

上海图书馆二次文献共享服务平台

技术手册



2015-12-30

上海双地信息系统有限公司

目 录

[1. 技术清单 2](#_Toc439199166)

[2. 数据框架 3](#_Toc439199167)

[2.1. 核心数据模式管理 3](#_Toc439199168)

[2.2. 中图分类法&学科分类法 3](#_Toc439199169)

[2.3. Solr词表管理 4](#_Toc439199170)

[2.4. 数据版本管理 4](#_Toc439199171)

[3. 数据工具 6](#_Toc439199172)

[3.1. 数据工具结构 6](#_Toc439199173)

[3.2. 数据规则 6](#_Toc439199174)

[3.3. 数据任务定义 8](#_Toc439199175)

[3.4. 数据转换脚本 8](#_Toc439199176)

[4. 权限控制体系 11](#_Toc439199177)

[4.1. 功能控制 11](#_Toc439199178)

[4.2. 产品控制 11](#_Toc439199179)

[5. 检索服务 12](#_Toc439199180)

[5.1. 篇数据检索 12](#_Toc439199181)

[5.2. 文献导航 12](#_Toc439199182)

[6. 整本浏览 13](#_Toc439199183)

[6.1. 报纸整本浏览 13](#_Toc439199184)

[6.2. 期刊整本浏览 14](#_Toc439199185)

[6.3. 行名录整本浏览 14](#_Toc439199186)

# 技术清单

|  |  |  |  |
| --- | --- | --- | --- |
| 项目 | 技术 | 版本 | 说明 |
| 整体框架 | SpringMVC | 4.0.9 | 支持Servlet3标准，支持JPA  为了兼容weblogic 12c |
| 数据库相关 | Hibernate | 4.3.7.Final | ORM基础框架 |
| Hibernate-Validator | 5.1.3.Final | ORM数据验证 |
| Spring-data-jpa | 1.8.0 | JPA技术 |
| Druid | 1.0.13 | 阿里巴巴连接池 |
| Oracle驱动 | 11.2.0.4.0 | Oracle官方驱动 |
| Solr检索服务 | SolrJ | 4.10.4 | SolrJ客户端4.10.4版本 |
| spring-data-solr | 1.4.10 | Spring Solr数据接口 |
| 日志相关 | Logback | 1.1.2 | 采用log4j一整套的日志系统，支持异步日志 |
| Log4jdbc | 1.2 |
| Log4j | 2.4 |
| Slf4j | 1.7.9 |
| 权限相关 | Shiro | 1.2.3 | Shiro的一整套权限控制体系 |
| shiro-spring | 1.2.3 |
| 缓存相关 | Ehcache | 2.5 | Ehcache缓存系统 |
| 测试相关 | Junit | 4.12 | 测试相关的内容 |
| Powermock | 1.6.1 |
| Mockito | 1.10.8 |
| Selenium | 2.44.0 |

# 数据框架

## 核心数据模式管理

我们首先通过下图可以看到Oracle和Solr的数据组织结构，整个数据框架就是支持这些元素的定义和验证；数据工具和检索服务全都依赖于这里的数据框架定义。



## 中图分类法&学科分类法

当然，我们的数据框架还管理一些其他的内容，例如中图分类法，学科分类法等等。我们的文献加工是按照中图分类法进行加工的，我们现在提供的是学科分类法的服务，我们通过规则定义了学科分类法通过中图分类法的表示方式。

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1级类目** | **2级类目** | **3级类目** | **4级类目** | **5级类目** | **中图分类号** | **排除** |
| 哲学、宗教 |  |  |  |  | A\*，B\*，D0-0，F0-0，G449\*，P315.96，R395.6 |  |
|  | 马克思主义、列宁主义、毛泽东思想、邓小平理论 |  |  |  | A\*，B0-0，D0-0，F0-0 |  |
|  |  | 马克思、恩格斯生平、著作及学习和研究 |  |  | A1\*，A71\*，A72\*，A81\*，B0-0，D0-0，F0-0 |  |
|  |  | 列宁生平、著作及学习和研究 |  |  | A2\*，A73\*，A82\* |  |
|  |  | 斯大林生平、著作及学习和研究 |  |  | A3\*，A74\*，A83\* |  |
|  |  | 毛泽东生平、著作及学习和研究 |  |  | A4\*，A75\*，A84\* | A49\*，A849\* |
|  |  | 邓小平生平、著作及学习和研究 |  |  | A49\*，A76\*，A849 |  |
|  | 哲学理论 |  |  |  | B，B-4\*，B0\* |  |
|  |  | 哲学基本问题 |  |  | B01\* |  |
|  |  | 马克思主义哲学 |  |  | B0-0 |  |
|  |  | 辩证唯物主义 |  |  | B02\* |  |
|  |  | 历史唯物主义 |  |  | B03\* |  |
|  |  | 哲学流派 |  |  | B08\* |  |
|  | 世界哲学 |  |  |  | B1\* |  |
|  | 中国哲学 |  |  |  | B2\* |  |
|  |  | 古代哲学 |  |  | B21 |  |

从上可见，学科分类法，每一个条目，都可以通过中图分类法的规则来标识，而且可以方便的转换为SOLR表达式来进行适配。

## Solr词表管理

我们数据框架集成了Solr词表的管理，用户可以定义词表，系统将会自动将词表应用到SolrCloud上，当然，如果希望生效，必须要重建Solr的索引。

## 数据版本管理

我们提供数据版本管理，按照入数据的批次可以将其进行版本分割，版本信息将从Oracle带入到Solr中。

我们提供DeleteVersion和UpdateVersion。启动DeleteVersion是单值数据，只支持一次删除；UpdateVersion是多值数据，支持数据的若干次的版本更新。

版本控制不可能支持任意的操作，必须基于一定规则基础上才能运行。有以下两条规则需要遵守：

1. 已经在低版本删除的数据，不需要在高版本再次删除

我们不允许已经在低版本删除了的数据，在高版本上再次删除，这显示是没有意义的，因此会跳过这个更新。

1. 数据赎回只允许针对删除版本执行，不允许跨版本赎回！

需要赎回的数据（就是已经删除了，再次启用的意思），必须只能在删除版本上执行，这个要求是为了逻辑上的一致性提出的，否则，非常容易造成版本控制的失效；

另外，如果不希望在删除版本上赎回数据，可以选择进行数据清理，然后在高版本上重入数据也是可行的。

# 数据工具

## 数据工具结构

数据工具是我们用来处理外部数据如何按照服务平台的规则顺利进入服务平台数据池的工具，我们通过一系列的定义和约束来保障数据进入的规范。

数据工具的组成结构如下图所示：



## 数据规则

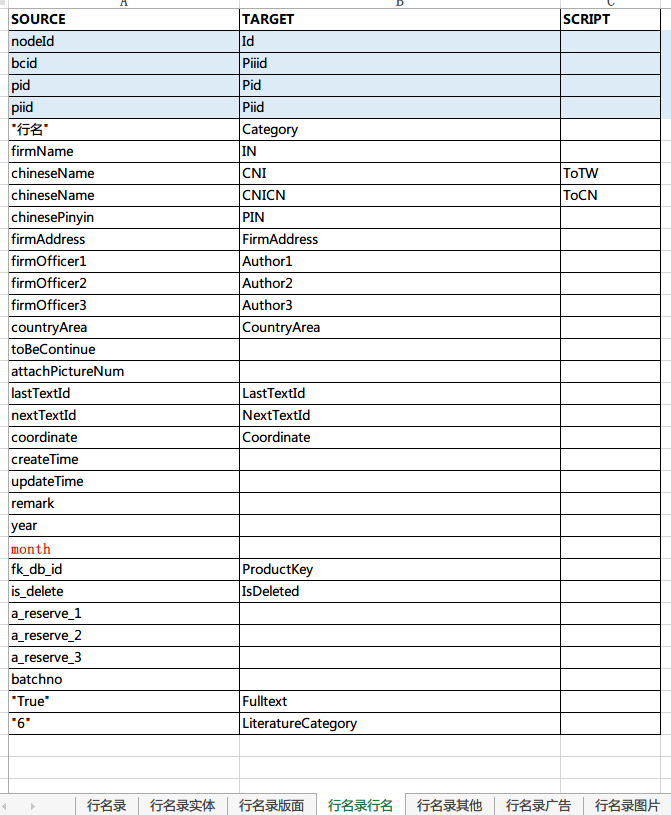
数据规则如下图所示，我们对源表和目标表定义了相应的字段对应，另外，额外加了一个脚本列，可以配置脚本库中脚本对数据进行处理和转换。

SHEET名称：代表规则名称

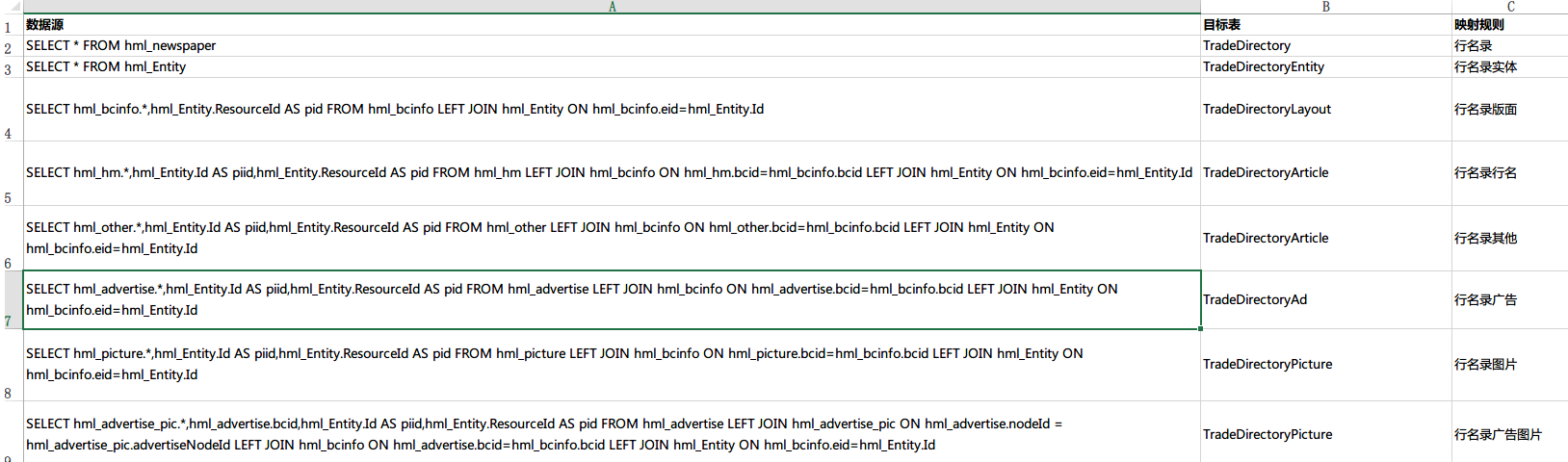
SOURCE段：代表源数据字段

TARGET段：代表目标数据字段

SCRIPT段：代表数据转换脚本（数据变换）



## 数据任务定义



数据源：是SQL脚本，可以直接提取数据；

目标表：指定对应的目标数据表

映射规则：之前定义的数据规则名称。

## 数据转换脚本

* 脚本编制说明：

1. 语言是：groovy，类似于java，比java强大灵活很多
2. 数据脚本如下

Object transform(Object instance, String sourceField) {

}

或

Object transform(Object value, String sourceProperty, String referenceProperty) {

}

Instance代表的是这一行数据对象，可以根据property来取值。

第一种函数，是直接对源数据对应的字段进行处理，就是sourceField，只有一个参数，使用的时候不需要带参数，只写名称即可；

第二种函数，可能是通过源表中额外字段进行处理，因此，多了一个referenceProperty参数，这个参数必须要在转会规则中指定，否则，工具因为不知道你具体要引用哪个属性而出错。

* 现有脚本库说明：

|  |  |  |
| --- | --- | --- |
| 名称 | 说明 | 参数情况 |
| BOOL | 布尔值处理 | 默认参数 |
| SmartYEAR | 期刊年份补全 | 默认参数，带一个参数为fkdbid，就是产品key  根据fkdbid智能判断年份补全 |
| GBK | 乱码解决方案（待定） | 默认参数 |
| ID\_GENERATOR | ID生成 | 除默认参数，带一个参数，是指向参数，就是主ID，我们需要根据主ID生成关联ID  写法就是  ID\_GENERATOR(ID\_GENERATOR(advertiseNodeId))  其实括号就是主ID属性，这里是广告节点ID |
| SmartToCN | 内容自动转简体中文  全英文转空  简体中文不变  繁体转简体 | 默认参数 |
| SmartToEng | 内容判断英文  全英文不变  其他转空 | 默认参数 |
| SmartToTw | 内容自动转繁体  全英文转空  简体变繁体  繁体不变 | 默认参数 |
| SmartToCNForXK | 现刊专用简体中文转换 | 默认参数外，多一个参数  说白了就是2选1判断语言，然后转换，做到最大转换 |
| SmartToEngXK | 现刊专用简体中文转换 | 默认参数外，多一个参数  说白了就是2选1判断语言，然后转换，做到最大转换 |
| SmartToTwForXK | 现刊专用简体中文转换 | 默认参数外，多一个参数  说白了就是2选1判断语言，然后转换，做到最大转换 |
| FULLTEXT | 有全文判断 | 默认参数 |
| UPLOADFULLTEXT | 有上传全文判断 | 默认参数 |

# 权限控制体系

## 功能控制

功能权限控制体系如下图所示：

我们的功能权限分配到具体的角色上，用户的功能是通过绑定的角色来体现的。部分角色是内置角色，用户的有些内置角色默认关联上。



## 产品控制

产品权限控制体系如下图所示，我们可以对角色进行权限配置，这种权限是默认的角色可以访问的产品配置，不限次数和时间。另外，对于组织和个人授予的权限，是需要配置相依的时间期限的，另外对试用用户还有一个整本和下载次数的限制。



# 检索服务

## 篇数据检索

检索服务如下所示：



我们的检索服务是通过数据框架的定义来进行的，检索内容是用户输入的内容了，检索的字段是系统管理员配置的，聚类也是一样在后台进行配置的，而且支持多级聚类配置，可以逐级缩小聚类范围，检索结果的呈现的字段内容是由数据框架配置的，聚类的呈现方式也支持自定义的图形化呈现方式。

## 文献导航

文献导航是针对文献的导航，主要功能架构如下：



关于学科分类导航就是利用之前的数据框架中学科分类和中图的映射关系进行的。

# 整本浏览

## 报纸整本浏览

报纸整本浏览采用了Flex4.6作为开发框架，使用了多种组件进行呈现，具体如下图所示：



## 期刊整本浏览

期刊整本浏览和报纸整本浏览类似的技术，不过更加简单一点，如下所示：



## 行名录整本浏览

行名录使用了报纸的浏览工具进行了支持，只是去掉了日历导航的支持，因为行名录不是按照日进行发布的。具体如下所示：

