

# 商品标签信息加工系统研究

胡珈诚<sup>1</sup> 甘克勤<sup>2</sup> 楚昊天<sup>1</sup> 张玳菲<sup>1</sup>

(1. 新南威尔士大学, 澳大利亚新南威尔士州 2052; 2. 中国标准化研究院, 北京 100191)

**摘要:** 为量化商品质量, 实施“质量强国”战略, 国家标准馆研制了一套商品标签信息加工系统, 实现了商品标签与多个实体数据的关联集成, 并录入标签表达的指标细节。描述了SSM+ServiceComb的微服务框架应用, 阐述了应用JSON数据结构, 并基于MongoDB非关系型数据库, 对商品标签的细节指标进行统一加工和存储, 最终展示了系统的实现效果, 实现了商品与多源数据的有效关联。

**关键词:** 商品标签加工; 多源数据; 数据关联; 半结构化数据存储; 微服务

中图分类号: TP311.1

文献标识码: A

DOI: 10.3772/j.issn.1674-1544.2020.02.010

## Research of Commodity Label Information Processing System

HU Jiacheng<sup>1</sup>, GAN Keqin<sup>2</sup>, CHU Haotian<sup>1</sup>, ZHANG Daifei<sup>1</sup>

(1. University of New South Wales, Kensington, NSW 2052; 2. China National Institute of Standardization, Beijing 100191)

**Abstract:** In order to quantify the quality of commodities and respond to the strategy of “quality power”, the national standard center has developed a set of commodity label information processing system, which realizes the association and integration of commodity labels with multiple data source, and records the index details expressed by labels. In order to realize the effective association between commodities and multi-source data, this paper describes the microservice framework application of SSM+ServiceComb, expounds the application of JSON data structure based on MongoDB unstructured database, realizes the unified processing and storage of the detailed indicators of commodity labels, and finally shows the implementation effect of the system.

**Keywords:** product label processing, multi-source data, data association, unstructured data storage, microservice

## 0 引言

商品质量是衡量商品使用价值的尺度。衡量商品质量的传统方法就是符合或达到国家标准和行业标准的即视为商品质量合格。随着标准化改革不断推向深入, 对商品质量的要求越来越高,

广大企业标准和团体标准的数量已经远远超过国家标准和行业标准, 同类商品执行的标准千差万别, 仅凭“合格”是难以量化商品质量的。

商品执行的标准信息和商品标签上的指标是量化商品质量的基础数据, 传统的商品信息填报和查询平台为中国物品编码中心提供的“中国商

**作者简介:** 胡珈诚 (1994—), 男, 澳大利亚新南威尔士大学在读研究生, 研究方向: 数据库和人工智能方向; 甘克勤 (1983—), 男, 中国标准化研究院副研究馆员, 研究方向: 标准文献信息化 (通信作者); 楚昊天 (1995—), 男, 澳大利亚新南威尔士大学在读研究生, 研究方向: 数据科学和人工智能方向; 张玳菲 (1996—), 女, 澳大利亚新南威尔士大学在读研究生, 研究方向: 数据科学和人工智能方向。

**基金项目:** 中央基本科研业务费项目“标准指标揭示应用研究”(252017Y-5299)。

**收稿时间:** 2020年3月2日。

品信息服务平台” (<http://www.ancc.org.cn>)。但是,该平台提供的商品信息较为通用,没有涵盖表达质量的具体指标以及商品执行标准信息。如以“火山鸣泉矿泉水”为例,该平台提供了瓶高、瓶宽、深度、产地等基本信息,却缺少钙、镁、钾、钠元素含量等质量指标,如图1所示。

对此,国家标准馆研发了商品标签信息的采集APP和商品标签信息加工系统。商品标签信息加工系统的数据集主要是利用商品标签采集APP拍照扫码采集的商品标签图片和物品编码<sup>[1]</sup>,主要关联整合了3个数据源:一是国家标准馆的

标准题录数据,加工时应用提示词,关联并修正标签的执行标准号(部分执行标准号的写法不规范);二是主流电商的线上商品信息,包括商品名称、分类、价格等,加工时应用提示词,将线上线下的商品信息通过“商品名称”手动匹配,实现关联;三是企业信息,通过调用百度地图API,加工时应用提示词,实现产品相关的企业实体信息的准确录入,同时获取企业的位置信息。此外,抽取商品标签中的指标细节,用非结构化数据方式进行存储,各实体整合关联。商品标签与各实体的关联如图2所示。

本文拟对商品标签信息加工系统采用技术框架、业务流程以及应用实例进行阐述和分析,探讨商品与多源数据的有效关联。

## 1 系统构建原理

### 1.1 集成多源数据源

商品标签信息加工系统集成了多源数据源,主要包括标准题录信息(国家标准馆API<sup>①</sup>)、主流电商商品信息(SQL SERVER<sup>②</sup>存储)、企业信息(百度地图API<sup>③</sup>)以及商品标签信息(MongoDB数据库<sup>④</sup>[2])。各数据库的操作主要集中在读取、异步请求和存储3个方面。

数据来源多样化,各个数据源之间交互



图1 中国商品信息服务平台关于火山鸣泉矿泉水的信息

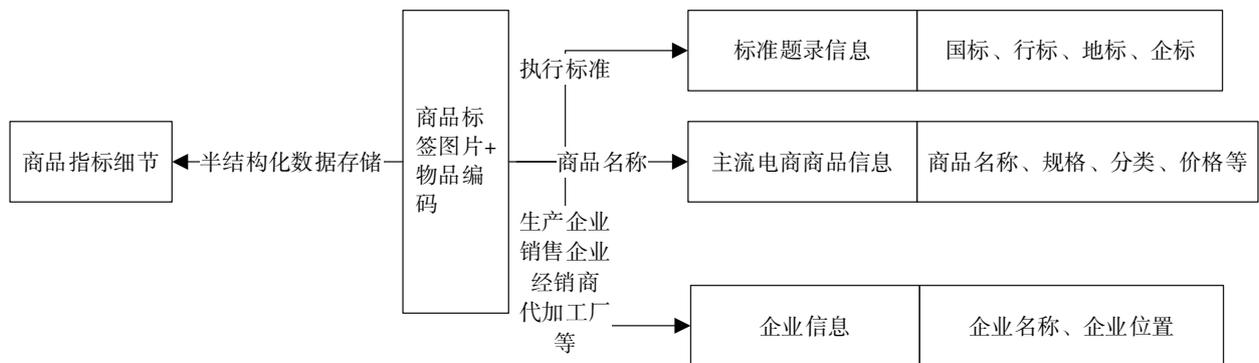


图2 商品标签与各实体关联

① 由国家标准馆门户 ([www.cssn.net.cn](http://www.cssn.net.cn)) 提供的可以读取标准题录数据的接口。  
 ② SQL Server是由Microsoft开发和推广的关系数据库管理系统(DBMS)。  
 ③ 由百度地图提供的可供开发者使用的外部接口。  
 ④ MongoDB是一个基于分布式文件存储的数据库。

繁琐。如果基于传统的SMM框架<sup>[3]</sup>（Spring<sup>①</sup>+Spring MVC<sup>②</sup>+Mybatis<sup>③</sup>）开发会丧失系统的可拓展性，难以整合可能的、更多的数据源。因此，商品标签信息加工系统引入了微服务，在保持原有Spring MVC开发风格的基础上（整合ServiceComb<sup>④</sup>对Spring MVC的注解支持机制，利用ServiceComb注解，在服务的实现类上打上@RestSchema注解以及配置相关接口契约），实现服务对客户端的响应，从而将不同的数据源请求独立形成微服务单元，实现微服务架构的开发<sup>[5-6]</sup>，降低了系统的耦合程度，便于新的数据源的装载和新功能板块的扩展。微服务架构如图3所示。

### 1.2 MongoDB数据库半结构化数据的存储

商品标签种类较多，指标结构差异较大，传统的关系型数据库字段名称必须要固定下来，然而商品的企业信息和指标种类存在不确定性，很难满足此类半结构化数据的存储。例如“火山鸣泉矿泉水”中有一个经销商和一个生产商，而“农夫山泉矿泉水”有3个不同的生产商，但没有经销商，并且这两种矿泉水的相关指标名称类别差距较大，无法固定表的字段，关系型数据库不能满足系统构建的需求。

MongoDB数据库作为半结构化数据的存储

数据库，对数据的存储具有读取速度快的优点，对半结构化数据的存储更加亲和。MongoDB中数据是BSON格式的，类似于JSON<sup>[7-8]</sup>数据格式，即“键—值（KEY-VALUE）”。JSON数据通常可以被用作MongoDB数据库存储半结构化数据的中间载体，其在被转化成DBObject数据类型后可以直接存入MongoDB数据库。

在商品标签信息加工系统中，对于无法存储在关系型数据库中的与商品有关的半结构化数据可以使用JSON数据格式进行处理，以KEY存储商品指标的标签名字，VALUE来储存对应标签的值即存储商品的信息，VALUE也可以内嵌新的键值以存储更为复杂的指标。JSON封装样例如图4所示，其中productInfo是整个数据的主键，记录与商品相关的商品指标数据，该商品的商品指标有12个，VALUE内部以新的“键—值（KEY-VALUE）”来存储；enterpriseInfo是商品相关企业信息的外键，记录与商品相关的企业信息，该商品关联的企业信息有两个，分别为生产商和运营商。

## 2 技术框架与业务流程

商品标签信息加工系统是基于JAVA SSM+ServiceComb框架开发的，数据库为SQL

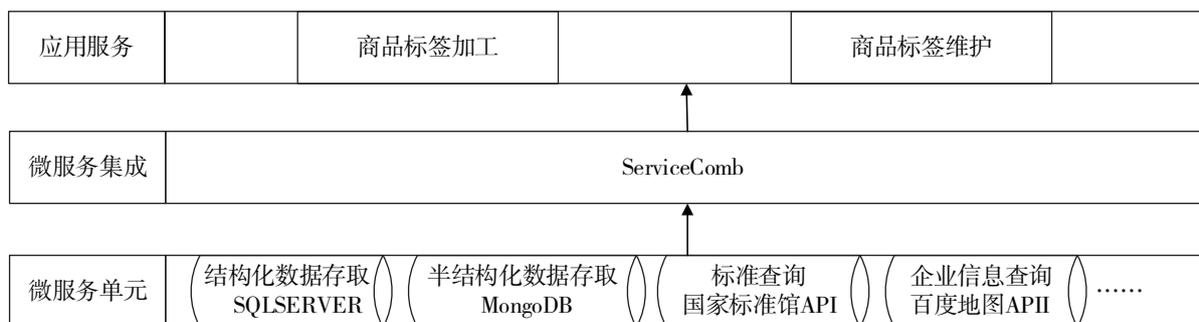


图3 微服务构架

① Spring 是一个开源框架，由 Rod Johnson 创建，为了解决企业应用开发的复杂性而创建的。Spring 使用基本的 JavaBean 来完成以前只可能由 EJB 完成的事情。  
 ② Spring MVC 属于 Spring Frame Work 的后续产品，已经融合在 Spring Web Flow 里面。Spring 框架提供了构建 Web 应用程序的全功能 MVC 模块。  
 ③ MyBatis 是一款优秀的持久层框架，支持定制化 SQL、存储过程以及高级映射，其主要实现了对传统 JDBC 操作的封装与利用反射机制打通 JAVA 类与 SQL 语句之间的相互转换。  
 ④ Apache ServiceComb 是由华为贡献的开源微服务解决方案。

SERVER 2008 以及 MongoDB 1.31.1, 服务容器采用 Apache Tomcat 7.0。

SSM集成 ServiceComb的技术框架如图5所示。

以物品编码定义的商品为加工的最小单元。具体业务流程为选中商品, 进入加工页面, 打开商品图片, 关联相关数据(包括执行标准号、商品名称及分类、企业名称), 进一步加工商品标签的指标, 最终保存数据, 标记商品为已加工状态。业务流程如图6所示。

### 3 实证分析

本文以对火山鸣泉矿泉水的商品信息加工作为案例展示商品标签信息的集成关联与指标细节

半结构化加工的具体实现, 最终给出成果数据, 验证本文设计构建的商品标签信息加工系统。

#### 3.1 主页面

该页面为商品标签信息集成关联加工主页面。数据加工人员依据以采集的商品标签照片, 完成以下两项工作: 多源数据源的集成关联和标签上指标细节的半结构化录入。标签信息集成关联加工主页面如图7所示。

#### 3.2 商品数据集成关联

数据加工人员根据商品图片填写商品名称。商品标签信息加工系统以提示词自动补全的方式将输入的信息与主流电商商品信息中的商品名称进行匹配, 而后由数据加工人员手动选择, 实现商品名称

```
{ "productInfo" :[{ "KEY" : "配料", "VALUE" : "天然矿泉水" },{ "KEY" : "净含量", "VALUE" : "470ml" },{ "KEY" : "保质期", "VALUE" : "18个月" },{ "KEY" : "锶", "VALUE" : "0.3-1.0mg/L" },{ "KEY" : "钾", "VALUE" : "0.8-1.8mg/L" },{ "KEY" : "钠", "VALUE" : "40.0-100.0mg/L" },{ "KEY" : "钙", "VALUE" : "10.0-30.0mg/L" },{ "KEY" : "镁", "VALUE" : "2.0-4.0mg/L" },{ "KEY" : "偏硅酸", "VALUE" : "15.0-30.0mg/L" },{ "KEY" : "碳酸氢根", "VALUE" : "150.0-350.0mg/L" },{ "KEY" : "溶解性总固体", "VALUE" : "200.0-500.0mg/L" },{ "KEY" : "PH值", "VALUE" : "7.8(±0.6)" }]} { "enterpriseInfo" :[{ "id" : "88929", "label" : "生产商" },{ "id" : "88954", "label" : "运营商" }]}
```

图4 JSON封装样例

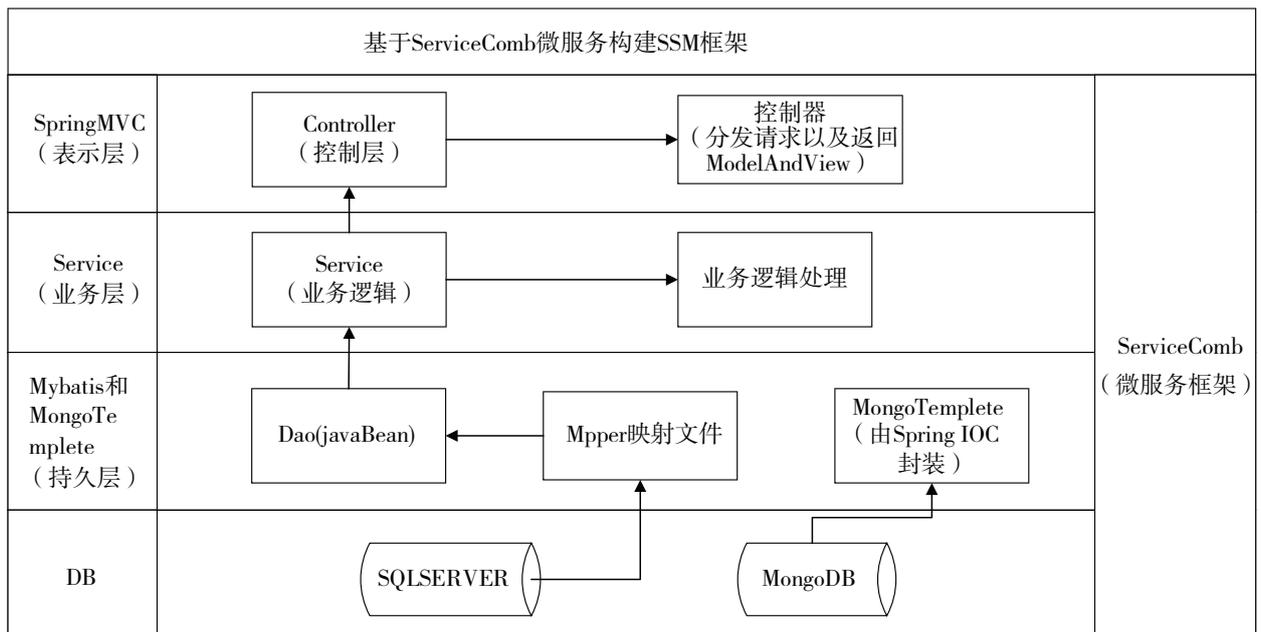


图5 SSM集成 ServiceComb 技术框架

与主流电商的商品信息的关联，如图 8 所示。

### 3.3 商品执行标准的集成关联

通过调用国家标准馆提供的题录数据 API，

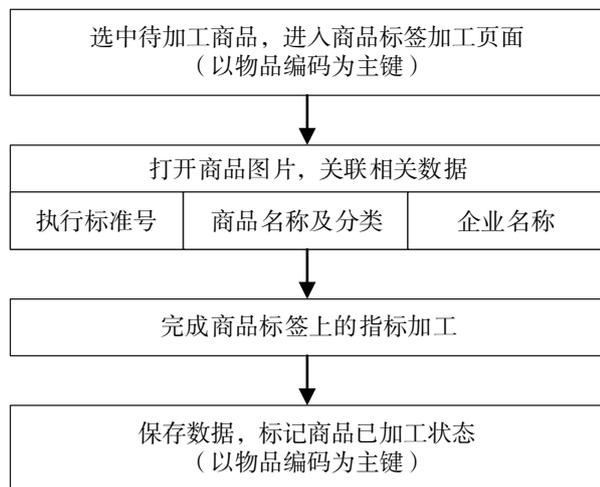


图 6 商品标签加工业务流程

商品标签信息加工系统以自动补全的方式实现了商品执行标准的集成关联。标准范围包括国家标准、行业标准、地方标准和企业标准以及国际标准，基本涵盖商品标签中定义的执行标准范围。加工过程以提示词的方式，实现标准的集成关联，如图 9 所示。

### 3.4 商品与企业信息的关联

基于百度地图 API，以提示词的方式，商品标签信息加工系统实现了被加工商品与企业信息的集成关联，如图 10 所示。

### 3.5 商品标签的细节指标

应用 JSON 数据格式，基于非关系型数据库 MongoDB，商品标签信息加工系统针对商品标签中反映的具体指标的半结构化信息进行存储，具体功能如图 11 所示。数据加工人员可以自由地



图 7 标签信息集成关联加工主页面



图 8 依据主流电商信息进行商品名称提示词功能

增加字段名称和数值。

### 3.6 成果数据

完成上述加工过程后，商品标签信息加工

系统形成了最终成果数据，如图12所示。其中：“属于”表示商品对分类的属于关系；“包含”表示商品对图片的所属关系；“发布”表示商品的企

产品标准号 添加

国家标准号 ▾

GB 8

GB 8898-2011

GB 8372-2008

GB 8537-2018

企业信息

名称  地址  标签(如生产商)

图9 商品执行标准补全提示词

企业信息 添加

名称  地址  标签(如生产商)

省火山鸣泉 key 生产商

火山鸣泉天然苏打水齐市总代理

其他 黑龙江省绿金集团火山鸣泉绿色天然矿泉水有限公司 加

图10 商品企业信息补全提示词

商品名称 火山鸣泉矿泉水 产品标准号 添加 产品图片

产品编号 692577370002 国家标准号 ▾ GB 8537-2018

产品分类

饮用水

水

饮料冲调

牛奶乳品

饮料

咖啡/奶茶

蜂蜜/柚子茶

冲饮谷物

成人奶粉

粮油调味

企业信息 添加

名称 地址 标签(如生产商)

江省绿金集 黑龙江省齐 生产商

山鸣泉生态 北京市朝阳 运营商

其他 水源地 添加

配料 天然矿泉水

净含量 470ml

保质期 18个月

锂 0.3-1.0mg/L

钾 0.8-1.8mg/L

钠 40.0-100.0mg/L

钙 10.0-30.0mg/L

镁 2.0-4.0mg/L

偏硅酸 15.0-30.0mg/L

碳酸氢根 150.0-350.0mg/L

溶解性总固体 200.0-500.0mg/L

PH值 7.8(±0.6)

水源地 黑龙江省克东县二克山(火山)西侧

提交

图11 商品标签中指标细节的加工

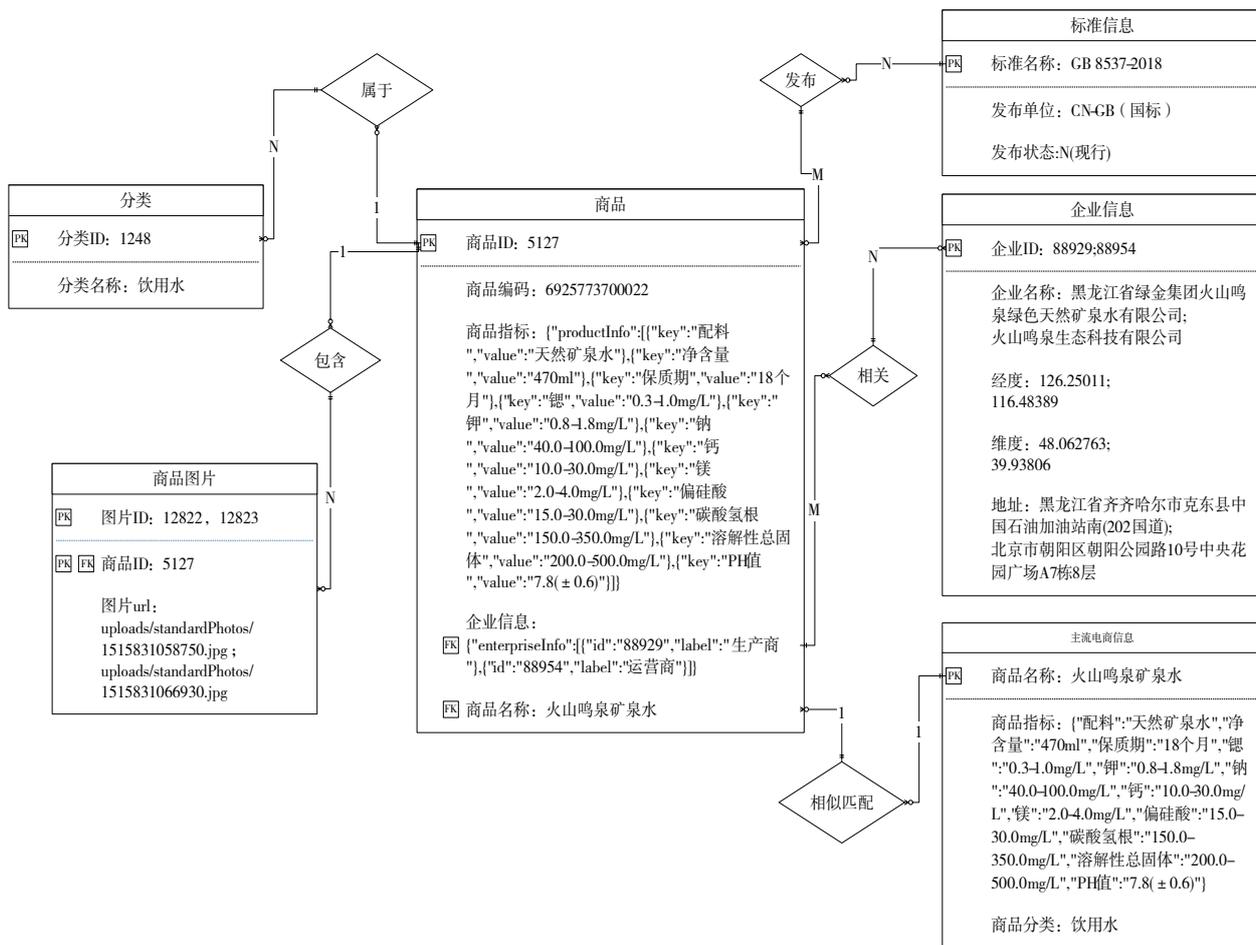


图 12 商品标签加工的成果数据示例

业标准被特定企业发布；“相关”表示商品的企业的具体信息与企业信息表的关联关系；“相似匹配”表示商品名称与主流电商商品名称的相似匹配。1,N/M代表数据库中实体之间的对应关系，1:N为实体间的一对多关系，N:M为实体间的多对多关系。

#### 4 结论

(1) 本文设计构建的商品标签信息加工系统集成了 SpringMVC 和 ServiceComb 的微服务框架，松耦合地集成了多个数据源，应用提示词技术实现了数据的关联和集成。同时，阐述了应用 JSON 格式，并基于 MongoDB 非关系型数据库的 Bson 格式，对商品标签的细节指标信息进行统一加工和存储，解决了关系型数据库无法存储结构不确定的数据格式的问题，最终顺利地完成了真实的商品标签信息的加工和存储。

(2) 商品标签信息加工系统与传统的加工系统的区别：一是基于微服务的多源数据的集成和关联；二是针对标签信息字段不确定的特点（如企业信息包括生产企业、销售企业、经销商、代加工厂、委托商等，指标信息包括脂肪含量、蛋白质含量等各种指标描述），将数据以 JSON 的“键—值 (KEY-VALUE)”的方式存入 MongoDB 数据库，解决了关系型数据库无法存储未知字段的数据问题。

(3) 实证分析表明，商品标签信息加工系统通过与多个数据源的成功关联，将商品和国家标准馆的商品标准数据整合，加工后的数据能够体现商品质量的高低。本系统目前已经进行推广，并且在投入生产加工之中，后期会进一步完善。

(4) 相关部门在为录入企业商品标签信息的系统提供本文描述的半结构化数据存储的方案时，建议为企业提供更多商品指标细节填报的入口，

使得商品的质量量化工作形成流程化和常态化。

### 参考文献

- [1] 张伟.多源异构大数据汇聚共享平台技术研究[C]//中国航天电子技术研究院科学技术委员会.第十五届中国航天电子技术研究院学术交流会优秀论文集.中国航天电子技术研究院科学技术委员会:航天电子发展战略研究中心,2018:131-138.
- [2] 吕林.基于MongoDB的应用平台的研究与实现[D].北京:北京邮电大学,2015.
- [3] 龚兰兰,凌兴宏.基于敏捷开发的SSM Web应用开发实践[J].实验技术与管理,2020,37(2):160-163,167.

- [4] Apache ServiceComb. Apache ServiceComb Java Chassis 用户手册[EB/OL]. [2019-05-30].<https://www.book-stack.cn/read/ApacheServiceCombJavaChassis/11.md>.
- [5] ZEN L, 刘云华, 马彬. 微服务顶级开源项目的创立和成长[J]. 软件和集成电路, 2019(4): 40-45.
- [6] 吴晓龙. 基于微服务架构的在线学习系统设计与实现[D]. 济南: 山东师范大学, 2019.
- [7] BOURHIS P, REUTTER J L, VRGOC D. JSON. Data model and query languages[J]. Information Systems, 2020(89): 101478.
- [8] 王华志. 基于JSON的异构数据源数据交换技术研究[D]. 武汉: 武汉理工大学, 2015.

(上接第47页)

出,要突出高校的主导地位,利用中心产学研协同的优势,建立中心独特的人才培养机制,改变以往“学历高”即为人才的评价方法,定点培养符合协同企业要求的人才。

### 参考文献

- [1] 唐小旭. 区域产学研结合技术创新研究[D]. 哈尔滨: 哈尔滨工程大学, 2009.
- [2] 王璐, 黄敏学, 肖橹, 等. 社会资本、知识利用与共有协同创新绩效[J]. 科研管理, 2018, 39(11): 79-87.
- [3] 李晓娣, 张小燕. 区域创新生态系统对区域创新绩效的影响机制研究[J]. 预测, 2018, 37(5): 22-28, 55.
- [4] 赵丽梅. 基于管理平台构建的高校学科之间协同创新运行体系研究[J]. 商业经济, 2018(9): 177-178, 181.
- [5] 宋伟, 康卫敏, 赵树良. 我国协同创新研究的知识图谱分析: 基于CSSCI(1998-2017)数据[J]. 西南民族大学学报(人文社科版), 2018, 39(6): 226-234.
- [6] 张健. 京津冀区域产学研协同创新绩效评价系统研究[J]. 商业经济, 2018(4): 35-36, 39.
- [7] 僧俐洋. 创新要素对中关村示范区协同创新绩效的影响研究[J]. 经营与管理, 2018(9): 84-86.
- [8] 徐喆, 李春艳. 我国科技政策演变与创新绩效研究: 基于政策相互作用视角[J]. 经济问题, 2017(1): 11-16, 102.
- [9] 严红. 政产学研协同创新绩效提升路径探索[J]. 科教文汇(上旬刊), 2018(10): 4-5.

- [10] 孙新波, 张大鹏, 吴冠霖, 等. 知识联盟协同创新影响因素与绩效的关系研究[J]. 管理学报, 2015, 12(8): 1163-1171.
- [11] 高少冲, 丁荣贵. 首席专家项目匹配度、组织网络特征与协同创新绩效[J]. 科学学研究, 2018, 36(9): 1615-1622.
- [12] 方齐. 科技服务业服务创新过程与绩效关系实证研究[J]. 科学学与科学技术管理, 2015(9): 117-125.
- [13] 方炜, 王莉丽. 协同创新网络的研究现状与展望[J]. 科研管理, 2018, 39(9): 30-41.
- [14] 郝向举, 薛琳. 产学研协同创新绩效测度现状及方法改进[J]. 科技管理研究, 2018, 38(11): 1-5.
- [15] 解学梅. 企业协同创新影响因素与协同程度多维关系实证研究[J]. 科研管理, 2015, 36(2): 69-78.
- [16] 朱健, 谢雨珊, 王辉. 地方高校协同创新平台的组织结构与运行机制: 基于湖南省行业产业类协同创新中心的案例分析[J]. 中国高校科技, 2018(9): 23-27.
- [17] 张运华, 梁郁, 吴洁, 等. 企业内部创新与产学研合作创新协同度模型及其应用研究[J]. 情报杂志, 2018, 37(2): 201-207, 193.
- [18] 何海燕, 王子文, 姜李丹, 等. 我国产学研协同创新影响因素研究: 基于OrderedLogit模型实证分析[J]. 华东经济管理, 2014(9): 106-110.
- [19] 丁欣茹, 胡永红. 协同创新绩效评价体系分析研究及启示[J]. 教育教学论坛, 2018(19): 72-74.
- [20] 王钰云, 陈军冰, 蔡宇青, 等. 产学研协同创新内涵探析及系统行为分析[J]. 江苏科技信息, 2016(16): 4-6.
- [21] 周晓阳, 王钰云. 产学研协同创新绩效评价文献综述[J]. 科技管理研究, 2014, 34(11): 45-49.