|  |
| --- |
| 武汉大学logo  **研究生学位论文** |
| 题 目： 基于Neo4j图数据库的研究团体搜索系统实现  学 院： 计算机学院  专 业： 计算机科学与技术  学 号： 2019282110194  姓 名： 陈小龙  导师姓名： 祝园园  导师职称： 副教授 |

目 录

[摘要 3](#_Toc64850479)

[Abstract 3](#_Toc64850480)

[1绪论 4](#_Toc64850481)

[1.1 研究背景 4](#_Toc64850482)

[1.2 国内外研究现状 4](#_Toc64850483)

[1.3 本文主要工作 4](#_Toc64850484)

[1.4 论文组织结构 4](#_Toc64850485)

[2 相关技术与概念 4](#_Toc64850486)

[2.1 MVC 4](#_Toc64850487)

[2.2 Restful 4](#_Toc64850488)

[2.3 社区搜索 4](#_Toc64850489)

[2.4 NoSql 4](#_Toc64850490)

[2.5 Vue 4](#_Toc64850491)

[3 研究团体搜索系统分析与设计 4](#_Toc64850492)

[3.1 系统需求分析 4](#_Toc64850493)

[3.2 系统整体框架设计 4](#_Toc64850494)

[3.3 社区搜索 4](#_Toc64850495)

[3.4 API模块 4](#_Toc64850496)

[3.5 接口设计 4](#_Toc64850497)

[3.6 Unmanaged 4](#_Toc64850498)

[3.5 Produce 4](#_Toc64850499)

[4 Neo4j算法扩展 5](#_Toc64850500)

[4.1 Equitruss 5](#_Toc64850501)

[4.2 Srimrank 5](#_Toc64850502)

[5 系统实现与结果分析 5](#_Toc64850503)

[5.1 数据集 5](#_Toc64850504)

[5.2 实验环境搭建 5](#_Toc64850505)

[5.3 实验与结果分析 5](#_Toc64850506)

[5.3.1 功能试验与结果分析 5](#_Toc64850507)

[6 结果与展望 5](#_Toc64850508)

[6.1 工作总结 5](#_Toc64850509)

[6.2 未来工作展望 5](#_Toc64850510)

[参考文献 5](#_Toc64850511)

[致谢 6](#_Toc64850512)

# 摘要

随着物联网、社交网络、分布式计算等技术的高速发展，互联网技术不断的改变着每一个人的生活。同时，计算能力、存储空间、网络带宽的发展更是日新月异。这就导致了人类各行各业积累数据的速度大大加快，积累的数据量更是以PB计。在可以预见的未来，这一趋势还将加快。

在数据量暴增的情况下，传统的关系型数据库不论是容量还是单机性能已经越来越难以满足数据存储的要求；更重要的是互联网时代产生的数据中很大一部分是非结构化数据，传统关系型数据库的存储模式难以满足大数据存储的扩展性和可用性。于是，非关系型数据库应运而生。

非关系型数据库主要有如下特点：模式自由、满足最终一致性(非ACID)、分布式、水平可扩展等。NoSql数据库可以按数据模型分成4类：键值存储库（如：Redis[1]、Memcached[2]）、列族数据库（如：Hbase[3]、BigTable[4]）、文档数据库（如：MongoDB[5]、Couchbase[6]）、图形数据库（如：Neo4j[7]、JanusGraph[8]、Hugugraph[9]）。其中图形数据库专门用来处理高度关联的数据，适用于推荐系统、社交网络、模式识别等领域。图形数据库具有高度的灵活性，并且支持复杂的图形算法，方便构建复杂的关系图。

在当今业界主流的图数据库中，Neo4j的社区最为成熟，应用也非常广泛。另外，Neo4j内置的算法库包含常用的图的算法[10]，包括路径搜索算法，如：Shortest Path、Single Source Shortest Path；中心性算法，如：PageRank、Betweenness Centrality；社区发现算法，如：Strongly Connected Components、Label Propagation。上述算法在实际生产中有着广泛的应用，但与此同时，Neo4j支持的算法也存在着较大的局限。特别地，对于社区发现算法而言，近年来有了新的发展，很多优秀的思想和改进算法被提出；相较于现有的算法，新的算法在某些方面有着更强的表现，遗憾的是这些新的算法却不被Neo4j支持。除此之外，在数据可视化方面Neo4j的功能仍然比较单一，不够灵活，这也是Neo4j的一个较为明显的缺点。

本课题旨在，基于Neo4j实现一个图数据处理系统，功能包括前端数据可视化、后端数据处理、图算法扩展。该系统的最终目标是扩展Neo4j的算法，并使用扩展的算法处理数据，最终在前端进行可视化展示。若上述目标实现，将提供一个相对完整的图数据从存储到计算再到展示的解决方案，并且在一定程度上解决Neo4j在数据可视化和算法扩展性方面存在的问题。

# Abstract

aasdasdasdasdasdassd

# 1绪论

## 1.1 研究背景

## 1.2 国内外研究现状

## 1.3 本文主要工作

## 1.4 论文组织结构

# 2 相关技术与概念

## 2.1 MVC

## 2.2 Restful

## 2.3 社区搜索

## 2.4 NoSql

## 2.5 Vue

# 3 研究团体搜索系统分析与设计

# 

## 3.1 系统需求分析

## 3.2 系统整体框架设计

## 3.3 社区搜索

## 3.4 API模块

## 3.5 接口设计

## 3.6 Unmanaged

## 3.5 Produce

# 4 Neo4j算法扩展

# 

## 4.1 Equitruss

## 4.2 Srimrank

# 5 系统实现与结果分析

# 

## 5.1 数据集

## 5.2 实验环境搭建

## 5.3 实验与结果分析

## 5.3.1 功能试验与结果分析

5.3.2 性能试验与结果分析

# 6 结果与展望

# 

## 6.1 工作总结

## 6.2 未来工作展望

# 

# 参考文献

[1]Redis Labs.Redis官网[EB/OL].https://redis.io/,2015-06.

[2]Dormando.memcached官网[EB/OL].https://memcached.org/,2015-08-03.

[3]Apache Hbase.The Apache Software Foundation[EB/OL].https://hbase.apache.org/,2007.

[4]Fay Chang,Jeffrey Dean,Sanjay Ghemawat.Bigtable: A Distributed Storage System for Structured Data[EB/OL].https://research.google/pubs/pub27898/,2006.

[5]Eliot Horowitz.The database for modern applications[EB/OL].https://www.mongodb.com/cloud/atlas,2020.

[6]Couchbase.couchbase官网[EB/OL].https://www.couchbase.com/,2011.

[7]Neo4j.Neo4j官网[EB/OL].https://neo4j.com/,2020.

[8]janusgraph.Distributed, open source, massively scalable graph database[EB/OL].https://janusgraph.org/,2017.

[9]王二铁.百度安全开源大规模图数据库HugeGraph[EB/OL].https://zhuanlan.zhihu.com/p/41240429,2018-08-03.

[10]Neo4j.the Neo4j Graph Data Science library[EB/OL].https://neo4j.com/docs/graph-data-science/current/algorithms/,2020.

[11]DB-Engines.DB-Engines Ranking of Graph DBMS[EB/OL].https://db-engines.com/en/ranking/graph+dbms,2020-11.

[12]Jermy Li.图数据库选型比较：Neo4j、JanusGraph、HugeGraph[EB/OL].https://blog.csdn.net/javeme/article/details/105000288,2020-03-20.

[13]Jermy Li.图数据库选型比较[EB/OL].https://blog.csdn.net/javeme/article/details/105000288,2020-03-20.

[14]Will Lyon.NBC News Analyzes Hundreds of Thousands of Russian Troll Tweets Using Neo4j[EB/OL].https://neo4j.com/case-studies/nbc-news/,2019.

[15]kbastani.Bank Fraud Detection by Kenny Bastani[EB/OL].https://github.com/neo4j-contrib/gists/blob/master/other/BankFraudDetection.adoc,2013-10-06.

# 致谢

感谢CCTV