

宁波工程学院

毕业设计(论文)文献综述



设计(论文)题目:	论文题目		
	论文题目第二行		
学 院 名 称:	电子与信息学院		
专 业:	网络工程		
班 级:	班级		
姓 名:	姓名	学 号:	20653B101
指 导 教 师:	徐栋	职 称:	工程师

定稿日期: 年 月 日

论文题目论文题目第二行

摘要：网络爬虫具有自动提取网页信息的能力，但存在着一些网页重复使用关键词以提高其搜索排名的问题。因此，我们提出了一种自适应的网页排名算法来构建一个爬虫系统，以解决上述问题。具体来说，我们根据抓取的网页访问关系生成一个关系矩阵，然后根据网页的数量迭代生成一个概率矩阵，最后将抓取的网页按照计算出的权重从高到低进行显示。此外，我们提出用锚文本的一致性来控制 PAGE-RANK 的迭代过程。该系统使用 Python 语言来实现网络抓取的功能。实验结果表明，该系统在数据收集方面具有很高的速度。与 Hints 和经典的 Page-rank 爬虫系统相比较，结果表明，所提出的方法在精确度和召回率方面都有优势。

关键词：ZJUTThesis，外文翻译，文献综述、开题报告， X_gL^AT_EX 模板

参考文献

- [1] Yong S, 王译霄, 宣晓婷, et al. 基于机器学习联合网络爬虫算法的果汁鉴伪技术的研究进展 [J]. 华中农业大学学报, 2022, 41(1):269–276.
- [2] 石经海, 苏桑妮. 爬取公开数据行为的刑法规制误区与匡正——从全国首例”爬虫”入刑案切入 [J]. 北京理工大学学报（社会科学版）, 2021, 23(4):154–164,172.
- [3] 戴均豪. 基于 NLP 技术的地质图向量化方法及其找矿预测应用——以甘肃寨上-马坞矿集区为例 [D].[S.1.]: 吉林大学, 2021.