



# 搜索引擎检索技巧

# 搜索引擎

搜索引擎(search engine), 1995年开始

搜索引擎以一定的策略从网络收集、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目的。

搜索引擎站——“网络门户”

# 1、搜索引擎的工作原理

◆信息的收集处理

◆信息的检索输出

## 2、搜索引擎的分类

搜索引擎按其工作方式主要可分为三种：

- ◆ 目录索引类搜索引擎（Search Index/Directory）
- ◆ 机器人搜索引擎（全文搜索引擎）（Full Text Search Engine）
- ◆ 元搜索引擎（Meta Search Engine）

## 2、搜索引擎的分类（续）

### 目录式搜索引擎

- ◆ 目录式搜索引擎：以人工方式或半自动方式搜集信息，由编辑员查看信息之后，人工形成信息摘要，并将信息置于事先确定的分类框架中。信息大多面向网站，提供目录浏览服务和直接检索服务。
- ◆ 该类搜索引擎因为加入了人的智能，所以信息准确、导航质量高，缺点是需要人工介入、维护量大、信息量少、信息更新不及时。
- ◆ 这类搜索引擎的代表是：yahoo!、Galaxy、Open Directory.....

## 2、搜索引擎的分类（续）

### 机器人搜索引擎

- ◆ 由一个称为蜘蛛（Spider）的机器人程序以某种策略自动地在互联网中搜集和发现信息，由索引器为搜集到的信息建立索引，由检索器根据用户的查询输入检索索引库，并将查询结果返回给用户。服务方式是面向网页的全文检索服务。
- ◆ 该类搜索引擎的优点是信息量大、更新及时、毋需人工干预，缺点是返回信息过多，有很多无关信息，用户必须从结果中进行筛选。
- ◆ 这类搜索引擎的代表是：AltaVista、Northern Light、Excite、Infoseek、Inktomi、FAST、Lycos、Google；国内代表为：百度等。

## 2、搜索引擎的分类（续）

### 元搜索引擎

- ◆元搜索引擎：这类搜索引擎没有自己的数据，而是将用户的查询请求同时向多个搜索引擎递交，将返回的结果进行重复排除、重新排序等处理后，作为自己的结果返回给用户。服务方式为面向网页的全文检索。
- ◆这类搜索引擎的优点是返回结果的信息量更大、更全，缺点是不能够充分使用所使用搜索引擎的功能，用户需要做更多的筛选。
- ◆这类搜索引擎的代表是WebCrawler、InfoMarket等。

# 常用的一些搜索引擎网址

◆ <http://www.google.com.hk>

◆ 1997年，（简体中文）提供有类目检索和网站检索两种方式。支持 A N D 和“-”等条件查询。以搜索精度高、速度快成为最受欢迎的搜索引擎，是目前搜索的领军人物。



# 常用的一些搜索引擎网址

◆ <http://www.baidu.com>

◆ 中国提供搜索引擎的主要网站中，超过**80%**由百度提供。**1999**年底，百度成立于美国硅谷，**2000**年百度回中国发展。是国内唯一商业化的全文搜索引擎，提供搜狐、新浪、**263**、**TOM**等站点的搜索服务。

# 常用的一些搜索引擎网址

◆ <http://www.metacrawler.com>

◆ 元搜索引擎，1995年，Washington大学开发，2000年被Infospace收购。

检索方法与技巧：模块分类、关键词检索；  
支持布尔逻辑检索；可以使用检索语法；

# 搜索引擎高级语法

## 1. 要求搜索结果中不含特定查询词

- ◆ 如果您发现搜索结果中，有某一类网页是您不希望看见的，而且，这些网页都包含特定的关键词，那么用减号语法，就可以去除所有这些含有特定关键词的网页。
- ◆ 例如，搜 神雕侠侣，希望是关于武侠小说方面的内容，却发现很多关于电视剧方面的网页。那么就可以这样查询：神雕侠侣 -电视剧
- ◆ 注意，前一个关键词，和减号之间必须有空格，否则，减号会被当成连字符处理，而失去减号语法功能。减号和后一个关键词之间，有无空格均可。

## 2. 精确匹配——双引号和书名号

- ◆ 如果输入的查询词很长，百度在经过分析后，给出的搜索结果中的查询词，可能是拆分的。如果您对这种情况不满意，可以尝试让百度不拆分查询词。给查询词加上双引号，就可以达到这种效果。
- ◆ 例如，搜索四级英语试题，如果不加双引号，搜索结果被拆分，效果不是很好，但加上双引号后，“四级英语试题”，获得的结果就全是符合要求的了。

◆ 书名号是百度独有的一个特殊查询语法。在其他搜索引擎中，书名号会被忽略，而在百度，中文书名号是可被查询的。加上书名号的查询词，有两层特殊功能，一是书名号会出现在搜索结果中；二是被书名号扩起来的内容，不会被拆分。书名号在某些情况下特别有效果，例如，查名字很通俗和常用的那些电影或者小说。比如，查电影“手机”，如果不加书名号，很多情况下出来的是通讯工具——手机，而加上书名号后，《手机》结果就都是关于电影方面的了。

# 搜索引擎高级语法

## 3. 把搜索范围限定在网页标题中——**intitle**

- ◆ 网页标题通常是对网页内容提纲挈领式的归纳。把查询内容范围限定在网页标题中，有时能获得良好的效果。使用的方式，是把查询内容中，特别关键的部分，用“**intitle:**”领起来。
- ◆ 例如，找四级英语的真题，就可以这样查询：  
**真题 intitle:四级英语** 注意，**intitle:**和后面的关键词之间，不要有空格。

# 搜索引擎高级语法

## 4. 把搜索范围限定在特定站点中——site

- ◆ 有时候，您如果知道某个站点中有自己需要找的东西，就可以把搜索范围限定在这个站点中，提高查询效率。使用的方式，是在查询内容的后面，加上“site:站点域名”。
- ◆ 例如，天空网下载软件不错，就可以这样查询：  
**msn site:skycn.com**
- ◆ 注意，“site:”后面跟的站点域名，不要带“http://”；另外，site:和站点名之间，不要带空格。

# 搜索引擎高级语法

- ◆ Site可将搜索定位于某一个域名上，有利于缩小搜索范围，提高搜索效率。如 **keywords site:edu**，效果很好！可以方便地搜索到各大学内有用的资源。
- ◆ **site: edu.cn.....**



# 搜索引擎高级语法

## 5. 专业文档搜索

- ◆ 很多有价值的资料，在互联网上并非是普通的网页，而是以Word、PowerPoint、PDF等格式存在。百度支持对Office文档（包括Word、Excel、Powerpoint）、Adobe PDF文档、RTF文档进行了全文搜索。要搜索这类文档，很简单，在普通的查询词后面，加一个“filetype:”文档类型限定。“Filetype:”后可以跟以下文件格式：DOC、XLS、PPT、PDF、RTF、ALL。其中，ALL表示搜索所有这些文件类型。例如，“**高等数学试卷 filetype:doc**”查找关于高等数学方面的试卷Word文档。点击结果标题，直接下载该文档，也可以点击标题后的“HTML版”快速查看该文档的网页格式内容。

# 搜索引擎高级语法

## 6. 把搜索范围限定在url链接中——inurl

- ◆ 网页url中的某些信息，常常有某种有价值的含义。于是，您如果对搜索结果的url做某种限定，就可以获得良好的效果。实现的方式，是用“inurl:”，后跟需要在url中出现的关键词。
- ◆ 例如，找关于Excel的使用技巧，可以这样查询：  
**Excel inurl:使用技巧**
- ◆ 上面这个查询串中的“Excel”，是可以出现在网页的任何位置，而“使用技巧”则必须出现在网页url中。
- ◆ 注意，inurl:语法和后面所跟的关键词，不要有空格。

# 搜索引擎高级语法

## ◆ inurl代替filetype

**filetype**可以搜索pdf、doc、ppt等格式的专业文档。同一样格式的文档自然有同样的后缀名。因此，可以用**inurl**来代替。

例：

信息检索 **filetype:pdf**

信息检索 **inurl:pdf**

◆想看看别人对**IPHONE 5**的评价，用  
“**IPHONE5 inurl:bbs**”搜索就能看到论  
坛里的说法。

◆甲流怎么样了，用“**甲流 inurl:news**”查  
找。

◆找音乐

**inurl:mp3** 晚风

# 搜索引擎高级语法

## 7、限定关键词只在标题中

- ◆ 例如“**allintitle:中国 苹果**”表示“中国”和“苹果”都必须出现在标题中
- ◆ “**intitle:中国 苹果**”表示“中国”必须出现在标题中，“苹果”可以出现在网页的任意位置，可以是标题也可以不是。
- ◆ “**title:**”用于针对标题进行搜索（百度、Yahoo）

例: **title:电路分析题**

# 搜索引擎高级语法

## 8. 限定关键词只在URL中

- ◆ 例如“**all inurl:高等数学 试题**”表示“高等数学”和“试题”都必须出现在URL中。
- ◆ “**inurl:高等数学 试题**”表示“高等数学”必须出现在URL中，“试题”可以出现在网页的任意位置，可以是URL也可以不是。

# 搜索引擎高级语法

## 9、在指定网站内搜索

- ◆ 如果想知道某个站点中是否有自己需要找的东西，可以把搜索范围限定在这个站点中，提高查询效率。在想要搜索指定网站时，使用site语法，其格式为：查询词+空格+site:网址。例如，只想看搜狐网站上的世界杯内容，就可以这样查询：

世界杯 site:sohu.com 。多个站点用“|” 隔开：

世界杯 site:www.sina.com.cn | www.sohu.com

site:和站点名之间，不要带空格。

# 搜索引擎高级语法

◆ filetype语法也可以与site语法混用，以实现在指定网站内的文档搜索。例如在北京大学和清华大学网站内搜索有关“硕士专业”的word文档，就可以用：

site: www.pku.edu.cn |  
www.tsinghua.edu.cn filetype:doc 硕士专业



# 搜索引擎的查询技巧

- ◆分析研究对象，选用搜索引擎
- ◆确定搜索途径，使用不同方式
- ◆正确选用搜索引擎的搜索选项
- ◆正确选用搜索引擎的各项搜索功能

# 根据反馈，修改检索

- ◆缩小搜索范围：利用二次搜索（重新查询、在结果中再查、在结果中去除）、使用检索语法构造恰当的检索表达式、选用正确的关键词。
- ◆扩大搜索范围：使用同义词或近义词、使用检索语法or和截词检索、使用多个搜索引擎和元搜索引擎

# 如何提高查准率？

- ◆提高专指度（用专指性强的自由词）
- ◆用逻辑乘and 或 not 相关检索项；如查找克林顿，但不要关于莱温斯基的信息。  
检索式：克林顿 -莱温斯基
- ◆用文献外部特征限制输出结果（在中文图书中查找）；
- ◆用“二次检索”、“条件检索”排除误检；
- ◆限制检索字段，指定邻接和优先关系。