

基于 Lucene 的网络创作及学习资料共享平台设计 *

袁健美,李旭,朱宁斌,黄晓凯

(湘潭大学 数学与计算科学学院,湖南 湘潭 411105)

摘 要:为向网络文学创作者提供开放的网络信息平台,本文引入两种广泛应用的搜索引擎技术——全文索引和目录索引,基于这两种技术提出网络创作及学习资料共享平台的设计思路,并详细叙述了网络创作平台的实现。该平台采用开放源码的软件工具包 Lucene,Struts2、Hibernate 框架进行设计制作,并通过全文搜索、目录搜索等技术实现网络作品的发布、查阅,以及学习和交流资源的信息共享。此类共享平台的设计,为文学用户提供了参与网络创作的渠道,使文学的创作和阅读、创作与交流实现了网络互动,既丰富了文学创作的形态,也促进了网络文学的发展。

关键词:全文索引;目录索引;Lucene;Struts2;Hibernate;网络创作平台

中图分类号:TP315

文献标志码:A

文章编号:1673-8454(2013)15-0024-03

一、引言

在网络载体的传播下,文学创作中增加了“网络文学”这一信息形式。网络文学,指以互联网为展示平台和传播媒介,借助超文本链接和多媒体演绎等手段来表现的文学作品、类文学文本及含有一部分文学成分的网络艺术品。随着互联网技术的迅速发展,网络文学爱好者的不断增加极大地促进了网络文学的创作发展,也增大了文学作品的共享需求。网络文学经历了从 BBS 到文学网站、博客、SNS 社区的平台形式上的转变,各种平台形式在不同时期形成各自的特点,分别扮演着不同的角色,见证了网络文学发展的整个历程。^[1]

由于网络文学具有读者可直接参与创作的特性,吸引了大量的爱好者和创作者,其中就有为数不少的大学生文学创作爱好者。近年来许多文学网站的创建和个人博客的兴起,为文学创作者和爱好者提供了众多的网络平台,如晋江原创网、起点中文网、小说阅读网、榕树下等等。但由于多数文学网站面向所有文学创作爱好者,使文学作品种类和数量过多,有时某些新创作的文学作品还没被用户搜索到就被其它类似的新作品给淹没了。如何在海量的网络文学中,搜索专门的文学作品,提高用户的搜索速率、效率,成为网络创作平台急需解决的问题。^[2]

Lucene 是一套用于全文检索和搜寻的开源程式库,提供了功能比较强大的应用接口,能够进行全文索引和搜寻。^[3]我们通过应用 Java 开发环境里的 Lucene 开源工具,进行 Java 资源检索库应用,设计并实现了以 Struts2 和 Hibernate 为框架的网络创作及学习资料共享平台。该平台实现了网络文学的创作和学习资料共享两方面功能,并改变传统的直接索引数据库模式,采用 Lucene 实现对数据库的索引并建立索引库,从而提高用户的搜索速率。

二、全文索引分析

1.全文索引模式简述

搜索引擎搜索的过程分为三个步骤:①捕获网页,每个独立的搜索引擎都有自己的网页捕获程序。网页捕获程序顺着网页中的超链接,连续地捕获网页。被捕获的网页被称之为网页快照,由于互联网中超链接的应用很普遍,理论上,从一定范围的网页出发,就能搜集到绝大多数的网页。②处理网页,搜索引擎抓到网页后,还要做大量的预处理工作,才能提供检索服务。其中,最重要的就是提取关键词,建立索引文件。其他还包括去除重复网页、中文分词、判断网页类型、分析超链接、计算网页的重要度/丰富度等。③提供检索服务,用户输入关键词进行检索,搜索引擎从索引数据库中找到匹配该关键词

* 基金项目:湖南省高等教育教学改革项目“深度开展大学生创新性实验计划与科研素质训练的研究与实践”(湘教通[2010]315);国家级大学生研究性学习和创新性实验计划“网络开放课程互动博客的排序算法分析与实现”(湘教通[2012]402);国家第一类特色专业建设项目“信息与计算科学”(高教函[2007]31号);湖南省研究生精品课程(KC2011B019)。

词的网页;为了用户便于判断,除了网页标题和URL外,还会提供一段来自网页的摘要以及其他信息。^[4,5]

2.全文索引与目录索引技术

全文索引作为目前应用广泛的搜索方式之一,主要从互联网提取各个网站的信息(以网页文字为主),建立数据库,并能检索与用户查询条件相匹配的记录,按一定的排列顺序返回结果。全文索引拥有自己的网页抓取、索引、检索系统,独立的“蜘蛛”(Spider)程序,能自建网页数据库,搜索结果直接从自身数据库中调用。目录索引将目录分类的网站链接列表,用户按照分类目录找到所需要的信息,不依靠关键词进行查询,因而搜索的准确性大大提高。^[6,7]

三、平台的设计与实现

1.系统结构设计

基于以上技术的分析,我们设计了基于Lucene的网络创作及学习资料共享平台。^[8,9]该平台以Struts2和Hibernate为框架,为网络文学创作者,如大学生文学创作者和爱好者,提供免费的网络创作和交流平台。该平台分为三个模块:用户服务层模块、网络文学创作模块、数据索引模块。系统整体结构如图1所示。

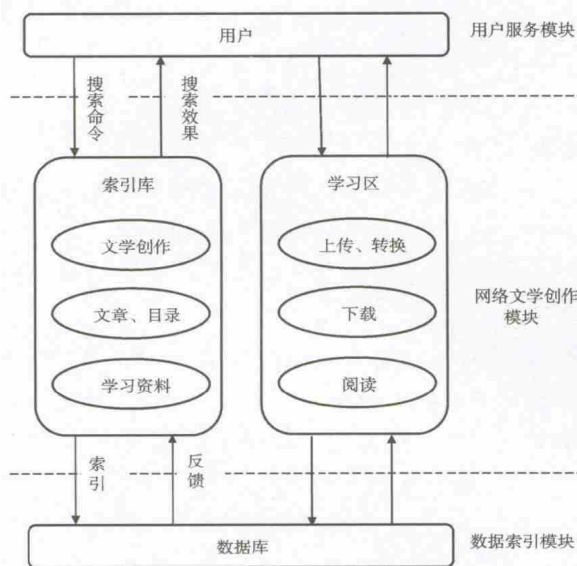


图1 系统结构图

(1)用户服务层模块

该层主要为进入文学平台的入口,也是为用户提供的接口。它基于B/S模式实现,系统对各类用户进行整合,使系统管理员、各用户可以跨平台访问该平台,也可利用该平台上传学习和交流资料,并共享给其他用户,提高用户在线阅读和下载学习资料的体验。

(2)网络文学创作模块

它是整个文学创作平台的核心部分。它与数据库的链

接由Struts2和Hibernate整合而来,Struts2是一个相当强大的Java Web开源框架,是一个基于POJO的Action的MVC Web框架。Struts2基于MVC架构,结构清晰,开发流程一目了然。文学创作模块将用户上传的TXT、Word、Excel、PPT、PDF等文档资料用OpenOffice+JodConverter转换成PDF文件再通过pdf2swf格式,最后通过FlexPaper实现在线浏览,实现类似百度文库的功能。同时,平台还支持多文件存储形式,将大文件划分成很多小文件,按每一章节对应一文本文件的方式存储。为方便创作者在修改某章节内容时不需修改所有章节内容,只需找到与之对应的文本进行修改,从而使系统的性能提高。

为了使用户新上传的文档能够及时被其他用户搜索到,而不用长时间等待,可采用Java多线程机制来解决上述问题。当用户上传一篇文档后,先存入数据库,然后开启另一线程来作为索引,这样就可以将用户上传文档和索引分开。建立索引对用户来说是不可见的。在用户不知不觉中就实现了索引。用户只需上传,其他事都是后台自动做的,对用户无影响,提高了用户体验。Office文档进行pdf和swf格式转换也是采用上述方法,对其分别开启线程,实现后台自动转换使用户无需等待。

(3)数据索引模块

网络文学创作平台采用Hibernate。Hibernate是JDBC的轻量级的对象封装,它是一个独立的对象持久层框架。它的工作原理是通过配置文件在Java对象和数据表之间建立起一个映射关系,只需要操作一些基本类,而达到使用数据库的目的,大大减少操作数据库的工作量。Hibernate的灵活性在于,它通过配置文件可以抽象底层具体的关系数据库,也可以有选择地将数据库操作预定义在配置文件中,从而可以很好地将持久层和逻辑层进行隔离。

2.系统平台实现

如图2所示,系统满足了用户进行文档搜索的要求,改进了现有文档搜索的方式,通过“Lucene+庖丁”实现对数据库的索引和分词并建立索引库,实现直接对索



图2 系统实现图

初中信息技术网络课程资源建设与应用研究报告

殷立斌,吴继承

(新华中学,广东 深圳 518109)

摘要:本课题重点研究在课堂教学实际需要的基础上,根据教材构建一个适合教师教学和学生自主学习的网络课程资源平台,通过基于平台的网络课程教学实践,归纳总结出初中信息技术课程教学的一般性规律,为信息技术课堂教学模式的改革提供了探索性的经验。

关键词:信息技术;网络课程;教学平台

中图分类号:G250.73

文献标志码:A

文章编号:1673-8454(2013)15-0026-05

2008年9月,我校向区教科所申请了区教育科研课题“初中信息技术网络课程资源建设与应用研究”。课题开题后,我校根据教学实际建设了网络课程资源平台,并在信息技术课堂上全面应用。在区教科所专家的关心和指导下,课题组全体成员扎实深入地开展了课题研究

和探索,达到了预期研究目标。

一、初中信息技术网络课程资源建设与应用研究背景及意义

1.信息技术课堂教学现状与课题的提出

在教育部“十五”课题中,初中学科资源建设及研究

引库中的文档进行索引以加快搜索速率,从而提高搜索效率。通过网络文学创作平台,各类用户在进行网络文学创作和学习资料共享时,利用索引库进行文档索引,达到提高用户搜索速率。同时,实现文档的上传和下载,平台提供各类文档格式的阅读,并在搜索输入框中使用AJAX实现搜索下拉提示框功能,提高用户体验。

基于Lucene的网络创作及学习资料共享平台,是大学生交流文学作品的信息应用环境,为大学生爱好者提供了一个文学交流的开放平台。用户通过平台提供的文档检索、上传和下载等服务,能方便快捷地在系统中检索到自己感兴趣的文档,相比以前的查找、搜索,减少了用户等待时间,提高了系统检索效率。各类用户通过网络文学创作平台提供的在线交流和在线阅读服务,方便了在线学习,从而促进了文学作品的传播。

四、结束语

网络创作平台主要采用将全文索引和目录索引相结合的方式实现搜索引擎应用,基于Struts2和Hibernate的框架系统,是一款面向大学生进行文学创作的开放信息和共享平台。随着互联网技术和Web2.0的不断发展,检索系统的优化设计成为越来越重要的需求。搜索引擎对于简化用户操作、节省系统资源、促进整个搜索体系

的发展具有重要现实意义。●

参考文献:

- [1]王莉,张延松.新世纪文学之网络文学研究述评[J].沈阳师范大学学报(社会科学版),2008(2):95-97.
- [2]刘彦平.关于网络搜索引擎及其优化的讨论[J].电子商务,2011(4):64.
- [3]蔡兵,胡敏.基于Lucene2.0的书目搜索引擎设计[J].重庆图情研究,2009(1):60-63.
- [4]陈丹,郭伟青.搜索引擎技术分析与研究[J].计算机系统应用,2008(3):23-26.
- [5]赵丹,朱巧明.中文搜索引擎的现状与展望[J].计算机与农业,2002(6):42-44.
- [6]王承君.Web搜索引擎的关键技术分析[J].潍坊学院学报,2009(4):27-33.
- [7]王巍,曾剑平,吴承荣,张世永.基于网络搜索引擎的网络话题分析框架[J].计算机工程,2009(3):257-262.
- [8]文振威,秦晓.个性化搜索引擎的研究与设计[J].计算机工程与设计,2009(2):342-344.
- [9]张巍.基于PageRank算法的搜索引擎优化策略研究[D].四川大学,2005:79-83.

(编辑:王天鹏)