

电子商务员基础知识：常用检索工具（2）电子商务考试 PDF
转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/514/2021_2022__E7_94_B5_E5_AD_90_E5_95_86_E5_c40_514733.htm 3.主题指南类检索软件

（1）Infoseek 检索系统 Infoseek 以前曾经是Netscape的缺省搜索引擎。1995年由Infoseek公司开发，它是第一个提供有偿服务的网络信息检索系统。Infoseek具备高级检索的功能。与其他检索系统不同，它不提供布尔逻辑式检索，而是使用特殊的符号。如果检索词是一组短语，而文件中也出现了这一短语，那么检索软件会认为这一文件的相关性评分最高。如果短语中有3个关键词，那么第一次检索并不把它作为词组对待；如果重复检索，则将这3个关键词作为词组对待。使用Infoseek 查询时，姓名和专有名词要大写，如 Steve Jobs、December、Star Wars等；在姓和名之间要用逗号“，”，如Clinton，Chelsea；用双引号"" ""来查找连在一起的词，但不能用单引号；用连接号“-”检索两个连在一起的词，如Cable - networks，用“-”连接的两个词表明这两个词必须同时出现；用“[]”检索在100词之内出现这两个词的文件，如[Web search]；用“（）”来检索彼此靠近、但可以任何顺序排列的词；要求或排除操作符与Alta Vista相同；“”号和“-”号的用法也与Alta Vista相同。Infoseek检索对大小写敏感，大写的单词被当成正确的名词来查找，相邻的大写单词被认为是一个词组。大写词组须用逗号分开，如：The Great Bambino，Baseball Hall Of Fame。Infoseek检索的优点在于它的速度快和使用方便。它既是搜索引擎，又是可检索的主题目录，可以检索UseNet NewsGroups、E-mail地址和Web的FAQ。由于软件

排除数据库中重复的URL，每一个URL只标引一次，因此在Infoseek上检索到的重复文件相对比较少，它特别适合缺乏检索经验和训练的用户。它的缺点在于数据库规模小，数据库综合性不强，而且缺乏高级功能，不支持布尔检索。

(2) Yahoo! (Yet Another Hierarchically Officious Oracle) Yahoo! 不是搜索引擎，而是严格的层次组织的主题索引。它已经开发了很长时间，有很多编辑人员来维护，所以质量非常高。当不知该去哪里的时候，在Yahoo! 上浏览是找到好站点的最好方法。它也是找到好的"初学者"站点的最好方法，从那里可以连到更专门的站点上。为了方便中国客户，目前，Yahoo! 网站新开设了中文版本（包括简体和繁体版本）。Yahoo! 是一个极好的联机商业机会搜索资源，它提供简单检索和细节检索。前者主要检索一级目录，后者可使用布尔逻辑式检索。Yahoo! 检索软件的一个独到之处是用户可以利用它的检索工具以关键词的方式查询它的目录，它的检索软件是由Open Text公司提供的，因此在检索时，一方面检索Yahoo! 的主题目录，另一方面检索Open Text提供的收有100万Web文件的Open Text数据库。Yahoo! 的使用很简单。只要输入查找单词，单击按钮即可。它将返回三种信息：满足查询条件的Yahoo! 的目录、满足条件的实际站点和其他检索工具。用户可以根据自己的需要确定搜索的范围是Yahoo!（缺省），Usenet或Email Address；选择搜索词之间的关系：OR还是AND（缺省）；决定是进行子串搜索（如输入head，可以查headlines），还是进行完整的单词搜索（如输入headlines才搜索headlines）。缺省是子串搜索。用户还可以控制每页显示的结果数目：10，25（缺省），50或100。

Yahoo! 的优点是速度较快，通过主题巡视查准率高；缺点是查全率低，相关性排序质量一般

4. 中文检索类软件

(1) 中文搜索引擎的特点

内码：由于历史与政治的原因，目前两岸三地在中文语言的使用上逐渐有了较大的差别，体现在计算机处理上也有很大的不同，其中最重要的区别是采用不同的字符集及内码体系，大陆用的是GB码，而台湾则用的是BIG5码，字符集的大小也不尽相同。作为一个要涵盖网上中文信息的中文搜索引擎就不能不正视这个问题，想办法去解决这个问题，如在搜索引擎内部将其统一在一个内码体系中（如UNICODE）或者是一个大字符集中（如GBK）等。

分词：英文单词的分隔相当清晰，用空格分隔。而中文的字词则不能简单地进行分隔，因此有不同的处理方法。一个是完全单汉字全文检索，即将文章中的每一个汉字都做索引，而将用户的检索提问根据单汉字匹配的原则去进行检索。此法虽然能够保证有很高的查全率，但却经常返回一些杂乱无章的检索结果。所以说，这是较容易造成答非所问的处理方法。除此以外的方法则是根据一定的原则和方法来对文章进行自动分词，然后按词建库，对用户的检索结果按词汇匹配来进行查询。一般来讲，这种处理方法较容易返回符合用户检索词概念意义的检索结果，但是分词技术的好坏直接影响到查全率。从目前中文搜索引擎实际情况来看，绝大多数的搜索引擎采用分词的处理方法，因为这样除了较准确之外还能提高整个系统的检索响应效率。中文分词的另一层意义则是指对于用户输入的查询语句进行自动分词，然后用分解后的词汇来进行检索，这种做法降低了用户在检索式组织上的工作量，提高了查全率和查准率。

分类：除了专门的分类搜

索引引擎（如Yahoo！等）之外，大量的关键词搜索引擎也采用了以分类目录进行导航的方法。这一现象在中文分类搜索引擎中更为普遍。但是由于文化的差异，中文搜索引擎分类类目具有与国外搜索引擎分类类目不同的特点。（2）网络语言翻译问题 在Internet中，英文在信息检索中得到了最广泛的应用，已成为因特网检索中的通用语言。熟悉英文检索体系和方法是顺利开展电子商务信息收集的基本条件。而就我国大部分用户来说，英语都不是很好，对所得到的信息不能很好地理解，有时甚至会产生误解，因此，使用字典和翻译程序来帮助自己把所得到的信息翻译成中文就显得很有必要。现在有很多字典和翻译程序来帮助理解不同国家的语言，而且不少软件能够实现所谓全屏幕英汉翻译。但是现在所能够见到的全屏幕英汉翻译软件并不能真正达到实用的翻译水平，而只能解决词汇的不精确的转换，它们对于语法和专业的问题要打很多折扣。有一些网站提供了在线的字典，可以直接上网查询，比如像

<http://rivendel.com/~ric/resources/dictionary.html>，包括有汉语、朝鲜语、日语等语言在内的100多种语言的在线对译字典。电子字典始创于台湾，其雏形“中华一号”于1985年6月问世。目前国内市场上流行的产品中，莱思康、好易通等产地为台湾；快译通、伟易达等产自香港，夏普、卡西欧等系日本所产。国内软件开发商从20世纪90年代开始研制电子字典，迄今上市的至少已有10余种产品。能运行于网络上的电子字典已有北京特科能公司《Roboword中日英多功能、多语态、动态词典》、上海朗道公司《朗道电子辞典》、中软总公司的《汉译机器翻译系统》、金山公司的《金山词霸》等多种。

某些中文平台也附有词典功能。例如北京四通利方信息技术有限公司开发出了Richwin For Internet平台，在全球首次推出了中文乃至日文内码间的自动识别，备有英汉光标快捷词典，便利了中国人在网上顺利阅读英文信息和外国人浏览中文信息。专用的中文浏览器，已有湖南国讯网络有限公司的“译王97”、桑夏集团的NetReader等，它们都具备在线翻译功能。虽然这些资源的使用可以帮助我们在没有其他帮助的情况下，有限地理解一些外语文章，但是通过这些工具一般无法获得高质量的阅读和理解。对于重要的信息，最好是能够设法寻找其他途径以求确认，并且最好能够有一定的英语（或法、德语等之一）基础再来使用它们。在线翻译或借助其他翻译软件的方法只能作为阅读网上信息的一个解决手段，不可能代替语言翻译。也就是说掌握基本的英语是访问Internet中国以外地域的最基本的条件，字典只能作为辅助的手段。

（3）常用中文搜索引擎 以下按关键词搜索引擎与分类搜索引擎两大类型介绍一些较有影响的中文搜索引擎。1）关键词检索站点 ChinaByte CSECK CSECK（图1-3-10）的检索功能非常强劲，除了国外的搜索引擎，它可能是我们评价的中文搜索引擎中检索功能最强的一个，因为它使用的搜索引擎核心是非常成熟的商用软体verity.但正因为其功能强劲且有些复杂，要充分利用其强劲的检索功能，用户必须详细阅读其帮助页面，这对于普通用户来说有些勉为其难。CSECK的界面虽然也算清晰明了，但在其页面上缺乏对于其检索功能的简单描述或者是一些辅助选项按钮，不能不说是一个小小的缺憾。像大多数的关键词搜索引擎一样，CSECK也有一个分类库供用户查询，但此库显然是疏于管理，缺乏更新，每一

个子类最多为30条网页。另外其结果显示中的网页摘要由于是由其搜索引擎按一定算法自动生成，虽然说有时比单纯引用前几行文字要更具有说明意义，但有时也使人莫名其妙。

网易 网易网站以其提供的一系列服务见长，如免费主页、免费E-mail都在国内具有一定的影响。它的搜索引擎也是颇有特色，它先将用户的检索式在自己的分类库中进行查询，如果没有检索出结果，系统将自动将提问式转向全文数据库进行检索。如果在分类库中检索出结果，用户对检索结果不满意，可以直接按检索结果页面底部的全文检索按钮，继续在全文库中进行检索，这对于一个非专业用户来说是非常实用的一种检索策略。另外，网易的分类库也做得相当不错（其分类类目与方法大致上与哇塞相同），这比许多带分类库的关键词搜索引擎要好。网易的功能简单、使用方便，并且提供一定的模糊检索功能，这也是非常能吸引普通用户的功能。网易的界面清晰，反应及库容量也是可以接受的。

天网中英文搜索引擎

天网是由CERNET在北京大学设立的一个较优秀的中文搜索引擎，主要以CERNET的网页信息为主。作为国内自行开发的搜索引擎，具有一定的水准。它像很多国外的优秀搜索引擎一样，系统不仅提供WWW网页的查询，同时也提供对新闻组（Newsgroup）内容的查询。天网将用户的检索式进行自动分词，然后进行检索。同时天网还提供模糊匹配检索模式，在模糊匹配中，系统不仅检出用户表达式中的关键词，并且将其同义词也一并检出。天网查询分为简单和复杂两种，简单查询时所输入的关键词可以是英文或中文词，英文词不必区分大小写字母；每次查询可输入最多10个关键词（OR关系）；英文词中支持通配符，如"?"匹

配任一字母，"*"匹配0到多个任意字母。天网的复杂查询支持简单查询所有的功能，但检索选项可以指明匹配关键词是用AND关系还是OR关系；可以设定每页显示查询结果的方式，如每页显示多少项以及如何介绍每个URL。天网支持通过E-mail的检索，即用户可将检索式通过E-mail发送检索提问式，而天网搜索引擎通过E-mail回送检索结果。同时天网还提供对ftp.pku.edu.cn、ftp.lib.pku.edu.cn、ftp.ncic.ac.cn、ftp.igd.edu.cn等FTP服务器的FTP文件检索。天网的优点是速度比较快，使用简便、索引很大，支持中英文两种查询方式。缺点是重复率太高，不够精确。

悠游中文搜索 悠游中文搜索（图1-3-13）是一个较有影响的中文搜索引擎，两岸三地是它的特色。它原先由美国、北京、重庆、香港四个网站组成，最近又增添了上海网站，同时台湾的网站也已推出。悠游每一个网站充分考虑到两岸三地的特点，所以每一个网站都有GB与BIG5两种页面，并且两个页面的内容还考虑到不同用户的需求而有所不同。同时各地网站主页面还有本地新闻等其他信息，各地网站还推出不同的服务内容等等。除了关键词检索功能外，悠游还提供了一个简单而排版清晰（类似于Yahoo！风格）的分类库。除了网页搜索以外，悠游搜索引擎还提供了网址（URL）查询功能。另外悠游搜索引擎很有特色的一点就是它能够将用户的检索式进行自动分词，然后再进行检索。因此用户在输入检索式的时候，可以按照自己的习惯直接输入，但有时如果用户键入一个长的检索式时，常会检索出一大堆检索结果。但是悠游搜索引擎在主页上未提供任何帮助信息，对于用户来说多有不便，用户只有在进入分类检索时才能找到一个帮助链接，但其帮助信

息也写得较为简单，参考价值一般，搜索引擎本身也不支持什么逻辑检索的功能。

2) 分类搜索引擎

搜狐 Suhu网站的功能类似于Yahoo!，是一个由我国自行开发的大型网上中文分类搜索引擎，目前已达到3万以上的日访问人数和30万的页读数，成为因特网上著名的中文网站。它以最先进的人工分类技术、友好的符合中文语言文化习惯的全中文界面，共分18个部类、近10万条链接构成的树杈型网页结构，为网上用户直观、轻松地提供所需要的内容。搜狐的目标不仅仅是作为一个网站，而且希望超越媒体，成为人们生活中不可或缺的电子商务市场。搜狐分层目录是专为中国用户设计的高质量的分类目录系统，它把超过5万个精选的中文站点归为18个大类，如：娱乐、计算机与互联网、工商经济、科学、艺术、文学、生活服务、科学技术、政治法律等等，而且在每个分类目录下，都建立全面丰富的目录树系统。分层目录系统不仅方便用户找到最需要的专门信息，也使广告客户更便捷地找到特定的用户群，达到最佳的宣传效果。

蕃薯藤中文搜索引擎、哇塞中文搜索引擎的发展有一个不断积累的过程，所以起步较早的网站技术与资金条件可能也更好一些，台湾地区的中文搜索引擎尤其是分类搜索引擎明显走在国内同行的前面。蕃薯藤与哇塞的共同特点是界面非常清晰，不像国内的分类搜索引擎那样过于繁复且尽是广告。从界面设计来看，哇塞是中文搜索引擎中界面做得最干净的一个，清晰利落，颜色配比也颇具匠心，令用户阅读起来相当方便。哇塞另外用清晰的小图标标明「酷」和「新」的站点，也使用户感到既非常实用又体贴备至。蕃薯藤的检索功能做得更强一些，它的通配符检索功能也使它在检索上比较方便用户

。在此要指出的是，中文检索的通配符与西文检索的通配符用户不同。在西文检索中，通配符的作用是能让用户检索出包含一个词不同时态和用法的结果集。而在中文检索中，通配符可以替代任何字符，较类似于相近检索（NEAR），或者说比相近检索功能更丰富一些，因此中文检索中的通配符功能还是相当实用的。F8F8" 100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com