网络引文的爬虫系统初探

摘要

随着互联网技术的高速发展，网络信息资源越来越受到人们的重视，成为人们获取信息和进行科学研究的主要信息来源。网络引文，即网络参考文献也逐渐受到众多学术工作者的青睐，并成为一种不可忽视的引文类型。统计分析网络引文，不仅可以了解各领域对网络引文的利用现状，而且还可以从中分析网络引文的来源网站分类、及网络信息使用习惯等诸多深层信息。

关键字：网络引文；引文分析；网络爬虫

Abstract

目录

1. 绪论
   1. 论文的研究背景及意义
   2. 网络引文爬虫系统的研究现状
   3. 论文的相关工作
   4. 论文的安排
2. 相关概念及技术背景
   1. 网络引文的概念
   2. 网络爬虫的概念
   3. 网络引文爬虫
      1. 网络引文的爬虫与通用爬虫的区别
      2. 网络爬虫的工作原理
   4. 本章小结
3. 网络引文爬虫系统ReferenceSpider的设计
   1. 网络引文爬虫系统ReferenceSpider的总体设计
   2. URL提取
   3. URL消重
4. 网络引文爬虫系统ReferenceSpider的实现
   1. 系统环境
   2. 测试报告
   3. 实验结论
   4. 本章小结
5. 总结与展望
   1. 本文总结
   2. 下一步工作

正文（前言、本轮、结论）

致谢

参考文献

1. 绪论
   1. 论文研究背景及意义
   2. 网络引文爬虫系统的研究现状
   3. 论文的相关工作
   4. 论文的安排
2. 相关概念及技术背景
   1. 网络引文的概念
      1. 引文

引文是现代学术论文的重要组成部分，它表征了本文献与被引用文献之间的引用关系。在学术研究过程中，任何一篇论文的写作，都离不开已有文献或资源的借鉴，这里的借鉴可以理解为我们通常所说的引用。引用的内容称为引文。

引文，又称参考文献，出现的形式有两种，一是作为附属内容置于文章或者论文的后面，二是作为文章内容的注释存在。引文是作者在撰写文章时由于参考并引用了他人的有借鉴意义的观点而形成的著录方式。

引文具有以下功能：

1. 体现作者的写作态度以及反映论文具有真实的科学依据；
2. 表明作者对于前人成果的尊重以及展示自己的创新和发展；
3. 为读者提供所引用参考文献以及作者简介的出处；
4. 表明学科的发展以及与其他学科之间的联系；
5. 为情报研究和文献计量研究提供依据。

因此，引文已经成为学术论文中不可缺少的内容。

网络引文分析和传统引文分析的比较 马晓佳

国外网络引文研究的现状及展望 杨思洛

* + 1. 网络引文

随着信息技术特别是网络技术的加速发展以及应用的普及，网络已经成为新的科学交流媒介，越来越多的学者通过网络来交流和获取信息、进行科学发现、保持和同行联系。网络环境下文献发生了很大的变化，形成了传统文献（print，P），和网络文献（web，W）两大类型。网络引文指的就是网络文献，即论文中引用的那些存在于网络中的文字内容。

网络引文分析和传统引文分析的比较 马晓佳

* 1. 网络爬虫技术