第 5 章　利用搜索引擎进行信息分析和决策

搜索引擎并非只是一种简单的信息检索工具，利用它提供的用户行为分析结果和网络舆情发展情况，可以更好地指导各种经济活动。具体来说，这主要是利用搜索引擎收集到的检索关键词、用户检索行为、用户网上问卷调研结果和新闻内容监测与分析等数据，提供两种主要功能：一是网络用户的检索行为分析。不同于一般的用户调研，利用用户在搜索引擎中的检索信息得到的用户需求往往更为准确和客观，更能反映用户真实的潜在需求。二是进行互联网舆情监测，通过对新闻报道的获取和分析，监测产品和企业的相关市场反应，为企业的市场公关策略、危机预警和处理等提供客观及时的咨询建议。

通常这些信息是以查询日志的方式存储在搜索引擎服务器中，但是现在大多数著名搜索引擎都已经公开了这些信息，而且大都开始提供基于此类原始信息生成的分析报告。这些信息对企业或者组织分析市场、分析决策有着重要的参考作用。

通过本章的学习，能够让读者加深对此问题的认识，同时也可以在各种经济工作中更好地应用搜索引擎来解决一些实际问题。

5.1　百度的信息分析决策功能

1. 百度热搜

百度热搜是百度搜索引擎利用用户检索关键词，将各类热门搜索词语和相关查询结果按照搜索量排列整理而成。通过搜索量的大小，用户可以了解互联网用户的信息需求特点和社会热点问题。百度每天都会根据前一天的搜索量自动计算，统计得到当日的热搜信息，其中重要栏目还包括热搜榜、实时脉搏和热点活动等内容。网址为 http://top.baidu.com，主页如图 5.1 所示。

其中，热搜榜主要根据用户在百度搜索引擎中的查询信息来给出当前最为热门的查询主题，并提供多种不同类别的分类统计，比如“小说”、“电影”和“汽车”等。实时脉搏则通过可视化界面展示当前的热门搜索词语，并可以直接单击打开相应的检索结果界面。热点活动则按照当前一些关注热点事件或者主题给出更为详细的分类展示，比如“两会大数据”，通过综合热搜榜、热门提案议案和热门知识等直观的方式提供一站式的两会热点信息查询，如图 5.2 所示。

2. 百度指数

较百度热搜而言，百度指数对这些关键词信息提供了更为详细和灵活的汇总结果，现在

图 5.1　百度热搜的主页界面(截取于 2022-11)

图 5.2　百度热搜中“两会大数据”结果页面(截取于 2022-11)

的数据统计都包括桌面 PC 端和移动终端两个体系的统计。同时还提供了包括诸如“最新动态”和“行业排行”等功能,网址为 http://index.baidu.com。用户必须登录百度账号才能使用,主页如图 5.3 所示。

其中行业排行里提供了网民对品牌的品牌指数、品牌搜索指数、品牌资讯指数、品牌互动指数排名和上升下降趋势等信息,反映品牌在行业中的位置和变化趋势功能。

比如想了解江苏省南京市 2021 年 361 和特步两个运动类品牌的数据对比情况,可以直接添加相关关键词,并选择时间和地区,相关结果如图 5.4 所示。

再如想了解江苏省地区在 2022 年高考结束后大众对高校的关注度,可以选择“行业排

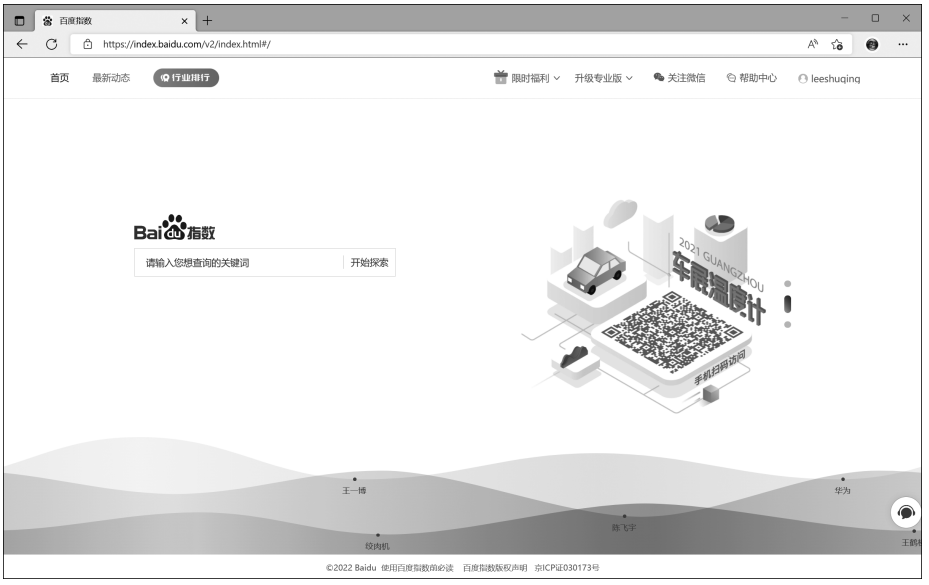

图 5.3 百度指数的主页界面（截取于 2022-11）

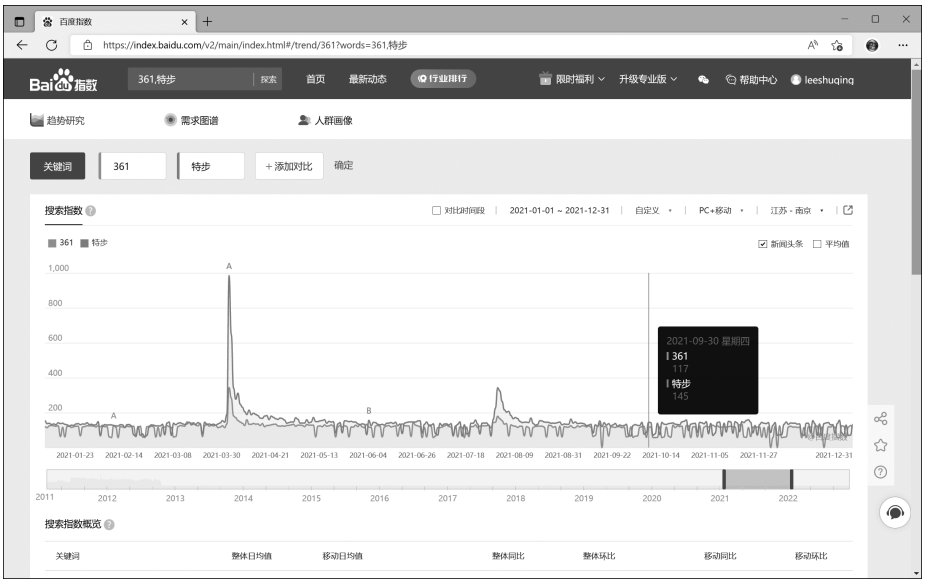

图 5.4 在百度指数中检索不同产品的相关数据统计结果（截取于 2022-11）

行”并进一步选择“高校行业排行”，设定检索时间和区域，即可看到相关统计结果，如图 5.5 所示。

3. 百度司南

百度司南于 2009 年开始提供服务，它也是利用搜索引擎用户搜索行为分析的数据结果来提供相关信息决策支持服务。主要的特点一个是基于大数据分析方法，另一个是侧重营销决策支持，目标是帮助广告主在网络上找到更多、更合适的潜在用户。不过，该服务为收费服务。网址为 http://sinan.baidu.com，主页如图 5.6 所示。

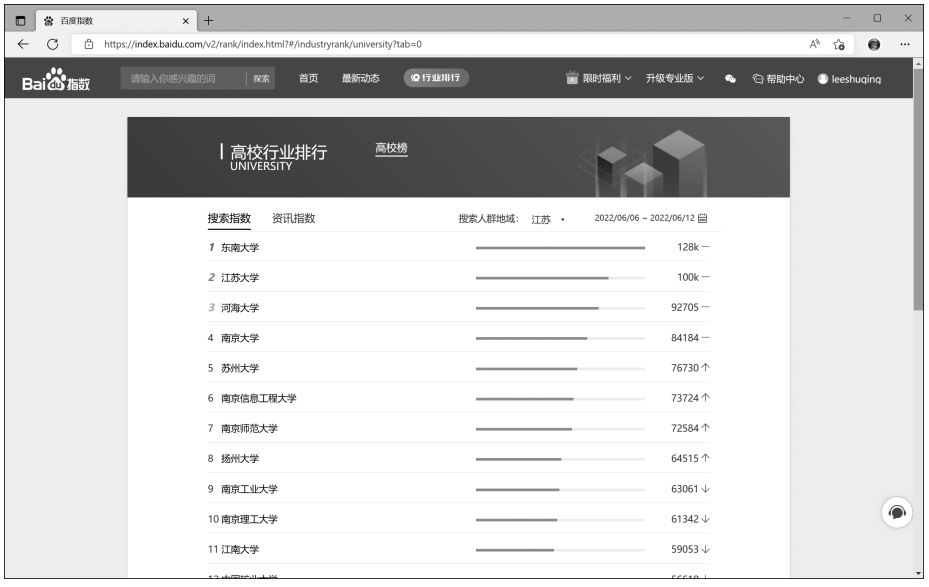

图 5.5　在百度指数中检索高校行业排行(截取于 2022-11)

图 5.6　百度司南的主页界面(截取于 2015-4)

具体功能包括：人群定义——系统推荐与自定义人群手段结合，司南锁定目标人群，对其网络搜索、浏览、属性行为数据进行分析；兴趣洞察——帮助企业了解指定消费者的具体兴趣点，为优化营销决策提供依据；搜索行为——记录并分析指定消费者的搜索行为，以帮助企业从搜索行为深入分析消费者隐含的主动意图，优化关键词策略；人口属性——包含年龄、性别、职业和学历等，挖掘比较不同指定消费者之间的人口特征差异，制定有针对性的营销方案；地域分布——分析指定消费者人群的地域分布，细化到市级，可依据不同的消费者偏好和媒体资源进行有差别的营销活动；媒体偏好分析——分析制定消费者人群对媒体类别、媒体站点的访问偏好性，从而优化线上营销的媒体选择，提升广告效果。

比如想了解“冰与火之歌”品牌的相关情况，可以在百度司南中检索，相关结果如图 5.7 所示。它提供了较为详细的品牌趋势、品牌份额、品牌认知和品牌发展指数等统计指标，同时还可以对人群进行分析，诸如人群兴趣点、搜索行为、人口属性、地域分布、媒体分析等。

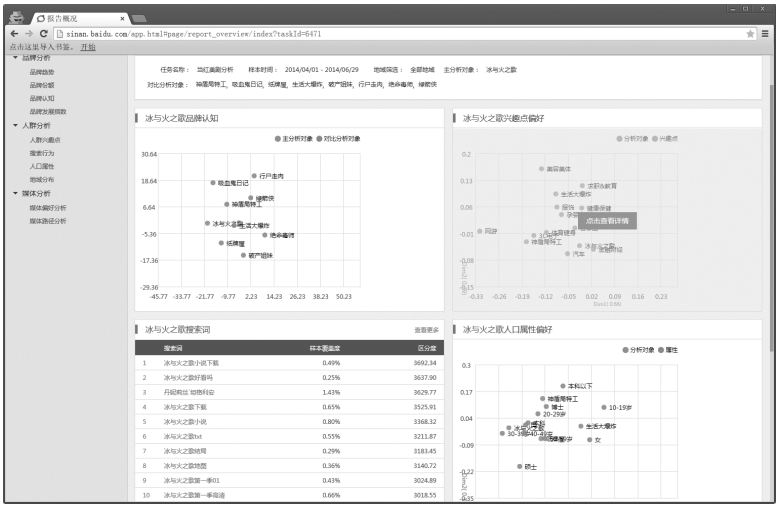

图 5.7　在百度司南查看相关品牌关键词的分析报告(截取于 2015-4)

5.2　Google 的信息分析决策功能

1. Google 趋势

Google 趋势集成了热榜检索和趋势检索两个主要功能，功能与百度热搜和百度指数类似，也利用用户的检索关键词信息，将某一专门领域内的相关内容，按照民众关注度的高低排列而成，主页如图 5.8 所示。

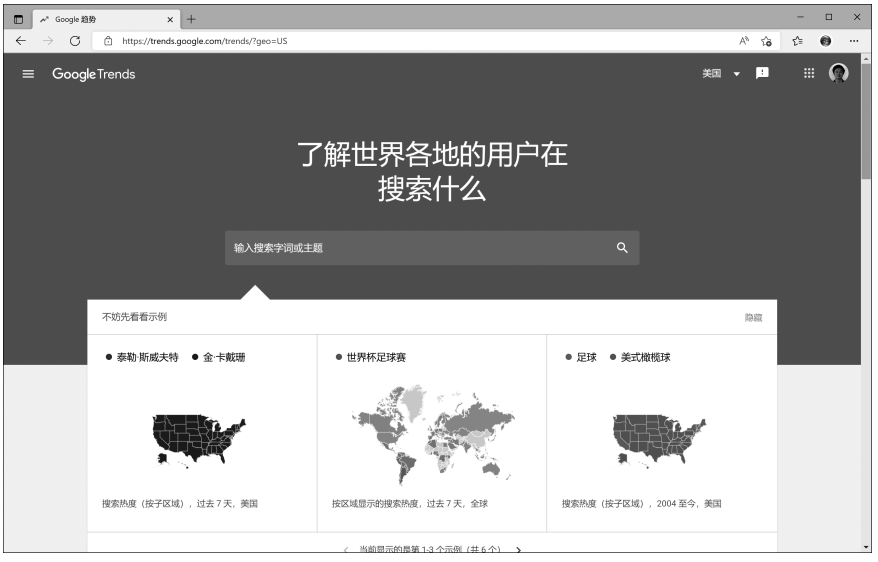

图 5.8　Google 热榜的主页界面(截取于 2022-5)

它提供了年度热词检索、趋势检索和 YouTube 趋势检索等功能，其中还进一步按照搜索、人物和运动员等指标进行了细分。但值得注意的是，由于 Google 退出了中国市场，所以中国国内信息的检索内容非常少，但是它所能提供的其他国家和全球总体情况检索非常有价值，如 2021 年全球总体热词检索结果如图 5.9 所示。

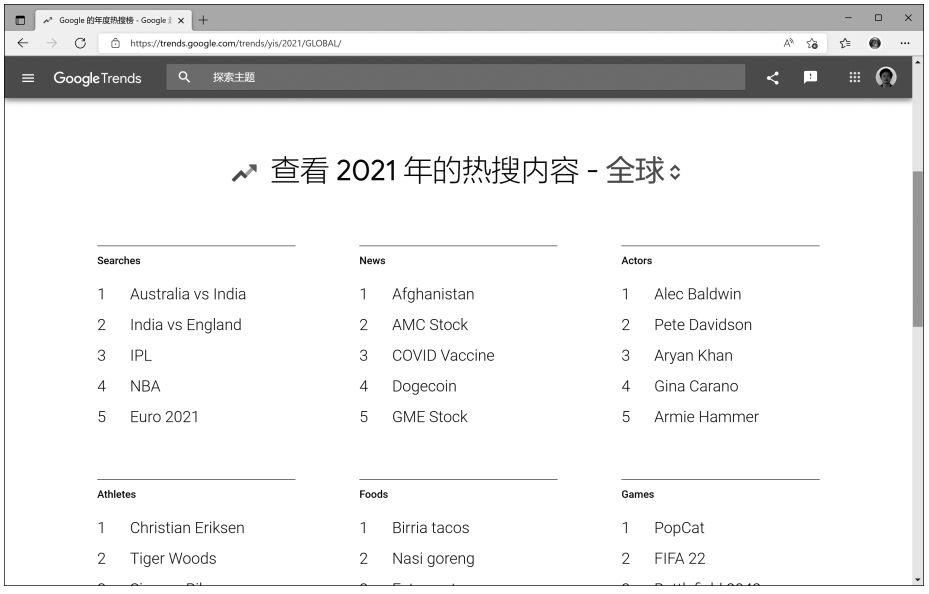

图 5.9　在 Google 热榜中检索 2021 年全球总体热词的结果页面(截取于 2022-5)

它甚至可以查看不到一小时的实时热词检索，如查看美国地区实时关注热点，如图 5.10 所示。

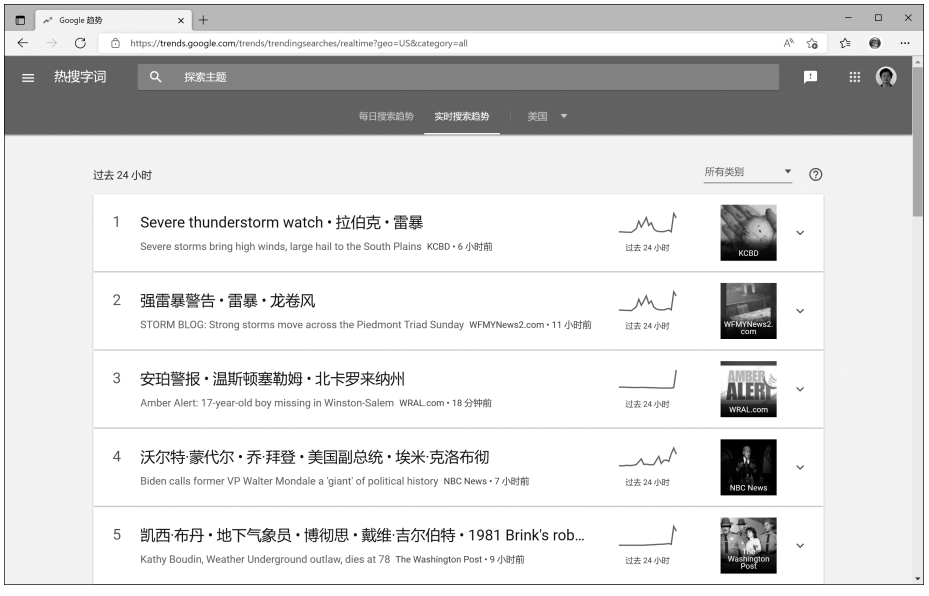

图 5.10　在 Google 热榜中检索 2021 年美国实时热词的结果页面(截取于 2022-5)

同时，它的探索(Explore)栏目还能进一步对用户自己输入的多个关键词做对比分析，

并从地区、时间、类目和搜索类型等方面来限定数据显示范围。该服务是于 2006 年推出的一项服务，它能根据用户检索关键词的次数，按照时间次序将其排列成按线性比例绘制的搜索量图表，具有良好的可视化效果。通过此类信息的分析，用户可以了解检索内容随时间变化的趋势，也就是相关检索关键词所反映的关注度变化趋势。它有两个特点：一是引入时间维度，这也是 Google 趋势最有价值的一个特点，使得用户可以按照时间序列进行分析；二是允许用户输入多个关键词，以此来观察它们在一段时间内的相对关注度程度。

如检索美国地区在 2021 年 nike 和 adidas 两个品牌的关注度对比情况，结果如图 5.11 所示。

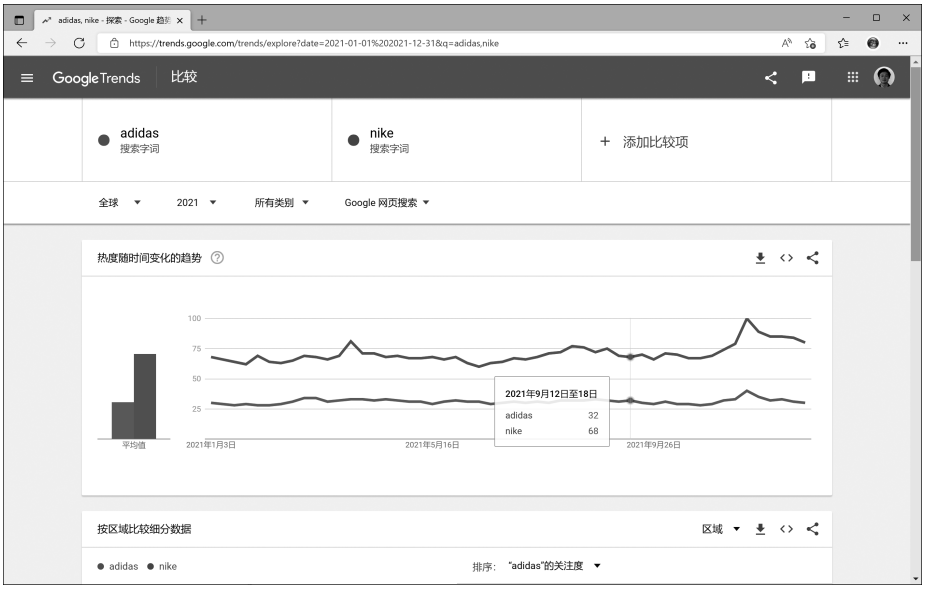

图 5.11　在 Google 热榜中检索自定义关键词相关的对比结果页面（截取于 2022-5）

2. 其他相关服务

利用信息分析功能可以实现一些很有价值的应用，如 Google 在 2008 年推出的利用互联网用户在搜索引擎中检索关键词统计信息得到的流感趋势搜索引擎（Flu Trends），它甚至比传统的流感监测系统还要快两周得到趋势预报结果。

Google 甚至还在一些重要时期专门针对重要事件推出相应的预测服务，如利用互联网用户的搜索结果来预测美国大选等。

5.3　评估站点的网络影响力（案例）

这个案例主要说明如何利用检索来评估不同站点之间的网络影响力。

所谓网站的影响力，在真实的检索需求中，可以间接表现为诸如网站网页的质量、网站知名度等类似概念。通常对于此类问题的判断，是根据多方面的综合因素才能做出相对正确的判断。这里以此案例来说明网络信息检索方法的应用。当然受限于网络信息资源本身检索的条件，这里也只能提供一种辅助参考的角度和思路，相关答案也仅供大家学习参考。

如果想比较大学的情况，如曼彻斯特大学和康奈尔大学，可以通过几款搜索引擎获取相

关信息检索结果，并进行如下必要地分析。

首先，可以按照互联网上存在的相关学校的域名网址数量，方法是直接搜索两个学校的域名，如 cornell.edu 和 manchester.ac.uk。可以发现两者区别已经出现，在百度 2022 年 5 月的检索结果中，康奈尔大学约为 673 万，曼彻斯特大学约为 579 万，康奈尔大学靠前，如图 5.12 所示。

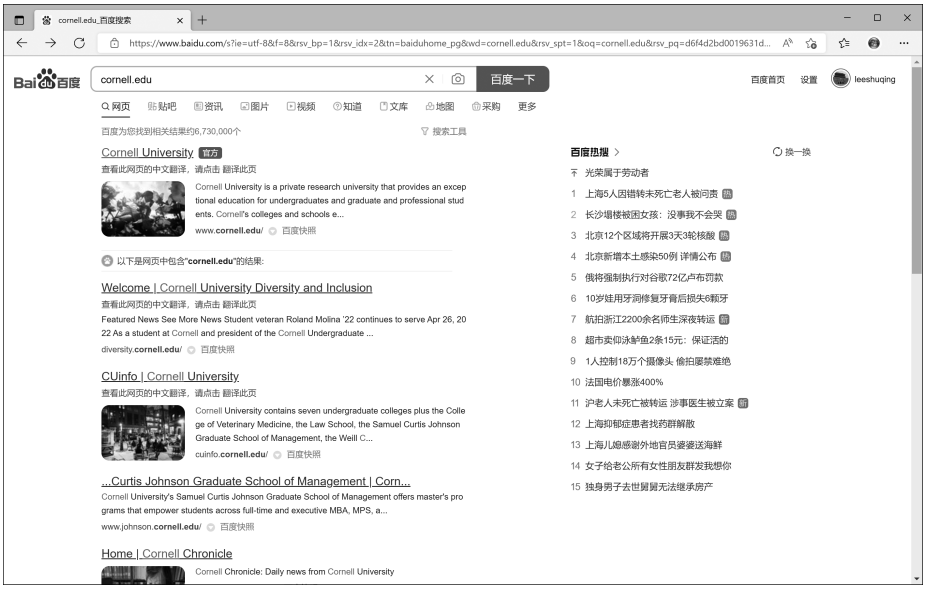

图 5.12　在百度中检索康奈尔大学的相关网址数量（截取于 2022-5）

这里需要说明两点：一是如果要考虑国外用户访问的规模，还需要使用诸如 Google 等搜索引擎，事实上如果要全面考虑还需结合更多搜索引擎的信息；二是在关键词选择上，这里去除了 www，这是因为在学校域名中，通常这个前缀可能表现为具体的学院或者下属机构名称。

这种功能也可以使用专门的外链查询来进行，即查询含有当前网站链接的网页。百度的 domain 字段检索可以实现这一功能，如 domain：cornell.edu 和 domain：manchester.ac.uk，可以发现，在百度 2022 年 5 月的检索结果中，康奈尔大学约为 674 万，曼彻斯特大学约为 580 万，康奈尔大学靠前。

其次，可以按照两个学校自己的网页规模来间接测度。方法是使用 site 限定域名进行检索，如 site：cornell.edu 和 site：manchester.ac.uk，可以发现，在百度 2022 年 5 月的检索结果中，康奈尔大学约为 84 万，曼彻斯特大学约为 33 万，康奈尔大学仍然靠前，如图 5.13 所示。

然后，还可以根据搜索引擎用户的搜索量，如 Google 趋势就可以测度比较各个搜索词语的搜索量，可以看到全球在近 5 年内，康奈尔大学的搜索关注度仍然比曼彻斯特大学更高，如图 5.14 所示。

考虑到用户极可能不会检索增加 the，而在曼彻斯特大学前面去除了 the。事实上，如果使用完整名称，差距更大。

如果做一个简单验证，通过查阅泰晤士 2020 世界大学排名，可以发现康奈尔大学排名

图 5.13　在百度中检索曼彻斯特大学的相关网页数量(截取于 2022-5)

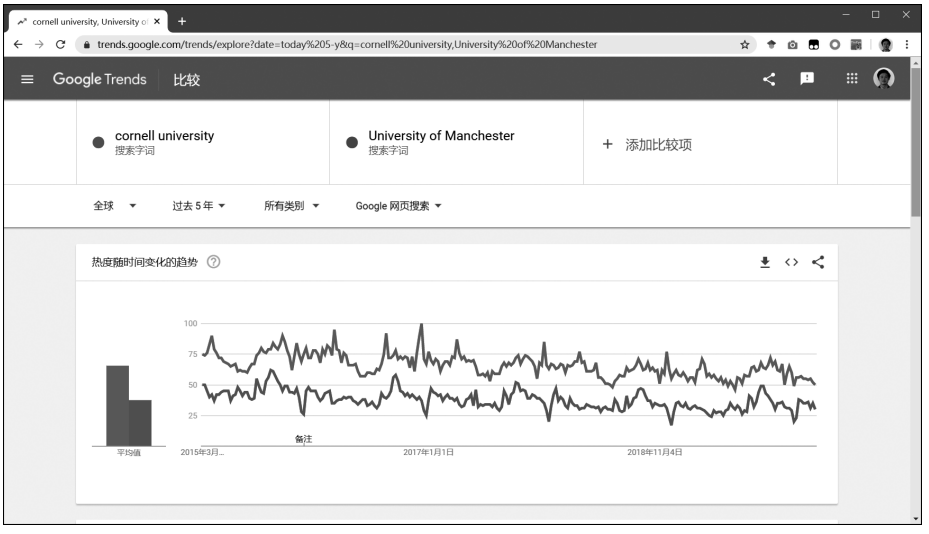

图 5.14　在 Google 趋势中比较两所高校搜索关注度(截取于 2022-5)

19,而曼彻斯特大学排名 57,和网络影响力基本呈现一定的相关性。

但是,我们也应该注意到这种方法只能作为一种参考,有时也未必总是正确,如直接使用大学名称作为检索关键词却呈现完全不一样的结果。何况搜索引擎返回的结果数量也是一种估算。这些其实很正常,事实上任何一种方法都不可避免地存在着误差。因此,我们应该综合比较多种方法,才能形成一种比较稳妥的决策依据。

这种利用搜索引擎检索信息的方法其实可以利用在很多其他方面,比如如何测度不同主题词语之间的相关度,传统的方法都是利用一些比较复杂专业的文本分析方法,但是这里可以使用类似的方法。

比如某“管理信息系统”,要与“Web 挖掘”“数据挖掘”“数据库”三个主题词比较相关

度，不妨直接利用这些词语的组合来检索试一下，可以明显发现“管理信息系统”与“数据库”关系更为紧密，而和其他两个词语的相关度则差很多，如表 5.1 所示，这基本符合专业认知。

表 5.1　在百度中检索不同词语组合的命中网页结果数量（截取于 2022-5）

管理信息系统 Web 挖掘	33,900,000
管理信息系统 数据挖掘	18,100,000
管理信息系统 数据库	100,000,000

当然，有效的测度必须要考虑更多的因素，比如综合多个搜索引擎结果，进行必要的语种综合判断等，有兴趣的读者可以自己多做思考。

5.4　练 习 题 5

1. 利用搜索引擎的信息分析功能，了解我国茶叶和咖啡的主要关注地区和趋势信息。

2. 百度指数中提供了两种用于分析网络主题关注度的指数，请分别说明概念及其含义。

3. 如何通过百度来推广企业的网站？举例说明。

4. 对比分析 Google 和百度在企业决策支持方面的不同。

5. 列举一或两个教材未提及的搜索引擎的企业决策支持功能。

6. 百度司南在决策分析上有哪些优点？

7. 结合自己的理解，说明什么是搜索引擎的信息分析与决策功能。

8. 利用搜索引擎的热词排行功能，查找网民关注化妆品品牌的 TOP 10。

9. 百度指数中的指数概况和热点趋势分别是指什么？

10. 上机实践练习：运用百度指数，分析你所关心的两种商品的趋势研究、需求图谱、舆情管家和人群画像。

11. 上机实践练习：搜索江苏、北京和上海网民的人物关注度排序和小说关注度排序。