

# 面向网络爬虫的网站优化策略

王晓楠 李杨 张海峰 张宇

(黑龙江省农业科学院农业遥感与信息研究所, 黑龙江哈尔滨, 150086)

**【摘要】**随着我国科学技术水平的提升,网络的发展速度逐渐的变快,在这一时代发展背景下,不管是网页的预处理数量,还是网页的搜索引擎品质都产生了很大的变化,其水平有着极为显著化的提高。在设计搜索引擎时期,设计人员需要对网页的便利性进行分析,进一步的去优化网络爬虫。本文主要就网络爬虫的特征进行探究,制定出较为完善且合理的网站优化措施,使得数据信息的预处理工作可以进展的更为顺畅。为了更为深入的探究面向网络爬虫的网站优化措施,本文以我国,某一农业企业网站优化为例,对企业网站的更新频率以及网站链接等要素进行综合性的探究,进一步的优化设计方案的内容。

**【关键词】**网络爬虫; 优化策略; 搜索引擎

网络爬虫是搜索引擎技术当中的一类核心性技术,其技术主要是以遍历策略为基准,借助网页链接来收集整合网页当中必要类的数据信息,同时把其数据信息下载存储到本地的相应设备上,使得数据信息的预处理程序以及脚本等的使用变得更为流畅。随着我国互联网技术的发展,社会各界已经开始注重网络爬虫技术的使用,并将其当做核心的搜索引擎技术。在众多的网络爬虫企业网站中,优化措施研究工作的开展已经成为了必然,是一种推广范围较广,形式多元化且成本节约的效果最为明显的一类方式。

## 1 网络爬虫的特征

网络爬虫的特征较为显著,首先,网络爬虫的程序具有一定的强壮性,执行力度也会比较强。在Web页和利用链接层面会凸显出爬行的良好性能,智能性以及自动性都比较明显。其次,网络爬虫可以把Web数据信息进行自动化的整合处理,对其进行简单性的存储操作,想要进一步的提升网络爬虫的性能,还可以适当的进行伸缩、分布等层面内容的补充,以此来达到完善性能的目的。

## 2 网络爬虫企业网站优化措施

### 2.1 网站导航

网络爬虫是用户们深入访问网站的重要组成部分,必须要保障网站结构的清晰程度,才可以使得网站的导航凸显出自身的最大效用,同时也可以给各个层次深入访问网站提供便利,其始终是网络爬虫的核心,需要对网站导航进行优化的设计。首先,需要使用文字链接去设置导航,文字是网络爬虫识别的关键性内容,但是其就flash和JS等内容的识别性会比较差,这主要是因为flash和JS内容识别的难度会比较高,所以,要借助文字链接的形式,合理的设置好网络导航,优化企业网站。其次,要设置导航的关键词,控制好目标关键词的长短,避免其和首页产生矛盾冲突等的问题。合理的使用锚文字,按照由左到右的顺序,把锚文字应用到导航关键词的设置工作中,凸显出锚文字的价值,若其栏目没有实际性的作用,那么就需要尽可能的少使用锚文字。最后,要合理的设置网站地图,网站地图和网站导航之间存在着一一定的相似性,所以,网站地图所产生的作用十分的重要。

不管是html还是xml,都可以应用主流搜索引擎。网络爬虫所无法处理的内容,都可以使用网站地图进行处理,解决好图片以及动态网页无法识别的问题。如果网站内的栏目以及内容数量比较大,且内容过于繁杂,那么就可以把网站地图融入到其内容,较好的处理并满足好用户们的深入性访问需求,完成网络爬行的抓取设置。

### 2.2 关键词

通过合理的设置关键词来分析该网站的主题以及核心的内容。首先,要合理的选择关键词,在选择关键词的时期,需要以用户们搜索的角度出

发,使得关键词和企业网站方向产品更加的具体化,使用一些针对性比较明显的词语。除此之外,想要避免同行之间所产生的竞争,就需要尽可能的少使用一些热门性的关键词。在实际操作时期,可以使用百度指数去分析当前网站内关键词的搜索量数值。

### 2.3 网站内容

网站的内容是网络爬虫的关键性内容,网络爬虫不管是在爬行时期,还是在抓取文件时期,都会对其内容进行相应的检测以及复制性的处理,这时会不再继续爬行。因此,对于网站内容来说,不仅要注意内容的丰富性,还应注重内容的创新,这样的网页被爬行的深度较高,而且对收录页面也较多。

### 2.4 Alt 属性

由于图片或flash在视觉效果方面具有一定的优势,当前大多数企业在网站设计中十分注重图片与flash的应用,为企业塑造良好的形象。但由于网络爬虫图片或flash信息识别具有一定难度。因此,用户对于存在图片或Flash信息的网站来说,并不能通过网络爬行来看这种网站,alt属性能够解决这一问题。由于Alt属性是用来对网页上的图片进行描述的,因此,将文字描述的形式引入Alt属性。

### 2.5 合理的链接

网络爬虫的主要路径是链接。因此,在网站建设过程中,注重页面栏目的互通性有助于实现网站各个页面的相互连接,防止死链的现象发生,为网络爬虫通过站内链接爬行网站提供保障。此外,由于一些大型论坛、门户网站空间博客更新速度快,网络爬虫的爬虫比较频繁,在这些站上留链接。

## 3 结语

面向网络爬虫企业网站优化策略的研究工作对于企业的发展来说至关重要,在分析该项策略的过程中,必须要进行网页抓取广度以及深度,并掌控好其所存在的各类关系,调整好网页的预处理工作量,进一步的提升搜索引擎的品质。在设计搜索引擎时期,工作人员需要对网页遍历策略进行探究,尽可能的优化企业网站的关键词以及网站的内容等,给企业网站的优化效果奠定一个坚实的基础。对我国某农业企业进行网站关键要素的分析以及优化设计,可以观察到,其网站在优化之后所取得的效果十分的显著。

**作者简介:**王晓楠,1984年生,女,专科,助理研究员,研究方向:数字农业、大数据采集与分析、系统研发、农业数据平台建设、信息技术在农业领域的应用等。

**基金项目:**黑龙江省农业科学院长期性基础性项目(项目编号:2019CQJC003);黑龙江省农业科学院院级课题项目(项目编号:2018YYF010)

### 参考文献

- [1] 穆喆.用云指建站把网站优化到极致[J].计算机与网络,2018(18):115.
- [2] 马玉.五大策略助你做好网站优化[J].计算机与网络,2017(9):20.
- [3] 侯建华.三大思路让网站优化文章写作锦上添花[J].计算机与网络,2017(18):33.
- [4] 焦大.企业网站优化人员培训要点[J].计算机与网络,2017(19):97.