

# 百度、Google、Sogou 三种图像搜索引擎功能的比较

高瑞 中国科学技术信息研究所 100038

## 摘要

本文介绍了图像搜索的几种途径,然后针对百度、Google、Sogou这三种常用的图像搜索引擎进行了用户界面设计、高级检索功能的比较,并通过实际测试分析了它们的检索数量、检索用时、查准率。

## 关键词

图像搜索; 百度; Google; Sogou

## Abstract

This paper introduces methods of image searching, compared baidu, Google, and Sogou in user interface and advanced search, analyze the quantity, time and accuracy through tests.

## Key words

image search; baidu; Google; Sogou

## 引言

百度、Google、Sogou作为互联网用户搜索信息时比较常用的几个综合性搜索引擎,提供了网页搜索、视频搜索、音乐搜索、在线翻译、贴吧、问答、地图等多种搜索功能。其中,图像搜索由于其直观性强,生动形象的特点,受到了广大用户的青睐,互联网上的信息量浩如烟海,选择一个好的图像搜索引擎,可以大大提高用户搜索图片的准确性,提高效率,节约时间。笔者分析了当前图像搜索的类型与评价标准,针对百度、Google、Sogou这三种搜索引擎进行了用户界面设计、高级检索功能等多方面的比较。

### 1 图像搜索的几种途径

搜索引擎是根据一定的策略、运用特定的计算机程序搜集互联网上的信息,在对信息进行组织和处理后,并将处理后的信息显示给用户,是为用户提供检索服务的系统。<sup>[1]</sup>图像搜索引擎是一种专门用来查询图像、图片并以缩略图、网站名或网

站内容等形式反馈检索结果的网上检索工具。其工作原理主要是利用其内部的蜘蛛程序自动对网站进行搜索,并对在互联网上浏览过的图像进行分析和判断,为图像加注释、存储抽取出的索引信息,建立图像索引库,从索引库中或者直接从大量存储图像的数据库中查找检索出具有相似特征的图像。<sup>[2]</sup>

图像搜索引擎识别图像的方式主要有两种,第一是自动查找图像文件,第二是人工干预找出图像并进行分类。

图像搜索的途径主要分为以下几种:

#### (1) 关键词检索

基于外部信息或人工赋予的关键词进行自由检索。外部信息包括图像的目录、路径、链接、文件名,以及图像周围的文本信息,根据这部分信息的检索式目前图像搜索引擎采用最多的方法。

#### (2) 分类目录浏览

图像搜索引擎把采集到的图片通过某种自动机制或者采用人工的方式进行分类标引,为互联网用户提供按主题浏览的检索方式。这种检索方式适合于以浏览为目的的用户,或者检索目标不明确的用户。

#### (3) 基于图像可视属性的检索

基于图像可视属性的检索又称基于图像内容的检索。图像分析软件自动抽取图像的颜色、形状、纹理等特征,并建立图像特征索引库,只要互联网用户将需要搜索的图像的特征描述出来,就可以找出与之具有相近特征的图像。这种基于图像固有属性的机械匹配,适合于检索目标明确的图像检索,其检索结果也比较准确。

目前,大多数的图像搜索引擎都支持关键词检索和分类目录浏览检索,部分可以支持基于图像的可视属性检索,但支持可视属性检索的范围十分有限。

## 2 三种图像搜索引擎的比较

### 2.1 用户界面设计的比较

互联网用户在浏览和使用图像搜索引擎时,首先对面的是用户界面,因此,一个图像搜索引擎的用户界面设计的好坏,在很大程度上影响了用户的选择。

百度的用户界面设计十分简洁易用,打开百度首页,点击“图片”链接,就可以进入百度的图像搜索界面,默认的搜索范围是全部图片,另外,百度提供了“新闻图片”、“壁纸”、“表情”、“头像”这些常用的图片范围,可供用户选择。在图片搜索的界面上,还有一些热点目录,如“爆笑趣图”、“精品推荐”、“美女明星”、“帅哥明星”、“精美壁纸”、“电影电视”、“风景名胜”,以及手机娱乐,如“热辣彩图”、“炫彩动画”。

Google的用户界面相对来说比较复杂。用户可以在Google的首页上选择“图片”链接,就可以进入Google的图像搜索界面。Google的图像搜索界面上显示了用户搜索的热门图片,并显示出该图片和该链接。另外,在“热门搜索”栏目中,Google给出了用户搜索频率最高的八项图片名称,在“上升最快搜索”的栏目中,Google给出了被搜索的频率上升速度最快的八项图片名称。在Google的图像检索首页上,还显示了一些常用的图片分类,如“生活时尚”、“潮流女星”、“闪亮男星”、“影视集锦”、“游戏动漫”、“精美壁纸”、“爆笑趣图”、“体育军事”、“风景名胜”。由于Google图像搜索首页的内容丰富,所以给用户提供了广泛的选择,另外,由于在首页上显示了多种图像的分类,这对用户搜索图片提供了参考范围。

Sogou的图像搜索见面与百度的比较相似。在Sogou的首页上点击“图片”链接,即可进入到Sogou的图像检索,默认的检索范围是全部图片,用户可以选择

择“大图”、“QQ头像”、“壁纸”、“套图”。另外,在图片搜索的首页上还显示了图片的分类目录,如“性感女星”、“魅力男星”、“非主流图”、“精美壁纸”、“爆笑趣图”、“网页素材”等。“一周热词”滚动显示了这一周被检索的频率最高的图片名称和链接,为用户搜索图片提供了参考。在“图片专题”栏目中,Sogou给出了一些近期热门的图片搜索专题,但更新速度还不够快,一些图片专题已经失去了时效性,其更新图片的速度有待改进。

总的来说,百度和Sogou的用户界面设计风格倾向于简约、实用,Google的用户界面包含的内容比较丰富。

## 2.2 高级检索功能比较

检索功能的多少是用户在选择搜索引擎时需要考虑的一个方面,在图像搜索引擎中,主要通过“高级检索”功能来实现。

百度的高级检索功能可以设置搜索结果类型,如“图片”、“新闻图片”,用户可以根据期望的图片尺寸选择图片的大小,如“大图片”、“中图片”、“小图片”,还可以选择不同分辨率的壁纸或者由用户自定义图片的大小。在颜色的选择方面,用户可以根据期望搜索的图片颜色来选择颜色。在图片格式方面,用户同样可以根据期望来搜索图片的格式,如jpg、gif、bmp、png格式。用户也可以指定网站进行图片的站内检索,只需要输入相应的网址即可。用户可以根据自己的偏好进行个性设置,如是否在搜索时显示搜索框提示等。

Google的高级检索功能可以对检索结果进行一定程度的过滤,用户只要输入一个或多个关键词,就可以进行一下选择:“必须和下列的全部字词有关系”、“必须和下列的字句完全符合”、“只要和下列的任何一个字词有关系”、“和下列字词无关”。在检索的内容类型方面,用户可以选择任意内容、资讯内容、脸部特写、照片内容、剪贴画、素描画。在图片尺寸方面,用户可以选择任意尺寸、大尺寸、中尺寸、图标、不同分辨率和像素的图片。在精确尺寸方面,用户可以根据需要输入不同的宽度和高度,或者根据用户计算机的屏幕尺寸来搜索图片。在纵横比方面,用户可以选择任意纵横比、竖版、正方、横版、超大。在文件类型方面,用户可以选择所有文件类型、jpg文件、gif文件、png文件、

bmp文件。在图片颜色方面,用户可以选择全部图片、黑白图片或全彩图片。在限制网站方面,用户可以只显示特定站点或域内的图片。在使用权限方面,用户可以选择未经许可过滤的图片、标明可供使用的图片、标明可用于商业用途的图片、标明要修改后方可使用的图片、标明要修改后方可用于商业用途的图片。

Sogou的高级检索功能比较简单。在检索结果方面,用户可以选择包含全部关键词的检索。在搜索结果的类型方面,用户可以选择全部图片、QQ头像、壁纸、套图。在图片大小方面,用户可以选择想要的图片尺寸,如全部大小、大图片、中图片、小图片或自定义大小。在图片颜色方面,用户可以选择想找图片的色调。另外,用户可以输入网址,在指定的网站内进行准确定位。

## 2.3 检索性能分析

图像搜索引擎的检索性能包括很多方面,如**检索到的图片数量,检索用时,查全率和查准率**,笔者通过具体的检索词进行测试,分析了百度、Google、Sogou这三种图像搜索引擎的检索性能。

### 2.3.1 检索数量

笔者选择了不同类型的8个关键词进行检索实验,分别是周杰伦、变形金刚、手机、油画、长城、靴子、奔驰、十月围城。经过多次检索,得到的结果如下列的表格所示。

表1 检索数量(单位:张)

	百度	Google	Sogou
周杰伦	1,390,000	215,000	4,619,516
变形金刚	523,000	524,000	3,252,939
手机	92,500,000	456,000	24,161,744
油画	1,170,000	428,000	5,702,067
长城	1,580,000	359,000	6,338,843
靴子	308,000	543,000	1,726,936
奔驰	1,340,000	348,000	4,802,883
十月围城	44,400	195,000	199,294

表1显示了这三种图像搜索引擎在检索数量上的对比,可以看出,Sogou的检索数量明显高于百度和Google,检索到的图片较为齐全,从而使得Sogou的查全率比较高,而百度能检索到的图片数量比Google多,Google能检索到的图片数量是三种搜索引擎中最少的。但总的来说,这三种搜索引擎能够检索到的图片总量都是非常可观的,可以充分满足用户的需求。

### 2.3.2 检索用时

表2显示了这三种图片搜索引擎的检

索用时,通过对比可以看出,Sogou的检索用时最少,而且每个关键词的检索基本上都在0.002秒,百度的检索用时比Sogou长,Google的检索用时最长,但是平均时间比较稳定,大约都在0.05秒左右。当然,检索用时与计算机的配置、网速有着密切的联系,笔者是在一个比较稳定的网络环境下对三种图像搜索引擎进行测试的。对于普通的互联网用户而言,三种搜索引擎的检索时间都是非常理想的,几毫秒的时间差距并没有对用户产生明显的影响。

表2 检索用时(单位:秒)

	百度	Google	Sogou
周杰伦	0.008	0.05	0.002
变形金刚	0.007	0.05	0.003
手机	0.016	0.04	0.002
油画	0.011	0.06	0.002
长城	0.008	0.04	0.002
靴子	0.006	0.05	0.002
奔驰	0.0011	0.04	0.002
十月围城	0.009	0.05	0.002

### 2.3.3 查准率

表3 前100张查准率(百分比)

	百度	Google	Sogou
周杰伦	86	88	92
变形金刚	89	77	87
手机	97	97	89
油画	95	99	95
长城	95	79	11
靴子	99	97	91
奔驰	93	99	98
十月围城	93	94	90
平均查准率	93.375	91.25	81.625

用户在搜索图片的时候,对于搜索结果,一般只会查看搜索到的前几页的图片,而不是查看所有图片,所以,笔者选择了排在前面的100张图片进行了查准率的测试。从表3可以看出,百度的查准率比Google略高一点,相对于Sogou来说,这两种图像搜索引擎的检索性能比较稳定,在查准率方面没有很大的差别。Sogou的查准率明显低于百度和Google,而且性能不够稳定,比如当用户检索“长城”这个关键词时,查准率只有11%,Sogou在查准率方面还有待进一步的提高。

## 3 结论

通过对百度、Google、Sogou这三种图像搜索引擎的用户界面设计、高级检索功能,以及关键词检索的测试和分析可以总结出它们的不同特点。

SM89516A 的 P0.2 和 P0.3 口与 DS1302 的 SCLK、I/O、CE 脚进行串口通信,组成实时时钟/日历电路,提供秒、分、时、日、月、年的信息,每月的天数和闰年的天数可自动调整时钟操作。

#### 4、程序设计

系统程序设计的功能主要有系统的初始化、用户温室环境参数实时采集和温室环境信息的 SMS 消息发送,用户 SMS 消息的接收及删除。其流程如图 3 所示。

系统的初始化工作包括定时器的设置、波特率的设置、TC35i 模块的初始化等。TC35i 模块的初始化包括模块的开启, SIM 卡 PIN 码的检测,短消息发送和接收的方式。

室温参数实时采集内容包括:正常状态下,单片机在规定的时间内采集环境信息并把数据传输到监控中心;一旦发生串口中断时,单片机读取 TC35i 中的短信息内容,对其进行解码,采集相关的信息。

温室环境信息的 SMS 消息发送和用户 SMS 消息的接收及删除:只要正确设置,系统就能接收用户的指令。一旦接收到短信息时, TC35i 就会给单片机一个中断,程序会读取 SIM 卡内存中 0~19 位置的短消息,读取短消息的 AT 指令就是 AT+CMGR=<index>, index 值为 0~19,成功读取到消息后,判断短信中是否包含指定的字符,如果没有包含指定字符,程序将退出中断服务。如果程序包含指定信息,则提取出短消息发送者的号码,用 AT 指令 AT+CMGS=139 × × × × × × × × 进行发送号码设置,用于发送短消息。由于是使用 Text 模式发送短消息,所以直接将采集到的环境信息发送给串口就可以了。发送短消息后,要将包含指定字符的那条短消息删除,避免下一次循环检测中被检测到,造成误差<sup>[5]</sup>。删除短消息的 AT 指令是 AT+CMGR :<index>, index 值为 0~19。

#### 5、结束语

本系统使用移动通讯网路作为远程通信的实现技术,降低了架设通信线路的高额成本,具有运行费用低廉,性价比高的特点。并可结合不同的控制系统创造出更智能的应用。该模块及文中提出的外围电路设计已经在供暖系统中得到了

实际的应用,取得了非常好的效果,真正实现热网的自动按需供热。

◀ 上接第 92 页

百度的用户界面设计简单,方便用户操作,高级检索功能齐全,合适高级用户的使用,关键词的测试表示其检索数量、检索用时、查全率方面的检索性能较好,在中文图片的检索方面,百度已经相当完善。

Google 的用户界面设计风格较为复杂,但有利于用户的分类浏览,高级检索功能与百度略有不同,但也十分丰富,便于高级用户的使用。在检索数量、检索用时、查全率方面, Google 比百度略低一点,但是基本上不影响用户的使用。总的来说, Google 作为图像搜索引擎,总体性能还是非常好的。

Sogou 是搜狐公司于 2004 年推出的全球首个第三代互动式中文搜索引擎,在图像搜索方面,其用户界面设计简单,方便操作。高级检索的功能比较简单,不能实现相对复杂的检索要求。在检索数量和检索用时方面, Sogou 的性能明显高于另外两个搜索引擎,但是在查准率方面, Sogou 的性能还不够稳定,有待提高。

#### 参考文献

- [1]潘斌,郭红霞.短信收发模块 TC35i 的外围电路设计[J].单片机与嵌入式系统应用.2004(7):36-37.
- [2]李栋梁,马洪连,李文成.基于 SMS 的远程数据采集系统的设计与实现[J].仪器仪表学报.2006(z1):55-56.
- [3]Siemens TC35 技术手册.
- [4]戴梅萼.微型计算机技术及应用(第二版)[M].清华大学出版社.
- [5]林凌,李刚,丁茹等.新型单片机接口器件与技术[M].西安电子科技大学出版社.2005.

#### 参考文献

- [1]付天香.Google、Yahoo 和百度的图像搜索比较[J].图书馆学刊.2009(2):103-106
- [2]梁小红.几种图像搜索引擎的比较[J].图书馆论坛.2004(8):103-105
- [3]颜丽君.新浪、Google、Yahoo 图像搜索引擎比较[J].图书情报工作.2003(10):83-87
- [4] <http://image.baidu.com/>
- [5] <http://images.google.cn/imgcat/imghp?hl=zh-CN>
- [6] <http://pic.sogou.com/?p=40030500&kw>