

基于SSH的多语种语言资源库管理系统研究

惠建新¹, 周杰¹, 张红卫^{2,3}, 吕波¹, 王凯¹

(1. 南京信息工程大学 电子与信息工程学院, 江苏 南京 210044;

2. 中国科学院 理化技术研究所, 新疆 乌鲁木齐 830011;

3. 中国科学院 研究生院, 北京 100039)

摘要:基于SSH技术的框架结构的思想,根据新疆多语种语言资源的特点和需求,研究与设计多语种语言资源库。该设计在传统的资源库管理系统技术的基础之上,运用SSH框架技术,显著提高了应用系统的开发效率,并且系统有较高的重用性和可扩展性。研究建立汉、英、阿、俄对照词汇和短语资源库,为多语种软件开发提供规范的软件界面和用户文档用语,减少重复翻译工作,提高翻译质量,具有重大意义;而且能使多语种软件产品符合阿拉伯语和俄语的使用习惯。多语种语言资源库管理系统的建设填补了多语种语言民族地区相关软件业领域建设的空白,对促进面向中亚西亚出口软件开发提供重要技术支持。

关键词:SSH;多语种语言资源库;IoC

中图分类号:TP311.5

文献标识码:A

文章编号:1673-629X(2010)06-0078-06

Research of Multilingual Language Resources Library Managing System Based on SSH

HUI Jian-xin¹, ZHOU Jie¹, ZHANG Hong-wei^{2,3}, LÜ Bo¹, WANG Kai¹

(1. College of Electronic & Information Engineering, Nanjing University of Information

Science and Technology, Nanjing 210044, China;

2. Technical Institute of Physics & Chemistry, Chinese Academy of Sciences, Urumqi 830011, China;

3. Graduated School of Chinese Academy of Sciences, Beijing 100039, China)

Abstract: According to the requirement and characteristics of multilingual language resources in Xinjiang, multilingual language resources library managing system is presented based on the SSH. On the basis of traditional language resources managing system techniques, the system adopts SSH technique, which eventually improves its applicability, interoperability and extendibility greatly. Study the establishment of Chinese, English, Arabic and Russian control words and phrases resources. Software development for multi-language software provides a standardized language interface and user documentation. Reduce duplication of translation work. Improve translation quality, with great significance. And make the multilingual software products comply with the use of Arabic and Russian customs. The building of the MLRLMS fills a gap on the software industry in multi-lingual language minority areas. It provides an extremely technical support in the promotion of the Central Asian export-oriented software development.

Key words: SSH; multilingual language resources library; IoC

0 引言

随着信息技术及其软件产业的快速发展,面向中

亚西亚的多语种出口软件开发是新疆软件产业发展的重要方向,得到自治区人民政府、国家科技部、自治区科技厅等各级政府部门、企事业单位、科研院所的高度重视和支持。面向中亚西亚的出口软件开发是以新疆多年来积累的多语种信息处理技术为基础,以国内特别是东部地区成熟的软件开发技术与产品为依托,以中亚西亚国家软件市场需求为目标,东联西出,发展具有自主知识产权的外向型软件产业。目前,多语种软件产品和服务的需求增长迅速,而多语种软件开发缺乏统一的标准规范、开发效率低、质量难以得到保证、

收稿日期:2009-05-25;修回日期:2009-09-07

基金项目:教育部留学归国基金项目(教外留[2007]1108号);新疆自治区科技攻关(含重大专项)和重点项目(200732143-1);江苏省高校自然科学基金项目计划项目(05KJD510122)

作者简介:惠建新(1982-),男,硕士研究生,研究方向为多语种信息处理、计算机应用软件;周杰,博士后,教授,博士生导师,研究方向为信息处理技术、通讯技术等。

不具有可持续更新服务等问题阻碍了多语种软件产业发展。对多语种语言资源的需求也越来越多,然而多语种语言资源库的研究在新疆地区目前尚处于起步阶段。

文中着重讨论了一种基于 SSH 的多语种语言资源库管理系统(Multilingual Language Resources Library Managing System,简称 MLRLMS)的实现。

1 多语种语言资源库定义及研究现状

多语种语言资源库是基于计算机技术与理论语言学研究的。语言资源是指语言信息处理用的语料库和语言数据库,以及语言词典和声电动画片等语言资料^[1]。语料库指的是为某一个或多个应用而专门收集的、有一定结构的、有代表性的、可以被计算机程序检索的、具有一定规模的语料的集合。多语种语言资源库的建设牵涉到语言学、计算机科学、概率统计学、文献学、版权学、管理学等多个学科,建设一个高质量的大型多语种语言资源库,是一个人力、物力、时间、金钱开销都很庞大的工程。多语种语言资源库的开发是一项软件工程,开发过程应前遵循软件工程的一般原则和方法^[2]。

多语种语言资源库在语言学研究领域有广泛的应用,它是语言应用研究的重要基础。语言资源库对多语种软件开发很重要,国外大型软件公司都有属于自己软件领域上的语言资源库。我国在此领域的研究上开展了一系列语言资源库的研究工作并制定了中文及各个少数民族语言软件使用术语规范^[3]。为了进一步推动我国少数民族语言信息技术研究,加速我国少数民族语文信息化进程,国家民委文化宣传司和中国科学院自动化研究所联合举办了“中国少数民族语言信息技术与语言资源库建设学术研讨”。云南等少数民族省和自治区已建立或准备建立本区域的少数民族语言文字资源库。新疆在维哈柯文语言资源方面研究较早,但服务于多语种软件开发的语言资源库研究建设尚属空白^[4]。

2 SSH 技术软件架构

随着计算机信息技术和 J2EE 的飞速发展,为了更好地创建企业应用程序,出现了许多的 Framework,一个典型的 J2EE 应用,至少包括以下 3 部分:表现层、业务逻辑层和数据持久层。文中用 Struts2 实现表现层,用 Spring 实现业务层,用 Hibernate 实现持久层,然后把这些框架无缝地整合起来,并应用到项目开发中。通过使用这些成熟的框架,能够减少重复开发工作量,

缩短开发时间,降低开发成本,增强程序的可维护性、可扩展性和可移植性^[5]。

2.1 Struts2 简介

Struts2 是 Apache 于 2007 年上半年推出的一个可扩展的 JAVA EE web 框架。当 Struts2 框架诞生后,基于 Struts1.x 和 WebWork 框架的 J2EE 实际项目开发都将转入使用 Struts2 框架,在这种背景下,Struts2 在短时间内迅速成为 MVC 领域最流行的框架。Struts2 吸取了 Struts1.x 和 WebWork 的各种优点,更改了一些不足的地方,譬如移走了原来 WebWork 中的 IoC 实现,引入 Spring 的 IoC 实现等。基于 Struts2 多层系统结构的软件克服了传统体系结构软件在功能和效率方面的缺陷,能够很好地适合当前企业需要。

Struts2 的目标很简单:使 Web 开发变得更加容易。图 1 显示了 Struts2 的体系结构,下面依据 Struts2 体系结构说明 Struts2 的工作机制。

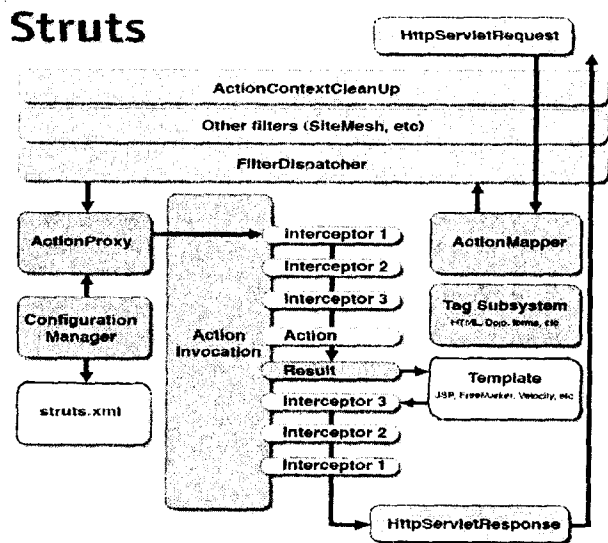


图 1 Struts2 体系结构

(1) 客户端初始化一个指向 Servlet 容器 (例如 Tomcat) 的请求。

(2) 这个请求经过一系列的过滤 (Filter), 如 ActionContextCleanUp 为可选过滤器。

(3) 接着 FilterDispatcher 被调用, FilterDispatcher 通过询问 ActionMapper 决定这个请求是否和某一个 Action 相关连。

(4) 当 ActionMapper 决定了一个 Action 的调用, FilterDispatcher 把请求处理交给 ActionProxy。

(5) ActionProxy 通过 Configuration Manager 询问框架的配置文件, 找到需要调用的 Action 类。

(6) 接着 ActionProxy 生成一个 ActionInvocation 的实例, 其负责命令执行。

(7) ActionInvocation 实例使用命名模式来调用, 在

调用 Action 的过程前后,涉及到相关拦截器(Interceptor)的调用,这样调用任何拦截器先于调用 Action。

(8)一旦 Action 执行完毕,ActionInvocation 负责根据 struts.xml 中的配置找到对应的返回结果。结果一般是调用一个在 JSP 或 FreeMarker 中绘制的模板,也有可能是调用另外的一个 Action 链。在表示过程中可以使用 Struts2 框架中继承的标签。在这个过程中需要涉及到 ActionMapper。

2.2 Spring 简介

Spring 是一个开源框架,是为了简化企业级应用开发而诞生的,它使用简单的 JavaBean 来实现以前那些只有 EJB 才能实现的功能,它是一个从实际项目开发经验中抽取的、可高度重用的应用框架。简单地说, Spring 是一个轻量级的 IoC(控制反转)和 AOP(面向切面编程)容器框架。这个简单的定义概括了 Spring 的重要功能和特点:

(1)轻量级。从大小和系统开支上说 Spring 都算是轻量级的,最为重要的一点是 Spring 是被设计为非侵入式的:基于 Spring 开发的系统中的对象一般不依赖于 Spring 的 API。

(2)IoC 即反转控制。Spring 提倡使用 IoC 来实现松散耦合,由容器控制对象之间的依赖关系,这样就减轻了组件之间的依赖关系,提高了组件的可移植性。IoC 是 Spring 框架的基础和精髓。

(3)AOP 即面向切面编程。Spring 对面向切面编程提供了强大支持,通过将业务逻辑从系统服务中分离出来,实现了内聚开发。系统对象只做它们该做的业务逻辑,而不用关心其它系统问题。

(4)容器。Spring 是一个容器,它包含并且管理系统对象的生命周期和配置。相对于传统的重量级 EJB 容器, Spring 容器是轻量级的。

(5)框架。Spring 框架使简单的组件配置和组合复杂的应用都成为可能。在 Spring 中,应用对象被声明式地组合,典型的组合方式是应用对象被声明式地组合在一个 XML 文件里。Spring 也提供了很多基础功能如事务管理、持久性框架集成等,将应用逻辑开发留给了开发者。

Spring 的所有这些特性使应用系统的开发更为简单,代码更为清晰,更易于管理,更易于测试。

2.3 Hibernate 简介

Hibernate 是一个开源的 ORM 框架,是现在最流

行的轻量级持久化框架之一。它将 JDBC 进行了轻量级的对象封装,使得开发人员可以方便地使用对象编程的思想来操纵数据库。利用 Hibernate,开发者可以按照所有 Java 的基础语义(包括关联、继承、多态、组合以及 Collection 架构)进行持久层开发。只要写好相关配置文件,就可以以面向对象的方式操作数据库,大幅度减少手工编码操作 JDBC 和 SQL 的时间。由于在编程中不会有特定数据库相关的代码,所以可以很容易地移植数据库而不用修改 Java 代码(只需修改 database schema 及 O/R 映射文件),大大提高了系统的灵活性。Hibernate 支持所有主流商业数据库和开源数据库,另外 Hibernate 支持多种缓存机制,可大大提高系统数据操作的效率。所以现在越来越多的 Java 开发人员把 Hibernate 作为企业应用和关系数据库之间的中间件,以节省和对对象持久化有关的 30% 的 JDBC 编程工作量。

3 MLRLMSBSSH 的设计及实现

3.1 系统的设计

多语种语言资源库的建立需要在研究维哈柯计算机术语规范的基础上,制定阿、俄界面术语规范,设计出符合多语种语言资源需求的存储模型,之后对大量原始资源进行采集、分类,提取所需界面术语及软件常用短语,通过验证之后进行入库。建立了多语种语言资源库之后,通过运用术语自动提取技术为用户提供界面术语检索以及界面术语自动翻译服务。系统开发流程如图 2 所示^[6]。

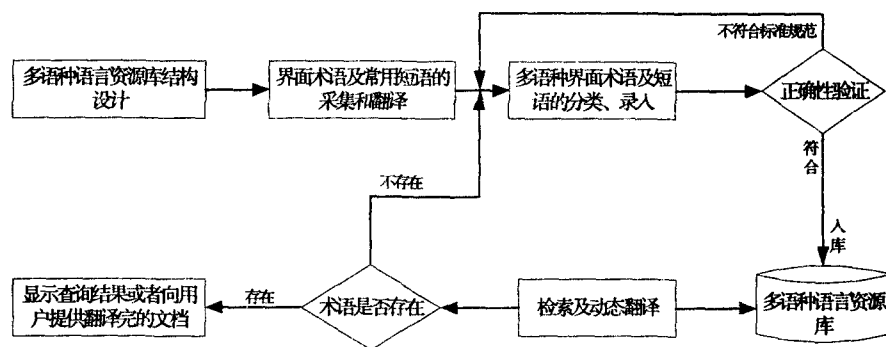


图 2 系统开发流程图

多语种语言资源库结构的设计将会直接影响系统的整体框架,多语种语言资源库结构的质量决定资源的检索速度、翻译质量,所以设计出符合多语种语言资源需求,能够清楚描述多语种界面资源和软件常用短语资源各项属性和参数的资源库存储模型非常重要。文中将基于国内外有关资源库和术语的标准规范,参考国际和国内知名的各类资源库结构模型,如:LOM 和 Dublin core 模型和其他资源库存储模型,结合新疆

以及中亚、西亚对语言资源库的需求,提炼、制定多语种语言资源库结构模型。实现资源的统一化、标准化,提高多语种资源的共享、重用以及查询效率^[7]。

多语种语言资源库管理系统提供了对多语种语言资源的管理和维护机制。多语种语言资源库管理系统一共包含五大功能模块:用户管理、术语管理及维护、检索管理、对语言资源库的备份及恢复、自学习管理。其具体的功能模块如图3所示。

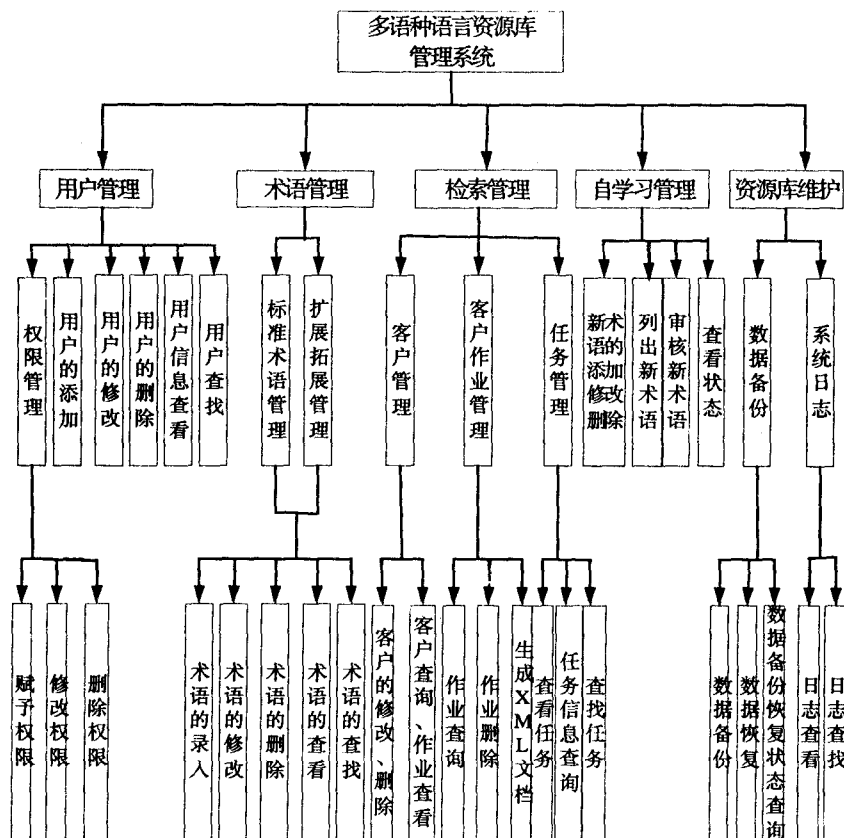


图3 系统功能模块

如图3所示,在用户管理模块中,用三种类型的用户,即术语管理员,审核专家,系统管理员。各种类型的用户根据其自身权限的不同来操作系统。采用用户-用户组-角色(三类用户)-权限来进行对语言资源库的用户管理。用户管理模块中可以对用户权限进行管理,可以增加、删除、修改、查看用户信息,也可以根据关键字查找用户^[8]。

术语管理模块又分为标准术语管理和扩展术语管理两大部分。术语管理模块主要完成了术语及短语的收集、翻译、录入工作,以及实现对已入库术语的增、删、查的维护工作。

检索模块中设有三个子模块:客户管理、客户作业管理及任务管理。在客户管理子模块中可以实现对客户相关信息的搜索,以及对客户信息的查改、修改及删除等。客户作业管理实现客户作业的搜索及作业的管

理。任务管理模块可以完成对任务信息的查询^[9]。

语言资源库管理系统将提取的界面术语与已有的三项标准术语空间进行匹配,确定查询的术语初始检索条件,未匹配的术语与已有的界面术语的同义词库进行匹配,确定查询的术语及短语。在语言库中检索匹配的术语及短语,如果存在完全符合的术语及短语,则将术语翻译标准 XML 格式输出;如果不存在这种完全符合检索条件的构件,则根据相似度匹配算法,找出最相近的术语,并将之标准格式输出给用户。

自学习管理模块主要是实现对未检索到的术语及短语的再审核。在这些模块中可以查看需自学习的内容,审核专家可通过该模块审核、增加、查看、修改、删除新术语。此外,还可以通过该模块查看专家审核术语的进度状态等。

多语种语言资源库维护模块设有两个子模块:数据备份、系统日志。系统管理员通过数据备份子模块可以对语言资源库的数据进行备份、恢复及查询。在系统日志子模块中可以查看、查找系统日志^[10]。

多语种语言资源库管理系统设计了资源库的结构,研究与分析了多语种语言资源的语种分类以及在软件界面用语中的使用频率等参数,建立能够详细描述多语种语言资源的各个属性和参数的存储模型和资源库结构^[9]。

3.2 系统的实现

本系统采用一种基于 J2EE 的框架 Struts2 + Spring2 + Hibernate3,采用三层 Web 体系结构模式,其中表现层采用 Struts2 技术,核心业务逻辑采用 Spring 框架,而数据持久层采用时下流行的 Hibernate O/R Mapping 技术。下面分别介绍这三种技术在多语种语言资源库平台中的应用^[11]。

3.2.1 Struts2 技术应用

Struts2 中视图可以 FreeMarker、Velocity 等模板技术,为简单起见在文中使用一组 JSP 页面来实现。这些 JSP 页面中不包含业务逻辑,也不包括模型的信息。模型的信息是通过控制器传递的。在 Struts2 中,使用 FilterDispatcher 和 Action 共同作为控制器。FilterDispatcher 是一个标准的过滤器,它负责过滤所有的

请求,并将这些请求转发给相应的 Action。Struts2 中的 Action 是一个 POJO,它本身不包含任何具体的业务代码,它仅仅是进行一些简单的控制。Struts2 的核心文件 struts.xml 内定义了 Struts2 的系列 Action,定义 Action 时,指定该 Action 处理结果与视图资源之间的映射关系。下面就判定资源内容完整为例给出 struts.xml 的部分代码(JSP 页面提交了名为 languageResoAud 的 action 请求):

```
<action name="languageResoAud"
class="languageResoAudAction" method="auditing">
<result name="success">/hijx/oneReso.Aud.jsp</result>
<result name="Incomplete">/hijx/announcement.jsp</result>
<result name="gohomepage">/hijx/oneReso.homeindex.jsp</result>
</action>
```

在上面的 struts.xml 文件中,定义一个名为“languageResoAud”的 action,它指定了三个 result,result 元素指定 auditing()方法返回值和视图资源之间的映射关系。当 languageResoAudAction 返回值为“success”时,提交的多语种语言资源内容完整,页面跳转到 oneResoAud.jsp,送交该资源到指定的审核专家组待审;当返回值为“Incomplete”时,资源内容不完整,页面跳转到 announcement.jsp,直接联系资源提交者通知资源内容不完整;当返回值为“gohomepage”时,页面跳转到 homeindex.jsp,结束本次编辑审核,返回本系统主界面。

3.2.2 Spring 技术应用

业务逻辑层使用 Spring 框架技术实现,其主要功能是执行程序的逻辑部分,调用持久层来处理事务,并控制异常。Spring 主要负责处理应用程序的业务逻辑和业务校验、管理事务;同时管理业务层级对象的依赖;而且在显示层和持久层之间增加了一个灵活机制;Spring 把 DAO 对象和 business service object 搭配起来,通过调用 DAO 和持久层进行通讯。applicationContext.xml 是 Spring 上下文,它是一个配置文件,用来向 Spring 框架提供上下文信息。Spring 上下文包括企业服务,例如 JNDI、EJB、电子邮件、国际化、校验和调度功能。下面是本系统中 applicationContext.xml 对应于信息公告的部分代码:

```
<bean id="bulletinsService" class="cn.edu.nuist.dao.implement.huijianxin.BulletinsService">
<property name="bulletinsMapDao">
<ref bean="bulletinsMapDao"/>
</property>
</bean>
```

```
<bean id="bulletinAction" class="cn.edu.nuist.dao.implement.huijianxin.BulletinsAction" singleton="false">
<property name="bulletinsService">
<ref bean="bulletinsService"/>
</property>
<property name="pagerService">
<ref bean="pagerService"/>
</property>
</bean>
```

3.2.3 Hibernate 技术应用

Hibernate 的目标是在 JDBC 之上提供一层封装,同时提供完善的透明持久化。Hibernate 不仅管理 Java 类到数据库表的映射(包括从 Java 数据类型到 SQL 数据类型的映射),还提供数据查询和获取数据的方法。建立 Hibernate 持久层,首先要将对象持久化,Hibernate 通过 XML 文件来映射对象。在这里注意各种关系的映射,如 one to one、many to one、many to many、one to many 等。例如本系统的用户提交 TEMP-RESO-USER(临时术语用户表)对应的映射文件 TempResoUser.hbm.xml。在 TempResoUser 类中定义了临时资源用户 tempRUIId,临时资源 tempResource 及相关的用户 User,文中截取映射文件的一部分加以说明,映射文件的属性和数据库表的字段是一一对应的,且类型一致。

```
<hibernate-mapping>
<class name="cn.edu.nuist.model.TempCompTerm"
table="TEMP-RESO-USER" schema="RESOURCE">
<id name="tempRUIId" type="java.lang.Long">
<column name="TEMP-R-U-ID" precision="10" scale="0" />
</id>
<generator class="assigned"/>
</id>
<many-to-one name="term" class="cn.edu.nuist.model.User" fetch="select" lazy="false">
<column name="User-ID" precision="6" scale="0" not-null="true" />
</many-to-one>
<many-to-one name="tempResource" class="cn.edu.nuist.model.TempResource" fetch="select">
<column name="TEMP-RESO-ID" precision="10" scale="0" not-null="true" />
</many-to-one>
</class>
</hibernate-mapping>
```

在本系统中,使用 hibernate.cfg.xml 对 Hibernate 进行配置,因为使用 XML 文件更加方便直观,当增加 Hibernate 映射文件的时候,可以在 hibernate.cfg.xml 里面增加。使用 XML 时部分配置代码可通过第三方

的开发插件协作生成,减少工作量。配置文件 hibernate.cfg.xml 代码如下所示:

```
<? xml version = '1.0' encoding = 'UTF-8'? >
<! DOCTYPE hibernate - configuration PUBLIC
"- //Hibernate/Hibernate Configuration DTD 3.0//EN" "http://
hibernate.sourceforge.net/hibernate - configuration - 3.0.dtd">
<hibernate - configuration>
<session - factory>
<property name = "connection. username">system</property>
<property name = "connection. url">
jdbc:oracle:thin:@huijx:1521:reso
</property>
<property name = "dialect">
org.hibernate.dialect.Oracle9Dialect
</property>
<property name = "myeclipse. connection. profile">Oracle</prop-
erty>
<property name = "connection. password">huijx</property>
<property name = "connection. driver. class">
oracle.jdbc.driver.OracleDriver
</property>
<property name = "show. sql">true</property>
<mapping resource = "cn/edu/nuist/model/IdAssignment. hbm.
xml"/>
<mapping
resource = "cn/edu/nuist/model/TempResoUser. hbm. xml"/>
.....
</session - factory>
</hibernate - configuration>
```

Hibernate 为 Java 提供了 ORM 持久化机制和查询服务,Hibernate 将我们在 Java 类里使用的 HQL 语句转换成 SQL 语句,利用 JDBC 驱动进而操作数据源,完成数据的增加、删除、修改、查询等最底层的数据库操作。

(上接第 77 页)

- [10] Sahoo S S, Sis Center K E, Halb W, et al. A Survey of Current Approaches for Mapping of Relational Databases to RDF [DB/OL]. W3C RDE2RDF Incubator Group Reporter, <http://esw. w3. org/topic/Rdb2RdfXG/StateOfTheArt>, 2009,01.
- [11] Chen Huajun, Wu Zhaohui, Mao Yuxin, et al. DartGrid: a Semantic Infrastructure for Bilding Database Grid Applications [J]. Concurrency and computation: Practive and Experience, 2006,18:1811 - 1828.
- [12] Cullot N, Ghawi R, Yetongnon K. DE2OWL: A Tool for Automatic Dabase - to - Ontology Mapping [C] // Ceci M, Malerba D, Tanca L, et al. SEBD. DBLP. [s. l.]: [s. n.],

4 结束语

文中提出和开发的 MLRLMS 系统在传统的语言资源库管理系统的基础上采用了 SSH 框架,探索了基于 SSH 框架给 Web 应用开发带来的巨大好处。研究建立汉、英、阿、俄对照词汇和短语资源库,为多语种软件开发提供规范的软件界面和用户文档用语,减少重复翻译工作,提高翻译质量,具有重大意义;而且能使多语种软件产品符合阿拉伯语和俄语的使用习惯。系统建设填补了多语种语言民族地区相关软件业领域建设的空白,对软件开发提供重要技术支持。

参考文献:

- [1] 图格木勒.蒙古语语言资源库建设相关技术研究[D].呼和浩特:内蒙古大学,2007.
- [2] 郭向勇,吴光斌.构建基于跨平台检索技术的校园网多媒体资源库[J].计算机工程,2002(5):252 - 254.
- [3] 钟 璐,王 辉,李锐强,等.基于语义 Web 的网络学习资源库本体实现[J].计算机工程,2007(8):282 - 284.
- [4] 孙启勤,周 卫.计算机在线翻译快速入门[M].北京:中国水利水电出版社,2008.
- [5] 李 刚.Struts2 权威指南——基于 WebWork 核心的 MVC 开发[M].北京:电子工业出版社,2007.
- [6] 曹 国,姚建初,吴义忠,等.支持协同设计的资源库的开发研究[J].计算机工程,2003(16):182 - 184.
- [7] 蔡雪燕.Hibernate 开发及整合应用大全[M].北京:清华大学出版社,2006.
- [8] 张德育.蒙古语远程教育平台中蒙古文教育论坛和资源库系统的研究与实现[D].呼和浩特:内蒙古大学,2008.
- [9] 曾 皓.多语种软件构件库的分类与检索[D].北京:中国科学院研究生院,2008:36 - 43.
- [10] 徐 明,黄云森,陈可期.教学资源库建设策研究[J].中山大学学报:自然科学版,2002,41:114 - 117.
- [11] 李 刚.整合 Struts + Hibernate + Spring 应用开发详解[M].北京:清华大学出版社,2007.

- 2007:491 - 494.
- [13] Bizer C, Cyganiak R. D2RQ - Lessons Learned[DB/OL]. Position paper for the W3C Workshop on RDF Access to Relational Databases, 2007, 09. <http://www. w3. org/2007/03/RdfRDB/papers/d2rq - positionpaper>.
- [14] Fisher M, Dean M. Automapper: Relational Database Semantic Translation using OWL and SWRL[C/OL]. Proc. of the IASK Int Conf. - E - Activity and Leading Technologies 2007(E - ALT2007), 2007, <http://www. progeny. net/People/MattFisher/files/papers/AutomapperIASKEALT07.pdf>.