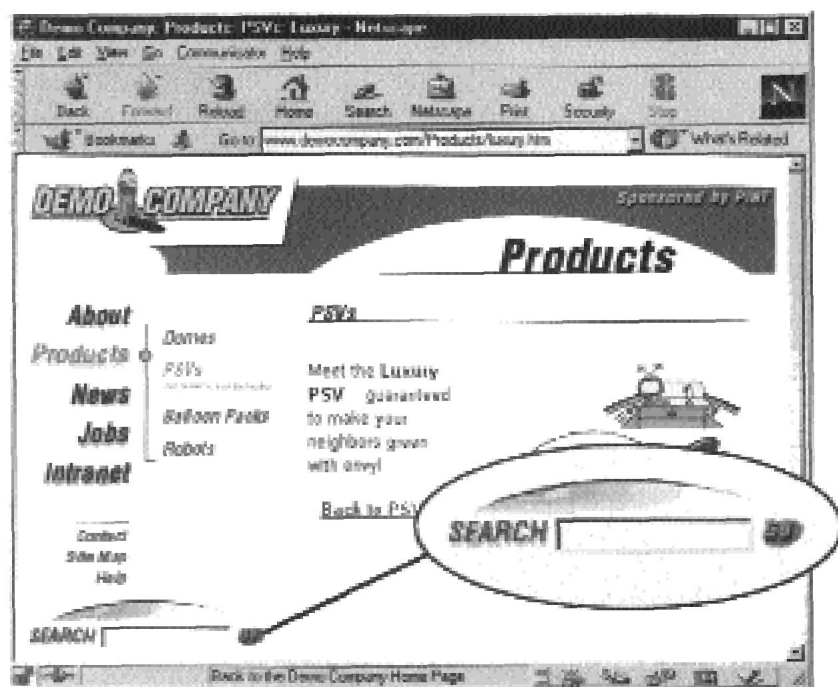


图7-8 页面内的搜索与专用的搜索页面

建议：当在站点上使用搜索时，请在所有的页面上包含一个搜索按钮或搜索区。

7.9.2 设计一个基本的搜索界面

站点的搜索工具应同网站的其他页面保持一致的界面。通常两者不同，是因为搜索工具是由那些采用同其他部分外观和风格不同的搜索模板的技术人员添加的；如果这种改变过大，使用这种搜索的用户就会感觉他们是不是已离开了原来的网站。观察图 7-9，你会感觉到保持一致性是多么的必要。

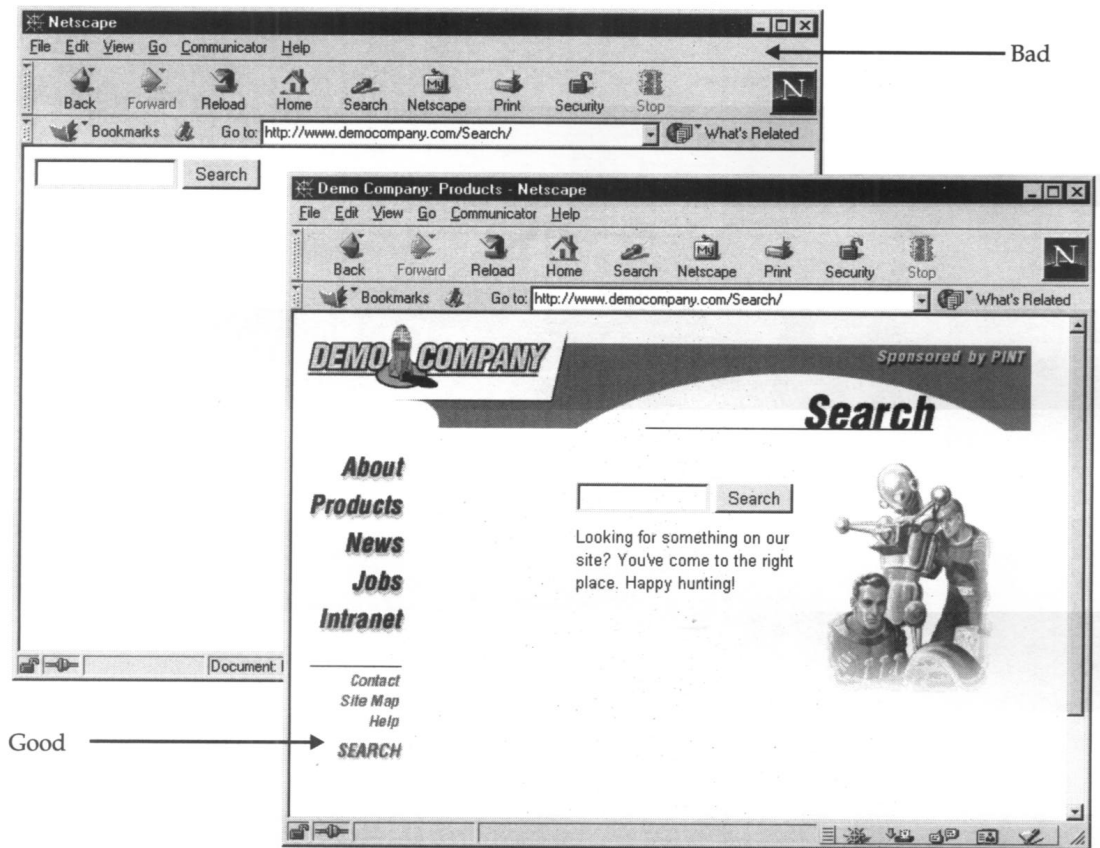


图7-9 搜索页面应该类似于站点中的其他页面

规则：搜索窗体和搜索结果页面应同站点的外观和感觉保持一致。

同样，搜索窗体应符合被搜索的数据类型。例如，如果用户正在搜索一个有色的对象，难道搜索窗体不应该提供描述颜色的手段？图 7-10中搜索个人太空交通工具的例子显示了搜索窗体怎样符合被搜索的内容。

考虑为一个网站设计搜索工具的典型提示，我们对用户的查找对象了解得越多，就越有能力帮助他们找到所需的信息。一种达到此目的的方法是分析用户输入。不管我们怎样估计用户在搜

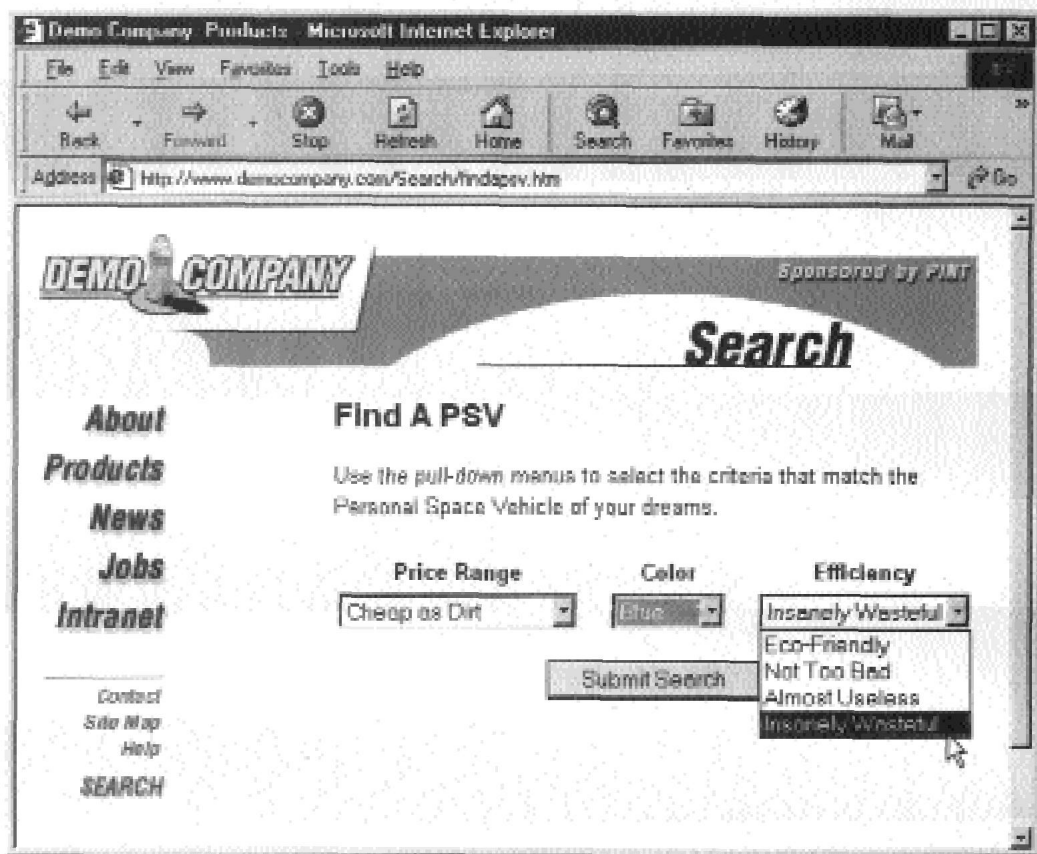


图7-10 随着搜索内容变化的搜索窗体

索什么，目的是帮助用户正确地缩小查询的范围。例如，如果我们搜索名字，就应该帮助用户在不同的文本框中输入第一个和最后一个名字，而不是在一个文本框中输入所有的名字部分。如果搜索1到10000间的一个编号，应该让用户知道输入的范围，将他们限制在范围内，同时对超出范围的提出警告。可以利用HTML中的TITLE属性或简单的JavaScript程序建立工具提示，而不是将提示直接写在屏幕上帮助用户知道输入的范围。一些适合搜索数据的搜索窗体显示如图7-11所示：

Employee Lookup Last name: <input type="text" value="Basinger"/> First name: <input type="text" value="Jon"/> <input type="button" value="SEARCH"/>	Part Number: <input type="text"/> <input type="button" value="SEARCH"/> <small>Enter part in range 1-9999</small>
Reverse Phone Number Search <input type="text" value="(345)"/> <input type="text" value="334"/> - <input type="text" value="1968"/> <input type="button" value="FIND"/>	

图7-11 搜索窗体

规则：搜索窗体应该同搜索的内容相匹配。

搜索窗体中主要的元素是搜索查询区。一个问题是搜索查询区应该有多长？一个搜索查询区应该至少可以容下一些搜索词汇而不用滚动。搜索查询区应可以提示建议查询的重点。如果查询是页面的主要重点并且用户将执行复杂查询，那么输入大小在 30到40 间是十分普通的，对公共搜索引擎的调查表明，在主查询中，查询区绝大多数使用 30、35或40大小。这个大小使得搜索查询区成为典型页面上横向宽度巨大的元素。当搜索仅是一个页面的第二主题时，搜索查询区的大小最好减半，即使用 15到18个字符，这恰好可以适合一些关键字的简单查询。当然，搜索框大小的设计应该时刻考虑到搜索词汇和页面的布局。

规则：主搜索文本框的大小最好是辅助搜索文本框的两倍。

另一个要考虑的是搜索窗体执行搜索的按钮。有时可以使用一个窗体按钮。有时可以使用用户按钮。用户往往凭直觉使用窗体按钮。按钮的标签也是各式各样的。有些标记为 Search，另一些则标记为 Find，有的甚至标记为 Go。采用什么方式依靠搜索的上下文。如果用 search标记区，那么再用 search标记按钮就显得多余（见图 7-12）。



图7-12 有标记按钮的标记区

搜索窗体应适合使用搜索的用户类型。例如，一个为儿童设计的搜索工具将有娱乐性，并且几乎没有什么指导；为工程师设计的搜索工具可能就会有許多文本区以利于他们调整搜索。简单的搜索窗体应同高级的搜索窗体区分开来。

7.10 高级搜索窗体的设计

高级的搜索窗体设计对设计人员来说是一个挑战，尤其是在有许多方法可供用户搜索的时候。首先，如果搜索允许使用 AND、OR和NOT运算符做布尔查询，搜索窗体要么设计为下拉方式以区分搜索词汇，要么提供对用户的详细指导，帮助他们建立布尔查询，如图 7-13所示。

对许多用户来说，建立布尔表达式是一件痛苦的事情。在基本的查询中要尽量避免使用。

高级的搜索窗体经常包含各种区以限制搜索的时间。例如，当对时间敏感的信息被搜索时，应能够限制日期范围。同样，也可以通过格式如图像、PDF和声音等来限制搜索数据的类型。一些搜索工具允许用户搜索某个站点的一部分，称为有界搜索（scoped Search）。不幸的是，用户可能不了解网站的构成，因此，将站点限制在一个搜索主题和分类上而不是整个站点上要好多。使用有界搜索方式最通常的做法是利用下拉框（见图 7-14）。



图7-14 利用下拉框进行有界搜索

建议：最好将搜索限制在一个主题、分类或观点上，而不是整个站点内容上。

使用高级的搜索窗体可能还允许用户限制返回的结果数量，设定结果返回方式和查询图元

信息如文档作者。图 7-15 是使用高级的搜索窗体的例子。

高级搜索窗体的一个非常重要的部分是要有指导。并不是所有的搜索引擎都采用同样的工作方式，必须直接在屏幕上或使用弹出式窗口详细地指导用户使用查询。不要将操作指导放在分离的页面上，这样可以避免用户将操作指导打印出来或迅速地记忆。除了使用操作指导外，还可以在页面上提供查询样例和区的使用。

robotic dog AND Astro SEARCH

robotic +dog -astro Advanced Search

Required Words
To include a word in all search results, add a plus sign (+) before the word.

Excluded Words
To exclude a word from all search results, add a minus sign (-) before the word.

robotic AND dog AND NOT Astro Advanced Search

AND
To include a word in all search results, use **AND** before the word.

AND NOT
To exclude a word from all search results, use **AND NOT** before the word.

OR
To include at least one but not all of a group of words, **OR** between the words.

图7-13 布尔查询

规则：高级的搜索工具应提供操作指导和样例。

7.11 结果页面设计

设计结果页面必须考虑两个极端的情况：没有结果以及结果信息太多。甚至当只有正确的信息被返回时，一个设计很好的结果页面也应该帮助用户区别什么是相关的。结果页面的首要原则是当人们不能确定某个结果的价值时，信息越多越好。设计很好的结果页面应该包含表 7-1 中的项目。

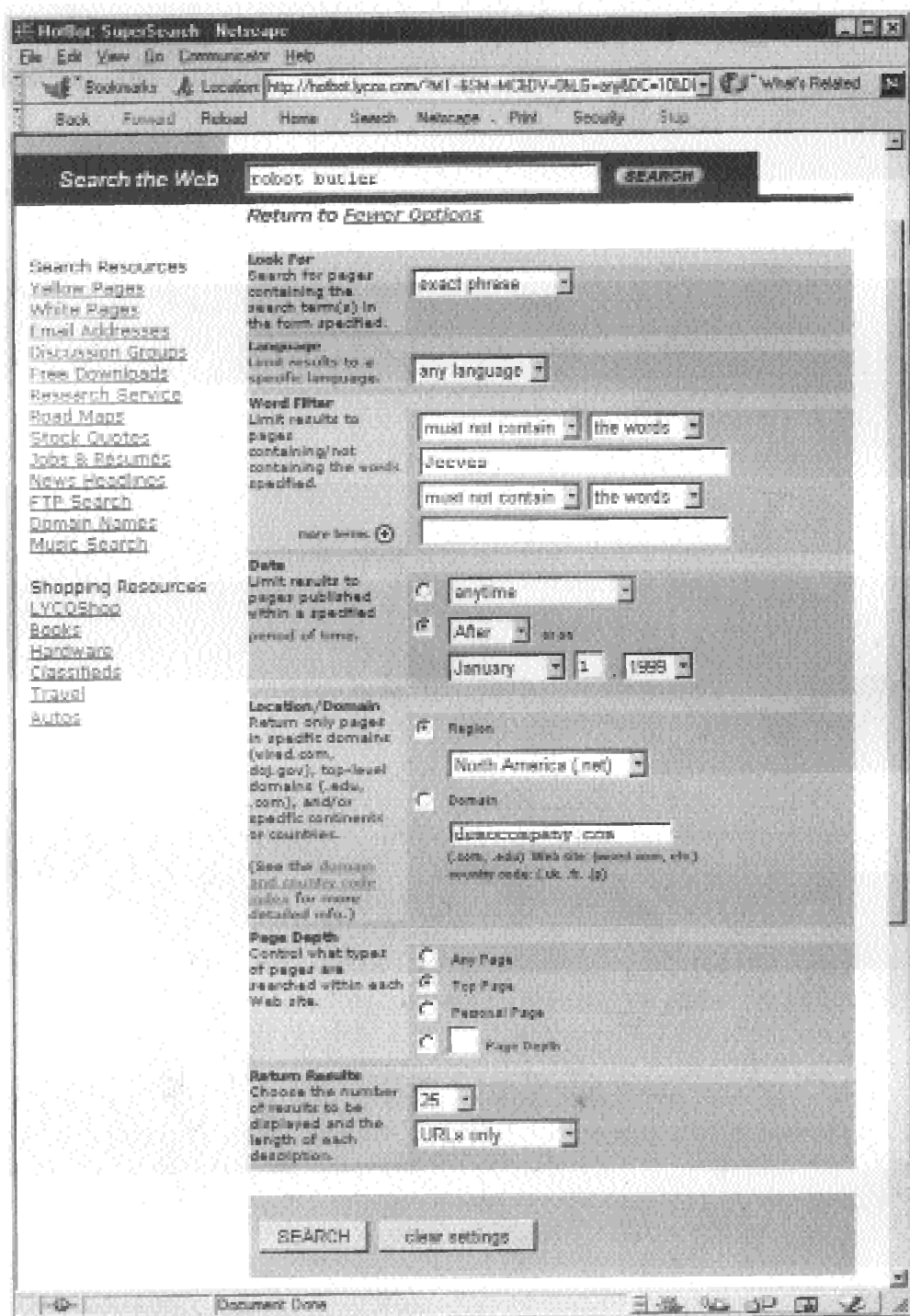


图7-15 高级搜索窗体必须仔细设计

表7-1 一般的结果页面元素

结果页面元素	注 释
原始查询	所使用的原始查询字符串应该突出显示在所有结果页面上，这样用户就不必回忆他们所使用的搜索字符串
搜索范围及找到的结果	应该指示搜索和返回的文档总数（例如，搜索了 10 000 个文档，有 20 个匹配结果）
当前结果的内容	应该有些指示，指出用户正在查看的是结果的哪一部分（例如，10 页中的第 2 页，200 个条目中的第 30~40 个）
页面或文档的标题	每个返回的条目都应该清楚地加上标题
返回页面的 URL	应该显示单个文档真正的 URL，因为可能给用户提供有用的信息
页面摘要	应该显示返回页面内容的简单摘要。这通常是从 <META NAME=" DESCRIPTION "> 或从文档文本的头几行中提取。用户可以选择显示还是隐藏页面描述
结果信息的日期或时间	至少应该显示返回页面的创建或最近修改的日期。有些搜索工具还指出索引的最近建立时间、搜索索引所花费的时间，以及查询的执行时间
返回页面的大小	应该指出返回的文档的文件大小。如果搜索的是很大的二进制文件，这一点尤为重要
结果类别	在某些搜索中，可能返回其他形式的数据，如 Adobe Acrobat、Microsoft Word 或多媒体数据。一定要用标签或图标指出数据格式
结果的相关性	应该清楚地指出相关程度。通常，搜索结果是从高到低排列的。应该使用百分数或横条显示条目之间的差别
关键词匹配	对于一个查询，当用户不能了解为什么返回某个特定页面时，他们可能非常恼怒，因此请显示匹配的关键词，如果可能的话，在摘要的上下文中突出这些词。如果可能的话，当用户选择一个文档时，查询条件应该清楚地高亮度显示
导航	应该提供在结果集中的移动导航。通常的按钮有“下 10 个文档”或“前 10 个文档”，其中步长取决于用户的爱好。有时也在结果集中使用第一页或最后一页来导航
提炼选择	应该提供提炼查询的能力。用户可能再次搜索结果集，甚至执行一个全新的搜索
帮助	解释结果格式的帮助信息应该是可用的

依赖于所使用的搜索引擎，这些项目并非都是可行的，特别是高级的相关性与匹配指示。然后，设计者应该尽量在结果页面中包含所有这些元素。

规则：结果页面应该提供尽可能多的信息，以使用户能够决定更进一步地阅读哪些条目。

搜索结果页面经常忽略任何站点导航方面的考虑。当用户访问结果页面时，他们不仅搜索——还可能切换到浏览模式查询结果。记住：用户是正在寻找某个答案，在寻找中，他们经常采用各种方法；因此要尽可能地在搜索结果中提供浏览设施，以防用户想要轻易地离开结果页面。图 7-16 呈现了一个结果页面，其中包含了列于表 7-1 的大多数元素。

搜索结果页面一个很明显但经常被忽略的方面是，必须认真考虑返回数据的格式。例如，只列举页面标题、URL 和描述，对于用户来说，可能并不足以做出选择。又如，如果用户执行一个产品搜索，输出的可能是如图 7-17 所示的满足用户标准的产品概述。

规则：搜索结果的格式应该适应返回的数据。

设计一个正面的搜索结果页面的关键点在于帮助用户找到并决定进一步阅读哪些条目。然而，与公共搜索引擎经常返回太多的结果不同的是，设计者必须认真考虑没有任何满足用户搜索标准的结果返回时的否定结果。

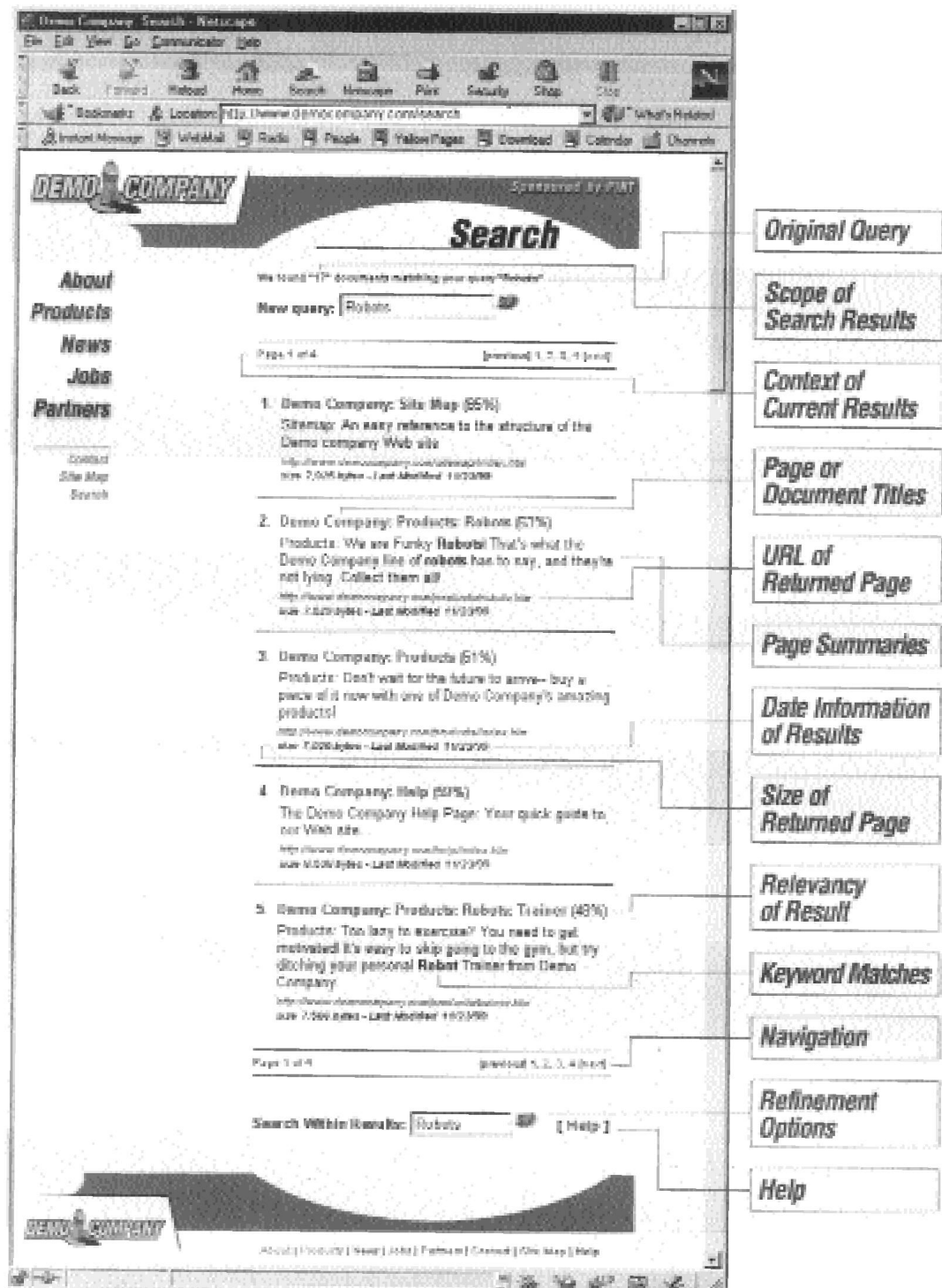


图7-16 搜索结果示例

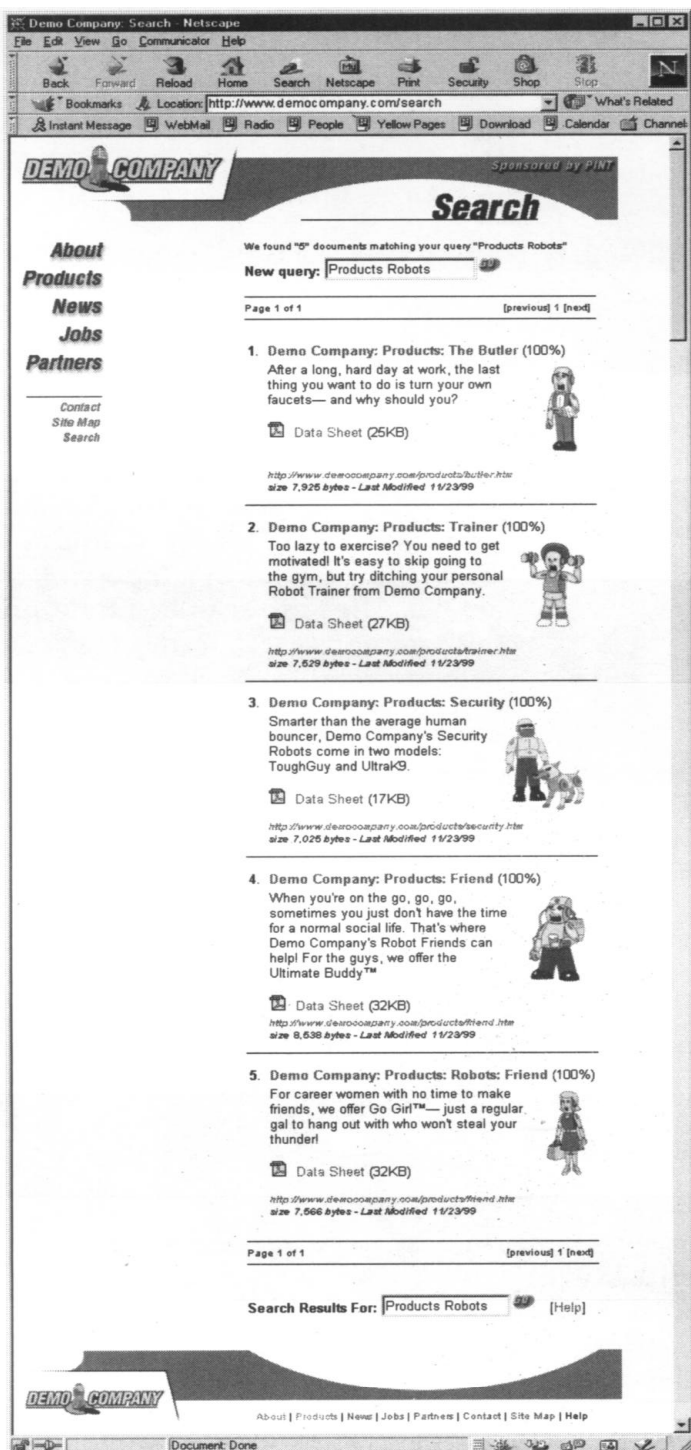


图7-17 根据搜索到的数据改变搜索结果

不定的结果页面

当没有任何匹配的查询结果时，结果页面应该努力帮助用户找到出错的地方。在某些情况下，可能就是没有任何满足条件的结果。一个好的否定结果页面应该尝试着帮助用户确定处于何种情况，并执行表 7-2 陈列的功能。

图 7-18 显示了一个否定的搜索结果页面，该页面提供了所有的能够帮助用户返回原路的功能。注意否定结果页面还需符合站点设计的要求。

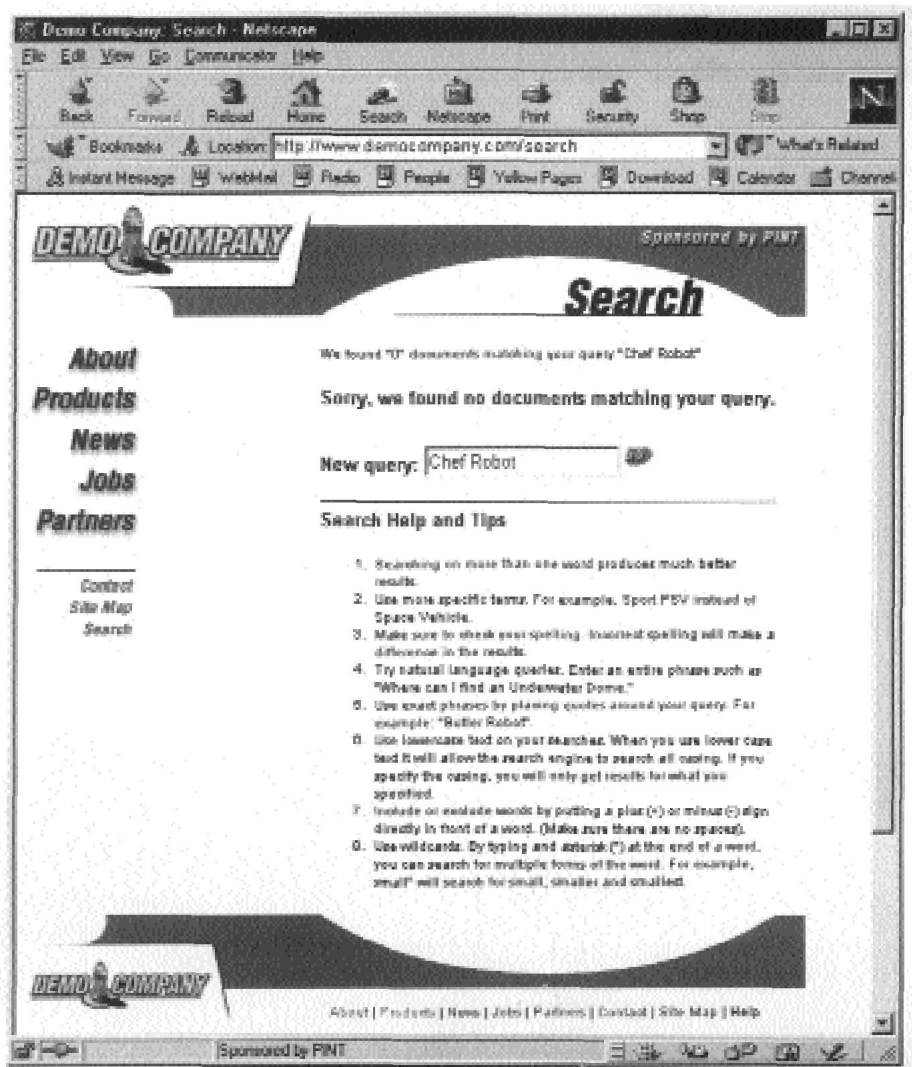


图 7-18 否定结果页面应该提供大量的帮助信息

规则：否定结果页面必须包括关于查询为什么失败以及潜在的修改该查询的信息。

表7-2 否定结果页面的特征

特 征	注 释
明确的失败消息	一定要让用户知道查询失败，以及可能的失败原因。指出所搜索的文档数目，并提供明确的消息指出失败的查询
再次搜索机制	与正确搜索结果页面类似，应该显示所使用的查询，并且应该在结果页面上提供再次搜索的选择。
帮助信息	否定结果页面最重要的方面可能是提供明确的有用的帮助。首先提供搜索为什么失败的提示，例如经常出现的搜索条件拼错。如果搜索引擎不能提供拼写检查，那么请考虑添加一个检查查询字符串的选项。如果可能的话，请显示接近的搜索条件。最后，确保关于如何使用搜索设施的帮助信息是可用的

类似于无效链接的页面，否定结果页面遇到的问题可能比我们想像的要多。务必监视否定的查询以便决定搜索设施的有效性。衡量否定查询的比例，并试着辨认一般无效的查询。如果站点遗漏了什么，否定查询可能显示出用户正在寻找的条目。

7.12 小结

虽然搜索设施主要是满足高级用户和频繁访问的用户，但初学者已经熟悉了公共搜索引擎。理解公共搜索引擎的工作及使用是设计本地站点搜索工具的第一步。设计者还应当理解用户是如何从公共搜索站点转移到本地站点的，并尽量将用户引导到他们所要查找的地方。搜索工具必须从用户的角度设计。最好的办法是考虑用户在站点中实际要搜索的是什么。不要落入盲目模仿全球搜索引擎自由文本搜索的陷阱。提供本地搜索时，务必提供基本的和高级的搜索窗体。认真地格式化搜索窗体并提供指导，这将帮助用户形成好的查询。万一事情出错了，请确保否定结果页面提供了额外的帮助以便让用户原路返回。一旦用户从搜索引擎中接受到正确的结果，就要确保提供足够的信息，以便他们能够逐渐缩小可能的选择范围。拥有太多的信息几乎跟根本没有任何信息一样糟糕。然而，请牢记搜索并不意味着一切。不要被公共搜索引擎的结果所困扰，不要假定用户将愿意利用本地搜索工具。像导航形式一样，搜索只是达到目标的一种手段，并不是目标本身。有许多方式帮助用户找到他们所要查找的东西。下一章将介绍其他的导航帮助，如站点映像、站点索引和帮助系统等。