

Southwest University of Science and Technology

本科毕业设计（论文）



基于NoSQL的网络文学监测分析系统优化

|  |  |
| --- | --- |
| 学院名称 | 计算机科学与技术学院 |
| 专业名称 | 软件工程 |
| 学生姓名 | 曾丹 |
| 学号 | 20134484 |
| 指导教师 | 杨春明 副教授 |

二〇一七年六月

基于NoSQL的网络文学监测分析系统优化

摘要

网络文学监测分析系统是一款具有统计分析网络文学作者作品地域分布、性别分布、类型分布、排行分布以及评论词云等功能的系统，能够通过该系统清晰明了的洞察网络文学的发展趋势。目前每日可采集到相关作者作品信息上千条，随着时间的推移，数据的累积，现已采集作者数达30万，作品数近30万，评论数165万。由于该系统采用的传统关系型数据库Mysql，关系型数据库需提前设计表结构以及表之间关系，导致部分查询需要至少3个表的联表查询，导致部分统计分析接口响应时间超过5分钟，十分影响用户体验。针对上述问题，结合当下已成熟的应对海量数据的解决方案，决定更换数据库并修改逻辑，采用NoSQL数据库存储数据。NoSQL数据库即非关系型数据库的总称，分别有Key-Value键值对形式、面向列、面向文档的存储形式存储数据，其中主流NoSQL数据库有Hbase、Redis、MongoDB等数据库。Hbase功能强大，但相较于其他数据库略微笨拙；Redis是一款内存数据库，若内存增长过快需定期清理，更适合于数据变化快的系统；MongoDB为最类似于关系型数据库的NoSQL数据库，十分适合注重海量数据访问效率提升的系统。经过比对，基于MongoDB数据库的系统优化，页面平均响应时间2秒，比对优化前后的页面响应速度，性能有了明显的提升。

首页1min以上，近期更新图标30+s，top作者20s+

作者统计页面，统计25s+；作者列表，25s+；

作品详情页面，

关键词：网络文学监测分析系统，NoSQL数据库，MongoDB数据库

**Base on NOSQL database on the monitoring and analysis of network literature system optimization**

**ABSTRACT**

The monitoring and analysis of network literature system which has the functions of statistical analysis of the geographical distribution, gender distribution, type distribution, ranking distribution and commentary clouds of the author's works, which can clarify the development trend of network literature through the system. At present, the daily collection of relevant works of thousands of works, with the passage of time, the accumulation of data, has been collected the number of 300,000, the number of works nearly 300,000, the number of comments 1.65 million. As the system uses the Mysql database, the relational database needs to design the table structure in advance and the relationship between the table, resulting in some inquiries need at least three tables of the table query, resulting in part of the statistical analysis of the interface response time of more than 5 minutes, Affect the user experience. In response to the above problems, combined with the current solution to deal with massive data, decided to replace the database and modify the logic, using NoSQL database to store data. NoSQL database that is non-relational database of the general term, respectively, Key-Value key pairs of form, oriented, document-oriented storage form of storage data, which mainstream NoSQL database Hbase, Redis, MongoDB and other databases. Redis is a memory database, if the memory growth too fast to be cleaned regularly, more suitable for data fast changes in the system; MongoDB is the most similar to the relational database NoSQL database, but the database is not the same as the database, Very suitable for focusing on massive data access to improve the efficiency of the system. By comparison, based on the MongoDB database system optimization, the page average response time is 2 seconds, compared to the optimization of the page before and after the response speed, performance has been significantly improved.

**Key words:** The monitoring and analysis of network literature system, NoSQL, MongoDB

1. 绪论

1.1课题的背景

伴随信息时代的高速发展，数据以爆炸式方式增长。面对海量数据的存储与处理，传统关系型数据库已无法满足用户需求，甚至严重影响系统性能以致用户体验呈直线下降。随着Internet网的普及，人们的生活方式也发生了巨大变化，由现金支付转为网上支付，由纸质书转为电子书等等。

网络文学监测分析系统是一款统计分析网络文学作者作品相关信息的网站，以图表形式清晰明了的展现给用户，能够快速了解与掌握网络文学作者作品的现状与发展趋势。

本文基于传统关系型数据库Mysql为网络文学监测分析系统主要的制约因素研究在海量数据环境下产生的NoSQL数据库在该系统的应用，比对各NoSQL数据库优缺点以及适用场景，通过更换数据库及系统逻辑实现，旨在优化网络文学监测分析系统性能，提升系统的响应速度，使得用户体验有质的飞跃。

1.2课题的研究意义

由于网络科技的高速发展与移动设备的普及，基本人手一台手机，达到足不出户便能知晓天下事。小说阅读是当今人们生活中重要的娱乐方式之一，李克强曾指出：读书不仅事关个人修为，国民的整体阅读水准，也会持久影响到整个社会的道德水平。随着2015年“两会”的圆满结束，“全民阅读”的发展进程进一步加快。在移动互联网时代的大背景下，移动设备已逐渐取代了传统的实体书，成为了人们主要的阅读工具。据统计，截止2015年，网络文学用户规模已达3.5亿人，而纸质图书作为最“正统”的阅读工具，用户仅占15%，因此阅读移动化已经势不可挡，也正是由于网络文学作品的大量出产，所以掌握网络文学作品发展趋势以方便对其管控和了解用户需求是尤为重要也是极其必要的。而网络文学监测分析系统正是以简单有效的图表形式展现给用户，可更方便了解网络文学用户的需求以及网络文学作品发展态势。

1.2.1 必要性

网络文学监测分析系统的性能随着爬取的网络文学作者作品信息数量递增，出现明显的下降，经过几次维护与修改，仍然未能从根源彻底解决问题，导致系统使用者往往想要查看数据分析的图表需要花费大量的时间等待，甚至有时会超过用户等待时间的最大限度，导致系统用户渐渐减少，如果这样的情况不能得到及时有效的解决，此系统将会成为一款无人使用的系统，就如同一栋烂尾楼，无人问津。为避免网络文学监测分析系统成为一死系统，所以对系统的优化不能停滞不前，要从根源解决问题才是根本，经过问题的排查和之前的维护与修改经验，最终确定是使用传统关系型数据库Mysql存储数据的原因，Mysql数据库需要提前定义好表字段以及表字段间关系，由于要满足三范式，所以获取一条数据往往需要关联多张表数据，在大数据量的情况下，查询效率差，因此更换数据库，基于NoSQL数据库完成对系统的优化是极其必要的。

1.2.2 重要性

基于非关系型数据库的网络文学监测分析系统与基于传统关系型数据库的网络文学监测分析系统相比较，具有以下6个优点：

1. 将关系型数据库Mysql替换为非关系型数据库MongoDB能够从根源解决系统性能瓶颈问题；
2. MongoDB采用的是面向文档存储方式，不需要提前定义好表字段及表之间关系，内部使用的是BSON语法，能够更直观易理解；
3. MongoDB海量数据下，查询效率高效快速，进而解决系统性能低下问题；
4. MongoDB使用简单，可以支持JSON、mapreduce、正则表达式以及Javascript函数等查询方式；
5. MongoDB相较于Mysql，节省系统资源，对CPU等资源的耗费较小；
6. MongoDB具有高可扩展性，支持分布式集群，通过分片数据缩放处理理论可实现更高的吞吐量。

1.3 课题研究现状

1.3.1 NoSQL数据库研究现状

NoSQL数据库最早可追溯到1991年Berkeley DB的第一版Berkeley DB是一个键值存储(Key-value Store)的Hash数据库，适用于数据类型相对简单但需要极高的插入和读取速度的嵌入式场合。NoSQL从2007年开始蓬勃发展。Google和Amazon的工程师们分别发表了有关BigTable和Dynamo数据库的论文，描述他们己经在使用的新型数据库的设计思想。BigTable提出了列存储模型，证明了数据持久存储可以扩展到数以千计的节点, Dynam。则提出了最终一致性的思想以达到更高的可用性和扩展性。分布式缓存系统Memcached证明了内存分布式数据存储可以达到极高的性能。事实上，Google的BigTable, Amazon的Dynamo和分布式缓存系统Memcached这三者可以说是所有NoSQL数据库的启蒙者。很多公司和组织受到他们思想观点的启发，开发了自己的开源 NoSQL数据库产品，例如Hypertable就是BigTable思想的开源实现。

现如今，国外的NoSQL数据库品种己经超过100个。除了Google的BigTable,Amazon的Dynamo，主流的数据库还有FaceBook的Cassandra, lOgen的MongoDB, Amazon的SimpleDB, Apache的HBase, Redis, CouchDB, Tokyo Cabinet/Tyrant, MemcacheDB等等。这些数据库在互联网中应用广泛，表现优异。Cassandra的用户包括Twitter, Cisco, Urban等多个公司，已知最大的Cassandra集群有超过400个节点，存储数据量超越300TB最受欢迎的MongoDB数据库在电子商务、政府、数据存储、实时统计、社交网络、游戏等众多领域都应用广泛，其中著名的有github, source forge, Disney等，国内的淘宝网、大众点评和视觉中国也在使用MongoDB国内的 N0SQL数据库领域在2009年也开始活跃起来。人人网下属UGC团队参考Dynamo设计文档，开发了自己的NoSQL数据库Nuclear。豆瓣推出了开源数据存储系统BeansDB。盛大创新院也基于Tokyo Cabinet开发了TCDatabase开源数据库。相比国外的NoSQL数据库，国内的这几款数据库并不成熟，也没有在生产中有更广泛的应用。国内大型的互联网公司多数还是选择国外支持度高的NoSQL数据库，除了淘宝、大众点评、视觉中国选用MongoDB外，新浪微博选用Redis存储数据，盛大也推出了基于MongoDB的“盛大云”MongoIC的数据库云服务。

在NoSQL蓬勃发展途中，有着成功案例的同时，NoSQL数据库也有一些失败的经历。2010年7月Twitter提出暂停使用Cassandra替代MySQL存储feed的计划，原因是Cassandra仍缺少大并发海量数据访问的案例及经验，新产品需要一定稳定期。2010年10月，使用MongoDB的foursquare网站由于数据碎片化和监控不力的原因，经历长达11小时的两次宕机。随着不断的完善，NoSQL也愈加成熟，得到广泛应用。

1.3.2 MongoDB数据库研究现状

MongoDB是一个性能卓越的开源的NoSQL数据库。它的存储模式自由，能够存储文档型数据等结构数据，灵活的存储方式在存储数据方面对传统的关系型数据库是一种强大的冲击，它逐渐的替代了Key/V alue存储方式。MongoDB的这些优点省略了开发过程中很多不必要的麻烦和繁琐的操作。同时，MongoDB能够很好的应对大多数互联网应用对大数据量、高并发的需求。MongoDB的水平扩展能力，使应用具备了处理上千万甚至上亿级别的数据处理能力。这些性能极大的满足了Web2.0和移动互联网系统对大数据存储的需求，同时，它的即插即用的特点也从某种程度上大大降低了相关行业网站的运行和维护的成本。两年来，跟随着Web2.0的发展势头，国内对NoSQL数据库出现了热忱的局面，MongoDB作为NoSQL中性能比较优越的一个，自然的成为了最有影响的一个。很多的互联网公司都将MongoDB作为数据库存储的最佳选择，在大数据的存储上也给专家、学者们在科研上提供了很好的存储服务，在行业内创造了很多具有不可估量的价值的互联网、移动应用。

1.4课题目标和研究范围

1.4.1 目标

在当前Web2.0时代的互联网应用和大数据时代的海量数据管理的背景下，NoSQL数据库相较于传统的关系型数据库更适合处理高并发读写、海量数据处理、可用性和高扩展性的应用需求。本文目的在于针对NoSQL中表现优异的MongoDB在网络文学监测分析系统中展开研究与应用，使系统功能响应时间小于5秒。

1.4.2 研究范围

1. NoSQL技术综述：NoSQL非关系型数据库技术使用松祸合的数据模式，支持水平伸缩，拥有在磁盘和(或)内存中的数据持久化能力，支持多种“Non-SQL”接口来进行数据访问。NoSQL的数据模型包括Key-Value键值对、Key一结构化数据、Key一文档、列簇式存储和图结构存储。NoSQL支持复杂的查询、弱事务机制，支持冗余备份保证单机可靠性、多种数据同步方式实现多机可靠性，支持哈希分区和范围分区来进行分布式扩展，提供强一致性和最终一致。
2. MongoDB服务机制分析和应用优化：MongoDB是当前应用广泛的NoSQL数据库。本文全面研究了MongoDB应用及优化方法与原则：
3. 针对MongoDB面向文档的数据模型，提出反范式模式设计方法；
4. 根据MongoDB的自动分片机制，提出基于分区粒度、写扩展和查询隔离的分片键选择机制、分片集群部署策略和磁盘碎片化问题的解决方案；
5. 针对MongoDB基于简单模式分类、多计划并行的查询策略及其存在问题，提出了加入语义关联、细粒度模式分类的改进查询算法。
6. MongoDB应用：网络文学监测分析系统抓取存储多个大型小说网站的作者和作品信息，存储量大且增长快。本文根据网络文学监测分析系统的存储、查询、扩展以及复制冗余等需求，给出了MongoDB的解决方案，并与之前应用的Mysql数据库时的查询效率以及系统功能响应时间进行了对比。

1.5课题实施方案

本课题将采用先对系统优化总体规划，然后系统功能优化分模块实施，需求调研与应用开发紧密结合的技术实施方案。

1. 在海量数据情景下，对NoSQL数据库的应用进行充分的调研，分析并抽取完成系统优化必要流程和步骤，对系统优化进行可行性分析。
2. 学习和掌握实现网络文学监测分析系统优化所需的必要技术支持，指定系统优化实现的技术方案；
3. 找出实现系统优化所要解决的技术关键点，给出解决该技术难点的可行解决方案；
4. 完成系统优化总体设计，包括实现数据库的更换，数据库的设计与数据的迁移，提取系统功能等；
5. 完成系统优化详细设计，包括功能模块优化具体实施方案的选择；
6. 根据前期的总体设计和详细设计，逐步分模块实现系统功能优化；
7. 完善系统所有功能优化，并加强系统在功能、安全和网络适应性等方面的性能。

1.6 本章小结

本章对课题所要研究的基于NoSQL的网络文学监测分析系统优化进行了简要介绍，与使用传统关系型数据库相比较，给出了本课题的研究意义，并且介绍了非关系型数据库的研究现状，明确了本课题所要实现的项目目标和研究范围，给出系统优化的实施方案，为后期系统的优化明确了方向和思路。

1. 系统开发关键技术介绍（舍弃）

网络文学监测分析系统原本采用的是SSM框架开发，即Spring+SpringMVC+Mybatis的集成开发框架，也是目前主流的开发框架，相较于SSH更轻量级更易于搭建与开发。由于系统数据与日俱增，Mysql数据库性能直线下降，所以本文主要介绍MongoDB的shell命令、Spring与MongoDB的整合以及Java中如何操作MongoDB。

2.1 MongoDB的Shell命令

MongoDB简单操作（增删查改，正则表达式，索引）

MongoDB聚合操作（aggregate，mapreduce）

MongoDB之JavaScript函数（function（element））

2.2 Spring与MongoDB的整合

Xml方式

2.3 Java操作MongoDB

2.4 本章小结

1. 系统需求分析

3.1 系统介绍

网络文学监测分析系统主要通过爬取各大型网络小说网站数据，然后分析其中例如地域，性别，书籍类别，点击量，评论数，推荐数，评论等等有效信息，以地图、饼图、折线图及柱状图或列表形式将统计分析后的数据以最简单最直观的形式展现给用户，使其能一目了然，能够快速获取其想要获取的数据信息。同时，通过系统抓取的数据可以有效的观察各大类型小说的发展趋势，以及当今小说流行与发展的状况和趋势，并且该系统还能提供其他好处，其一能够结合统计分析结果做书籍推荐，其二能够结合统计分析结果分析出网络文学作品现状，可方便做宏观调控，以方便管理。

3.2 系统优化可行性分析

系统优化可行性分析可以有效的避免对系统优化的错误投资，是保证系统优化成功的基本前提。本课题对基于NoSQL的网络文学监测分析系统优化进行了以下几方面的可行性分析：

经济可行性：本系统的开发人员没学生，由双导师指导，传统关系型数据库Mysql及非关系型数据库MongoDB均为开源软件，故此系统优化无需开发经费，开发成本低廉。而本系统的实现可给该系统带来无限好处，一是避免系统成为死系统，造成前期投入的人力物力造成浪费，二是从根源解决系统功能响应缓慢问题，大大提升系统性能，三是使系统有了更好的扩展性，随着以后数据的膨胀，只需加机器及简单配置，四是突破自我，研究除已熟悉的关系型数据库外的额外非关系型数据库，加强自己能力的同时也是为学弟学妹们打好跨出已知知识领域的第一步。因而本系统优化是有必要也是有意义的，在经济上是可行的。

技术可行性：技术可行性是指分析系统优化的技术条件是否满足需要。本课题将实现基于NoSQL的网络文学监测分析系统优化，系统将采用的开发技术包括Java、Spring、SpringMVC、MongoDB、JavaScript、JQuery、Echart等技术。目前，这些技术已得到广泛应用，可以借助大量的学习资料，因而系统优化在技术上是可行的。

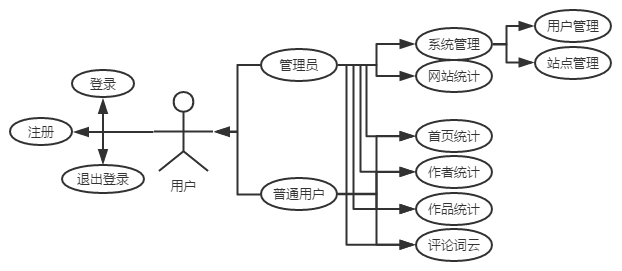
从以上两方面分析本系统优化是完全可行的，且系统操作简单且易于理解，但由于前期开发时数据量较小未过多注意系统性能问题导致随着数据的膨胀系统性能直线下降，严重影响用户体验，所以为了改善系统性能，优化是势在必为的。

3.3 系统中术语与缩写解释

**表3-1 术语/缩写及解释**

|  |  |
| --- | --- |
| 术语/缩写 | 解释 |
| SSM | Spring+SpringMVC+Mybatis开发框架 |
| MVC | Model-View-Controller将业务逻辑、数据、界面显示分离的思想 |
| JSON | 一种轻量级的数据交换格式 |
| Echart | 数据统计图表工具 |

3.4 系统优化功能性需求分析



**图3-1 系统用例图**

通过该系统用例图，可以看出网络文学监测分析系统共有2个角色：管理员及一般用户，已实现的功能如下表3-4-1所示：

**表3-2 功能及描述**

|  |  |
| --- | --- |
| 功能 | 功能描述 |
| 注册 | 用户需填写用户名密码、真实姓名、联系电话、邮箱及申请理由等信息注册账号 |
| 登录 | 已通过审核用户可以登录。用户数据存储于Mysql数据库，密码采取MD5加密算法，存储加密后密文 |
| 审核用户 | 仅管理员拥有的功能。管理员根据情况审核注册用户是否通过，并设定账号有效期天数 |
| 删除用户 | 仅管理员拥有的功能。管理员可以删除已失效或需要删除的用户数据 |
| 新增用户 | 仅管理员拥有的功能。管理员可以手动添加新用户 |
| 修改用户 | 仅管理员拥有的功能。管理员可以更改用户信息，如是否通过认证，用户名称，账号有效期等信息 |
| 系统参数修改 | 仅管理员拥有的功能。管理员可以根据需要修改页面初始大小值PageSize以及页面页脚显示信息 |
| 删除数据站点 | 仅管理员拥有的功能。管理员可以删除无用站点 |
| 新增数据站点 | 仅管理员拥有的功能。管理员可以新增爬取数据的数据站点 |
| 修改数据站点 | 仅管理员拥有的功能。若网站信息发生更改，管理员可以及时更改站点信息，以便不影响数据的抓取 |
| 网站统计 | 仅管理员拥有的功能。包括网站网址及其含有的文学作者、作品数量统计及最新更新时间，方便管理员管理网络文学监测分析网站，并且知道网站是否处于正常运行状态 |
| 首页统计 | 包括采集网站数、作者总人数、作者总数；作者区域分布百分比及地方人数地图、作者性别占比、Top10作者；作品Top类别占比、Top10作品；作者作品近期更新状况曲线图 |
| 作者统计 | 包括作者区域分别、性别占比、作者作品数统计、作者所有作品点击量评论数推荐数统计；作者详情列表，可按作品数、点击量、评论数、推荐数做升序降序排序功能 |
| 作者检索 | 包括按网站站点名称、按区域、按作者名称检索 |
| 作者所有作品分析 | 包括作者基本信息以及针对该作者的所有作品评论进行分词，提取出现最频繁有用的词语构建词云，可查看每部作品重要信息，每部作品的点击量占比及每部作品点击量、评论数、推荐数柱状图 |
| 作品统计 | 包括作品Top类别统计、作品点击量评论数推荐数统计、作品最新更新时间；作品详情列表，可按作品点击量、评论数、推荐数做升序降序排序功能 |
| 作品检索 | 包括按网站站点名称、按作品名称检索 |
| 作品分析 | 包括作品基本信息以及针对该作品的所有评论进行分词，以词频为准统计，分析出最重点且有用的词云信息，还有距最近更新时间往前推一周的该作品的点击量评论数推荐数曲线图 |
| 退出登录 | 需要切换账号或者其他情况下，可正常退出登录 |

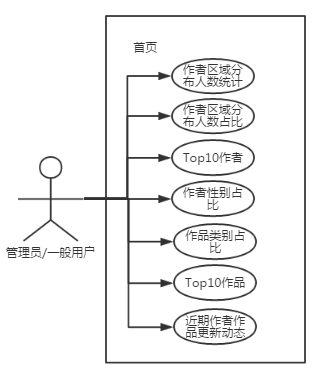
其中首页统计分析、作者首页统计分析及作者列表页面统计和作者详情页面统计分析、作品首页统计分析及作品列表页面统计和作品详情页面统计分析、网站统计页面是本次基于NoSQL数据库针对网络文学监测分析网站优化的重点，其他地方可以暂不做优化。

3.4.1 优化指标

**表3-3 描述及指标**

|  |  |
| --- | --- |
| 描述 | 指标 |
| 首页 | 共有8个功能，页面响应时间应小于5s |
| 作者统计页面 | 共有7个功能，页面响应时间应小于5s |
| 作者分页页面 | 共有5个功能，页面响应时间应小于3s |
| 作者详情页面 | 共有5个功能，页面响应时间应小于3s |
| 作品统计页面 | 共有5个功能，页面响应时间应小于5s |
| 作品分页页面 | 共有5个功能，页面响应时间应小于3s |
| 作品详情页面 | 共有3个功能，页面响应时间小于3s |
| 单条件查询 | 20ms以内 |
| 聚合条件查询 | 例如分组、排序等，500ms以内 |

3.4.2 首页

网络文学监测分析系统首页包括作者区域分布人数统计和人数占比，Top10作者，作者性别占比，作品类别占比，Top10作品及近期作者作品更新动态。用例图如下：  


**图3-2 首页用例图**

首页或作者统计页的作者区域分布人数统计用例规约描述如下：

**表3-4 作者区域分布人数统计用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者区域分布人数统计 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页或查看作者统计页面 |
| 前置条件 | 1a、管理员/一般用户在导航栏处点击“首页”或“作者统计”  1b、管理员/一般用户点击导航栏处“首页”，然后点击最热作者处的“更多” |
| 基本事件流 | 1、作者区域分布地图显示统计的作者人数 |

首页的作者区域分布人数占比用例规约描述如下：

**表3-5 作者区域分布人数占比用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者区域分布人数占比 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页 |
| 前置条件 | 管理员/一般用户在导航栏处点击首页 |
| 基本事件流 | 1、作者区域分布占比列表显示数据 |

首页的Top10作者用例规约描述如下：

**表3-6 Top10作者用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | Top10作者 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页 |
| 前置条件 | 管理员/一般用户在导航栏处点击“首页”，然后点击“最热作者” |
| 基本事件流 | 1、默认按作者点击量降序排序展示排名前10位作者信息，包括作者作品数、点击量、评论量、推荐量  2、点击“评论量/推荐量”，按作者评论量/推荐量降序排序展示排名前10位作者信息 |

首页或作者统计页的作者性别占比用例规约描述如下：

**表3-7 作者性别占比用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者性别占比 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页或查看作者统计页面 |
| 前置条件 | 1a、管理员/一般用户在导航栏处点击“首页”或“作者统计”  1b、管理员/一般用户点击导航栏处“首页”，然后点击最热作者处的“更多” |
| 基本事件流 | 1、正常显示男女和未知性别比例的饼图 |

首页的作品类别占比用例规约描述如下：

**表3-8 作品类别占比用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作品类别占比 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页 |
| 前置条件 | 1a、管理员/一般用户在导航栏处点击“首页”或“作品统计”  1b、管理员/一般用户在导航栏处点击“首页”，然后点击最热作品处的“更多” |
| 基本事件流 | 1、默认统计前9种类别占比，其余按其他类别归并统计 |

首页的Top10作品用例规约描述如下：

**表3-9 Top10作品用例规约描述**

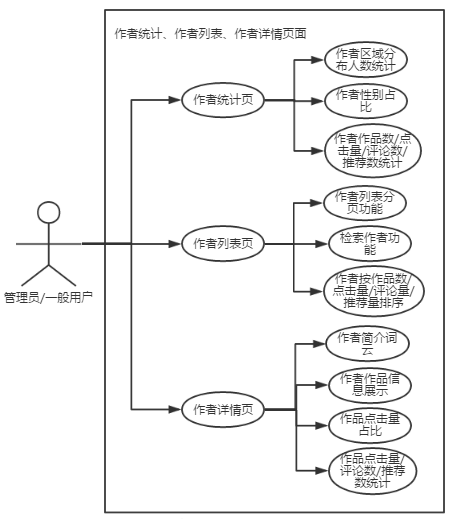
|  |  |
| --- | --- |
| 用例名称 | Top10作品 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页 |
| 前置条件 | 管理员/一般用户在导航栏处点击“首页”，然后点击“最热作品” |
| 基本事件流 | 1、默认按点击量降序排序展示排名前10个作品信息，包括作品点击量、评论量、推荐量  2、点击“评论量/推荐量”，按作品评论量/推荐量降序排序展示对应排名前10个作品信息 |

首页的近期作者作品更新动态用例规约描述如下：

**表3-10 近期作者作品更新动态用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 近期作者作品更新动态 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看首页 |
| 前置条件 | 管理员/一般用户在导航栏处点击“首页” |
| 基本事件流 | 1、按日期显示作者/作品更新数 |

3.4.3 作者统计、作者列表及作者详情页



**图3-3 作者统计、作者列表及作者详情页用例图**

作者统计页的作者作品数/点击量/评论量/推荐量统计用例规约描述如下：

**表3-11 作者作品数/点击量/评论量/推荐量统计用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者作品数/点击量/评论量/推荐量统计 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者统计页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作者统计”  1b、管理员/一般用户点击导航栏处“首页”，然后点击最热作者处的“更多” |
| 基本事件流 | 1、正常显示作者作品数/点击量/评论数/推荐数分布的柱状图 |

作者列表页的作者列表分页功能用例规约描述如下：

**表3-12 作者列表分页功能用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者列表分页功能 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者列表页面 |
| 前置条件 | 管理员/一般用户点击导航栏处“作者统计”，然后点击“所有作者” |
| 基本事件流 | 1、默认加载前三页数据，默认按作者总点击量降序排序  2、点击“上一页/下一页”，加载上一页/下一页数据  3、点击页面大小下拉框，选择每页展示数据的数量，页面展示数量改变  4、输入跳转页面，点击“跳转”，页面跳转 |

作者统计页的检索作者功能用例规约描述如下：

**表3-13 检索作者功能用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 检索作者功能 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者统计页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作者统计”  1b、管理员/一般用户点击导航栏处“首页”，然后点击最热作者处的“更多”  2a、点击“作者”输入框，输入作者名称，点击“检索”  2b、点击“区域”输入框，输入区域名称，点击“检索”  2c、点击网站列表，选择任一网站，点击“检索” |
| 基本事件流 | 2a、显示模糊匹配作者名称的作者信息  2b、显示模糊匹配作者区域的作者信息  2c、显示选择网站下的作者信息 |

作者列表页的作者按作品数/点击量/评论量/推荐量排序用例规约描述如下：

**表3-14 作者按作品数/点击量/评论量/推荐量排序用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者按作品数/点击量/评论量/推荐量排序 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者列表页面 |
| 前置条件 | 管理员/一般用户点击导航栏处“作者统计”，然后点击“所有作者” |
| 基本事件流 | 1、点击“作品数量/总点击量/总评论量/总推荐量”，作者列表按作品数量/总点击量/总评论量/总推荐量升序排序，再次点击“作品数量/总点击量/总评论量/总推荐量”按其降序排序 |

作者详情页的作者简介词云用例规约描述如下：

**表3-15 作者简介词云用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者简介词云 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者详情页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作者统计”，然后点击“所有作者”，任意选择一个作者进入作者详情页面  1b、管理员/一般用户点击导航栏处“首页”，然后点击“最热作者”，任意选择一个作者进入作者详情页面 |
| 基本事件流 | 1、正常展示作者简介词云 |

作者详情页的作者作品信息展示用例规约描述如下：

**表3-16 作者作品信息展示用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者作品信息展示 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者详情页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作者统计”，然后点击“所有作者”，任意选择一个作者进入作者详情页面  1b、管理员/一般用户点击导航栏处“首页”，然后点击“最热作者”，任意选择一个作者进入作者详情页面 |
| 基本事件流 | 1、正常展示作者信息  2、正常展示每部作品信息详情 |

作者详情页的作者作品点击量占比用例规约描述如下：

**表3-17 作者作品点击量占比用例规约描述**

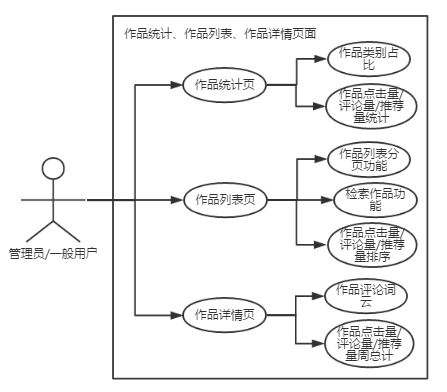
|  |  |
| --- | --- |
| 用例名称 | 作者作品点击量占比 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者详情页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作者统计”，然后点击“所有作者”，任意选择一个作者进入作者详情页面  1b、管理员/一般用户点击导航栏处“首页”，然后点击“最热作者”，任意选择一个作者进入作者详情页面 |
| 基本事件流 | 1、正常展示每部作品点击量占比的饼状图 |

作者详情页的作者作品点击量/评论数/推荐数统计用例规约描述如下：

**表3-18 作者作品点击量/评论数/推荐数统计用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作者作品点击量/评论数/推荐数统计 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作者详情页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作者统计”，然后点击“所有作者”，任意选择一个作者进入作者详情页面  1b、管理员/一般用户点击导航栏处“首页”，然后点击“最热作者”，任意选择一个作者进入作者详情页面 |
| 基本事件流 | 1、正常展示每部作品点击量/评论数/推荐数统计柱状图 |

3.4.4 作品统计、作品列表及作品详情页



**图3-4 作品统计、作品列表及作品详情页用例图**

作品统计页的作品点击量/评论量/推荐量统计用例规约描述如下：

**表3-19 作品点击量/评论量/推荐量统计用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作品点击量/评论量/推荐量统计 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作品统计页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作品统计”  1b、管理员/一般用户点击导航栏处“首页”，然后点击最热作品处的“更多” |
| 基本事件流 | 1、正常显示作品点击量/评论数/推荐数分布的柱状图 |

作品列表页的作品列表分页功能用例规约描述如下：

**表3-20 作品列表分页功能用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作品列表分页功能 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作品列表页面 |
| 前置条件 | 管理员/一般用户点击导航栏处“作品统计”，然后点击“所有作品” |
| 基本事件流 | 1、默认加载前三页数据，默认按作品总点击量降序排序  2、点击“上一页/下一页”，加载上一页/下一页数据  3、点击页面大小下拉框，选择每页展示数据的数量，页面展示数量改变  4、输入跳转页面，点击“跳转”，页面跳转 |

作品统计页的检索作品功能用例规约描述如下：

**表3-21 检索作品功能用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 检索作品功能 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作品统计页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作品统计”  1b、管理员/一般用户点击导航栏处“首页”，然后点击最热作品处的“更多”  2a、点击“作品”输入框，输入作品名称，点击“检索”  2b、点击网站列表，选择任一网站，点击“检索” |
| 基本事件流 | 2a、显示模糊匹配作品名称的作品信息  2b、显示选择网站下的作品信息 |

作品列表页的作品按点击量/评论量/推荐量排序用例规约描述如下：

**表3-22 作品按点击量/评论量/推荐量排序用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作品按点击量/评论量/推荐量排序 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作品列表页面 |
| 前置条件 | 管理员/一般用户点击导航栏处“作品统计”，然后点击“所有作品” |
| 基本事件流 | 1、点击“总点击量/总评论量/总推荐量”，作品列表按总点击量/总评论量/总推荐量升序排序，再次点击“总点击量/总评论量/总推荐量”按其降序排序 |

作品详情页的作品评论词云用例规约描述如下：

**表3-23 作品评论词云用例规约描述**

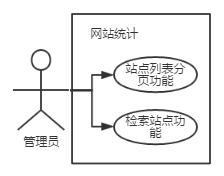
|  |  |
| --- | --- |
| 用例名称 | 作品评论词云 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作品详情页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作品统计”，然后点击“所有作品”，任意选择一个作品进入作品详情页面  1b、管理员/一般用户点击导航栏处“首页”，然后点击“最热作品”，任意选择一个作品进入作品详情页面 |
| 基本事件流 | 1、正常展示作品评论词云 |

作品详情页的作品点击量/评论量/推荐量周统计用例规约描述如下：

**表3-24 作品点击量/评论量/推荐量周统计用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 作品点击量/评论量/推荐量周统计 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看作品详情页面 |
| 前置条件 | 1a、管理员/一般用户点击导航栏处“作品统计”，然后点击“所有作品”，任意选择一个作品进入作品详情页面  1b、管理员/一般用户点击导航栏处“首页”，然后点击“最热作品”，任意选择一个作品进入作品详情页面 |
| 基本事件流 | 1、正常展示作品点击量/评论量/推荐量周统计 |

3.4.5 网站统计



**图3-5 网站统计用例图**

网站统计页的站点列表分页功能用例规约描述如下：

**表3-25 站点列表分页功能用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 站点列表分页功能 |
| 角色 | 管理员 |
| 用例说明 | 管理员查看网站统计页面 |
| 前置条件 | 管理员点击导航栏处“网站统计” |
| 基本事件流 | 1、默认加载前三页数据  2、点击“上一页/下一页”，加载上一页/下一页数据  3、点击页面大小下拉框，选择每页展示数据的数量，页面展示数量改变  4、输入跳转页面，点击“跳转”，页面跳转 |

网站统计页的检索站点功能用例规约描述如下：

**表3-26 检索站点功能用例规约描述**

|  |  |
| --- | --- |
| 用例名称 | 检索站点功能 |
| 角色 | 管理员/一般用户 |
| 用例说明 | 管理员/一般用户查看网站统计页面 |
| 前置条件 | 管理员/一般用户点击导航栏处“网站统计”，然后点击“作品”输入框，输入站点名称，点击“检索”，模糊匹配 |
| 基本事件流 | 显示检索匹配的站点信息 |

3.5 系统非功能性需求分析

非功能性需求是允许软件以满足用户的业务需求部分以外的要求。非功能性需求的是一个很容易被忽视的模块，但它对一个系统的开发和维护是非常重要的。系统非功能性需求应包括系统性能，可靠性，可维护性，健壮性等方面的问题。

本课题基于NoSQL的网络文学监测分析系统优化的非功能性需求定义如下：

实用性：系统满足网络文学作者作品统计分析要求；

易用性：界面美观，图表化易于理解，系统设计符合用户操作习惯；

可扩展性：系统优化结构设计合理，方便系统功能扩展；

可靠性：系统运行稳定，基本未发生故障；

健壮性：系统在发生故障时应有合理的处理和容错机制；

安全性：系统严格控制系统用户的使用权限及使用期限；

性能需求：系统数据量大，需要满足10万作者作品、300万评论时统计分析时间在5秒以内。

3.6 系统运行环境（表格形式,需添加硬件信息）

本课题所研究基于NoSQL的网络文学监测分析系统优化将SSM框架结构改为Spring+SpringMVC+Mybatis+MongoDB框架，使用Java技术开发，将传统关系型数据库Mysql替换为非关系数据库MongoDB，数据库安装于Linux服务器上，开发者系统为Window8.1，系统优化完成后部署在Linux服务器中。

3.7 本章小结

网络文学监测分析系统主要通过爬取各大型网络小说网站数据，然后分析其中例如地域，性别，书籍类别

1. 系统概要设计

4.1 系统优化策略

优化的原理

优化的方法

缓存技术；索引技术

4.2 系统结构设计

4.2.1 总体结构

4.2.2 功能模块

4.3 数据库设计

4.4 接口规范设计

4.5 本章小结

1. 系统优化详细设计与实现

5.1 系统具体实现技术

5.2 技术难点解决

5.2.1 数据的迁移

5.2.2 Spring与MongoDB的整合

5.3 系统模块设计与优化

5.3.1 文学作者统计分析模块

5.3.2 文学作品统计分析模块

5.3.3 文学网站统计模块

5.3.4 评论词云分析模块

5.4 本章小结

1. 系统测试

6.1 测试原则

6.2 测试环境

6.3 测试结果

6.3.1 单元测试

6.3.2 功能性测试

6.3.3 非功能性测试

6.4 本章小结

参考文献

[1] 沈姝. NoSQL数据库技术及其应用研究[D].南京信息工程大学,2012.

[2]刘卓. 基于NoSQL的空间数据云存储的研究[D].河南大学,2014.