中国包装产业大数据知识图谱应用系统的设计

廖立君[1]，吴岳忠[2,3]，李长云[2,3]
（1.长沙学院 计算机工程与应用数学学院，长沙 410022；
2.湖南工业大学 计算机学院，湖南 株洲 412007；
3.湖南省智能信息感知及处理技术重点实验室，湖南 株洲 412007）

摘要：目的 针对目前包装产业存在的产业链长、数据大而散、包装领域知识不全面等问题，设计一个中国包装产业大数据知识图谱应用系统。**方法** 从行业高度定义涵盖包装领域全生态的分类体系，结合人工智能中知识图谱的最新技术，对政府信息、工商信息、行业信息、学术论文、全球包装专利等互联网上各类数据进行自动采集汇聚，抽取出知识信息，融合成一个涵盖资讯、政策、会议、标准、论文、专利、企业、产品、高校、机构和专家等十几类信息的包装知识图谱知识库。**结果** 系统主要功能包括数据采集、知识图谱和终端应用，实现了包装产业大数据的图谱探索、产业链图、数据报告和关联搜索。**结论** 该系统使用方便，可从多视图、多维度获取包装产业相关数据，提升行业的数字化和信息化水平，加速中国包装行业的智能化产业升级，促进包装产业逐步向智能、绿色、集约、创新方向发展。

关键词：包装产业大数据；知识图谱；图谱探索；关联搜索；全生态链
中图分类号：TB489 **文献标识码：**A **文章编号：**1001-3563(2019)21-0140-11
DOI：10.19554/j.cnki.1001-3563.2019.21.021

Design of Big Data Knowledge Graph Application System for China Packaging Industry

LIAO Li-jun[1], *WU Yue-zhong*[2,3], *LI Chang-yun*[2,3]
(1.School of Computer Engineering and Applied Mathematics, Changsha University, Changsha 410022, China;
2.School of Computer Science, Hunan University of Technology, Zhuzhou 412007, China;
3.Key Laboratory for Intelligent Information Perception and Processing Technology, Zhuzhou 412007, China)

ABSTRACT: The work aims to design a big data knowledge graph application system for China packaging industry in terms of the problems existing in the packaging industry, such as the long industrial chain, large and scattered data, and incomplete knowledge in the packaging field. From the industry level, the comprehensive classification system covering the whole field of packaging was defined, and the latest technology of knowledge graph in artificial intelligence was combined to automatically collect and aggregate various types of data on the Internet such as government information, business information, industry information, academic papers, and global packaging patents, and extract knowledge into a knowledge base of packaging knowledge graph covering information, policies, conferences, standards, papers, patents, companies, products, universities, institutions and experts et al. The main functions of the system included data acquisition,

收稿日期：2019-04-30
基金项目：国家重点研发计划（2018YFB1700204）；湖南省重点领域研发计划（2019GK2133）;智能信息感知及处理技术湖南省重点实验室开放课题（2017KF07）
作者简介：廖立君（1973—），女，硕士，长沙学院副教授，主要研究方向为包装大数据、软件自动化和工业物联网。
通信作者：吴岳忠（1981—），男，博士，湖南工业大学副教授，主要研究方向为知识图谱和包装大数据。

knowledge graph and terminal application, which realized graph exploration, industry chain diagram, data report and association search of big data in packaging industry. The system which is easy to use can obtain relevant data of packaging industry from multi-view and multi-dimension to improve the digitalization and informationization level of the industry, accelerate the upgrading of intelligent industry in China packaging industry, and promote the development of the packaging industry to gradually move toward smart, green, intensive and innovative directions.

KEY WORDS: big data in packaging industry; knowledge graph; graph exploration; association search; whole ecological chain

随着社会信息化的发展，人们获取信息的来源主要是互联网，特别是在大数据、“互联网+”和“工业4.0”迅速发展的浪潮中，谁拥有数据就拥有了开启未来大门的钥匙。综上可知，如何从浩如烟海的互联网数据中提取有价值的信息成为了目前的研究热点。

2016 年 12 月 20 日，中国包装联合会发布《中国包装工业发展规划（2016—2020 年）》[1]并提出了发展重点：面向建设包装强国的战略任务，坚持自主创新，突破关键技术，全面推进绿色包装、安全包装、智能包装一体化发展，有效提升包装制品、包装装备、包装印刷等关键领域的综合竞争力。2017 年 7 月 8 日，国务院《新一代人工智能发展规划》[2]明确提出了“建立新一代人工智能关键共性技术体系”的重点任务，特别强调了“研究跨媒体统一表征、关联理解与知识挖掘、知识图谱构建与学习、知识演化与推理、智能描述与生成等技术，开发跨媒体分析推理引擎与验证系统”的关键共性技术。在包装行业，存在产业链长、数据大而散、包装领域知识不全面等问题，其领域数据具有大量、高速、多样、价值、真实性等大数据的 5V 特征。知识图谱可以把海量、多源、异构数据汇聚得到一个丰富的语义关系网络，因此，找到一种新的知识引擎技术对引领包装行业互联网+的形成，加速中国包装行业的产业升级和结构转型具有非常重要的意义。

现已有很多学者在包装领域应用计算机技术、大数据分析技术和人工智能技术，并取得了一些成果。王志伟[3]对智能包装技术及应用进行综述并讨论了智能包装与智能包装系统；马爽等[4]研究了专家系统在智能包装中的应用；Klose[5]研究了跨系统存储数据的方法；赵瑞可等[6]构建了产品包装设计数据管理系统；王昕兵[7]设计了一个不同情境下常态产品包装交互系统；邓礼全[8]设计了针对于饮料企业的包装管理信息系统；李同英等[9]提出了一种基于增强学习的自适应共振结构神经网络算法，可应用于分布式包装实时数据库；金颖磊等[10]研究了基于可拓语义分析的文化创意产品设计方法；吴隔格等[11]研究了基于本体的包装设计辅助决策支持方法。Li 等[12—13]研究了基于多源异构数据构建包装知识图谱的方法和基于多层神经网络的包装域实体命名技术；朱文球等[14]提出了一种能处理自然语言查询的、基于知识图谱的中国包装产业数据库查询方法；杨芳权[15]设计了一个基于包装产业大数据知识图谱的智能问答系统；张华等[16]基于信息设计背景进行了包装大数据可视化研究。文中拟基于知识图谱对包装产业大数据进行有效的研究和应用，构建包含包装学科、包装教育、包装产业等信息的综合性大数据知识图谱，从而推动包装行业逐步向绿色、集约、创新、可持续方向健康发展，加快我国包装工业从世界包装大国向世界包装强国的转变。

1　系统总体架构

1.1　系统需求分析

中国包装产业大数据知识图谱主要是针对包装行业全产业链的企业、高校、人员等电子资源和外部资源，以各种应用子系统和功能组件为服务实现方式，通过门户网站对外提供包装产业、行业相关的资源和应用服务，为用户提供登录认证、资源检索与获取、个性化等服务。应用系统主要面向 3 类用户人群：政府、行业协会人员；企业、从业人员；包装行业研究人员。

1）政府和行业协会用户。主要包括包装行业协会、政府部门等，针对政府和行业协会类用户缺乏包装行业完整权威统计数据及报表，希望了解包装产业发展趋势的需求，提供趋势分析功能；对包装产业热点追踪的需求，提供热点统计分析功能；对希望了解包装产业上下游产业链情况的需求，提供上下游产业分析功能。

2）企业用户。主要包括包装供应企业、包装需求企业和包装销售企业等，针对企业用户关心的包装产业最新动态，提供资讯订阅功能；针对包装供应及销售企业希望了解自己下游客户、上游供应商企业信息的需求，提供图谱搜索分析、知识卡片功能；对具有创新能力的企业，希望了解包装产业技术进展的需求，提供学术科研检索功能。

3）个人用户。主要包括包装研究人员、包装从业人员等，针对个人用户希望一站式检索最新资讯、会议等包装信息的需求，提供产业检索功能；针对如何

直观地发现包装知识之间的关联信息，提供包装知识关联搜索分析和知识推荐功能。

另外还存在一些公共需求，包括相关政策与法规的查询，文献及标准的下载，对如何更加智能方便地获取包装知识的需求，提供智能问答搜索功能。

以上业务功能均建立在包装产业知识图谱之上，因此对于知识图谱的构建和更新，需要提供可视化的图谱定义及编辑工具、数据自动化采集和集成工具。另外，包装产业大数据知识图谱还可以在后期应用于人工智能等其他应用，提供灵活的扩展能力。

1.2 系统功能划分

从需求出发，中国包装产业大数据知识图谱应用系统包括数据采集、知识图谱、终端应用和用户管理等 4 大功能模块。数据采集模块主要包括自动采集和信息众筹等 2 种方式，自动采集方式是对政府信息、工商信息、行业信息、学术论文、全球包装专利等互联网上各类数据进行自动采集汇聚；信息众筹方式是先由高校、企业、机构填报众筹信息，然后经系统管理员审核生效。知识图谱模块主要是进行包装领域知识图谱的构建和更新，包括本体定义、知识抽取和知识融合等功能。终端应用主要分为 4 个特色功能模块，即图谱探索、产业链图、数据报告和关联搜索。用户管理主要是对系统用户进行登录、注册及其权限等管理。具体系统功能见图 1。

1.3 系统体系结构

基于知识图谱构建的自动化程度，可以分为人工构建方法、自动构建方法和半自动构建方法。考虑到完全由人工构建本体需要耗费大量的人力物力，且构建的本体难以随着互联网的信息变更而更新以致迅速老化；完全自动化的知识图谱构建虽难以实现，但自动构建可利用各类知识获取技术、机器学习技术以及统计技术等自动地从已经存在的数据资源中获取本体知识，进而降低本体构建的成本。综上，在包装大数据实际应用中，通过借助包装领域专家的指导，可实现一个半自动的知识图谱构建与更新。中国包装产业大数据知识图谱应用系统体系结构有 5 层，包括知识存储层、信息采集与清洗层、知识抽取层、知识计算层、图谱应用层。知识存储层通过图数据库 MongoDB 存储和全文检索服务器 Elasticsearch 分布式文件索引；信息采集与清洗层主要对文本、百科、行业垂直网站等进行信息采集与清洗；知识抽取层通过 D2R 映射、文本知识学习等方式进行；知识计算层包括知识编辑、知识融合、知识计算等；图谱应用层是包装大数据的功能界面入口，对数据和服务集成，实现包装产业大数据的各接口组件能够为内外部用户提供多层次上集成，实现重用，以满足不同类型的用户对包装知识访问的服务需求，为各机构企业提供高效便捷的知识访问支撑。具体见图 2。

1.4 系统业务流程

根据系统实际需求和功能分析，文中从用户角色进行业务流程分析，包括系统各类管理员和用户及拥有的相应权限。例如数据管理员负责采集数据的清洗和抽取工作。包装产业大数据知识图谱业务流程具体见图 3。

1.5 系统数据模型

结合业务需求和流程分析，系统的数据库采取 Mongo DB 和 My SQL 混合数据库，整个系统数据模型见图 4。

2 主要功能模块设计

2.1 数据采集

1）自动采集功能。这是系统的主要数据采集方

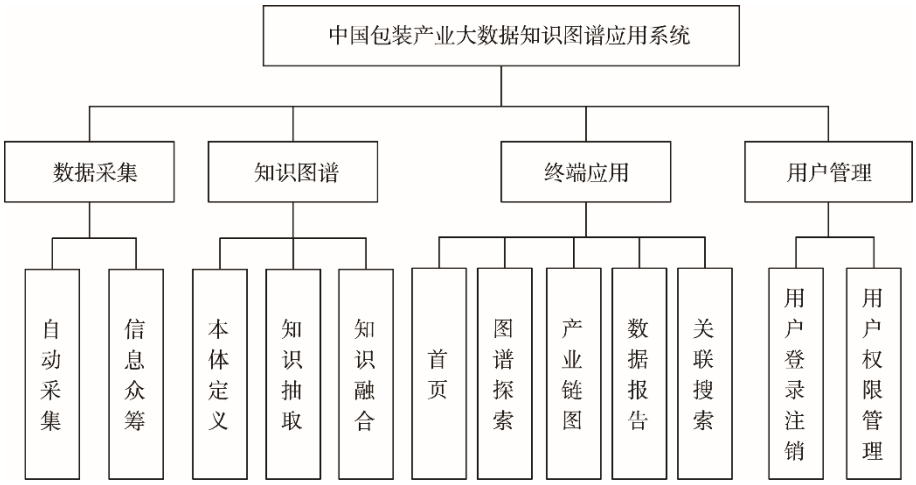


图 1 系统功能

Fig.1 System function

图 2　中国包装产业大数据知识图谱应用系统体系结构

Fig.2 Construction of big data knowledge graph application system in China packaging industry

式。系统通过 7×24 h 持续智能化采集网络包装数据，支持自监督/无监督方式自动学习语料采集，采集互联网上的百科、互联网网页等各类语料，其准确度≥85%、召回率≥70%。此外，可提供可配置的采集策略，并支持结构化、半结构化和非结构化信息的手动和自动抽取，包括分布式采集爬虫部署、可视化采集配置、采集数据预处理与清洗、采集系统检测与管理、采集流速控制、采集数据更新机制、采集服务优先级设置和采集服务数据存储。自动采集配置见图 5。

2）信息众筹功能。企业、高校、机构用户可以自主进行自身信息的填写，结合系统获取工商、专利、论文等权威数据，共同构建中国包装产业大数据知识图谱。

2.2　知识图谱

使用 Li 等[12]提出的基于多源异构数据构建包装知识图谱的方法，通过本体定义、知识抽取和知识融合的领域知识图谱构建架构，按照从行业高度定义的涵盖包装领域全生态的分类体系，围绕包装产业相关的“人”、“物”和“事”，构建了湖工大包装工程大数据知识图谱，见图 6。

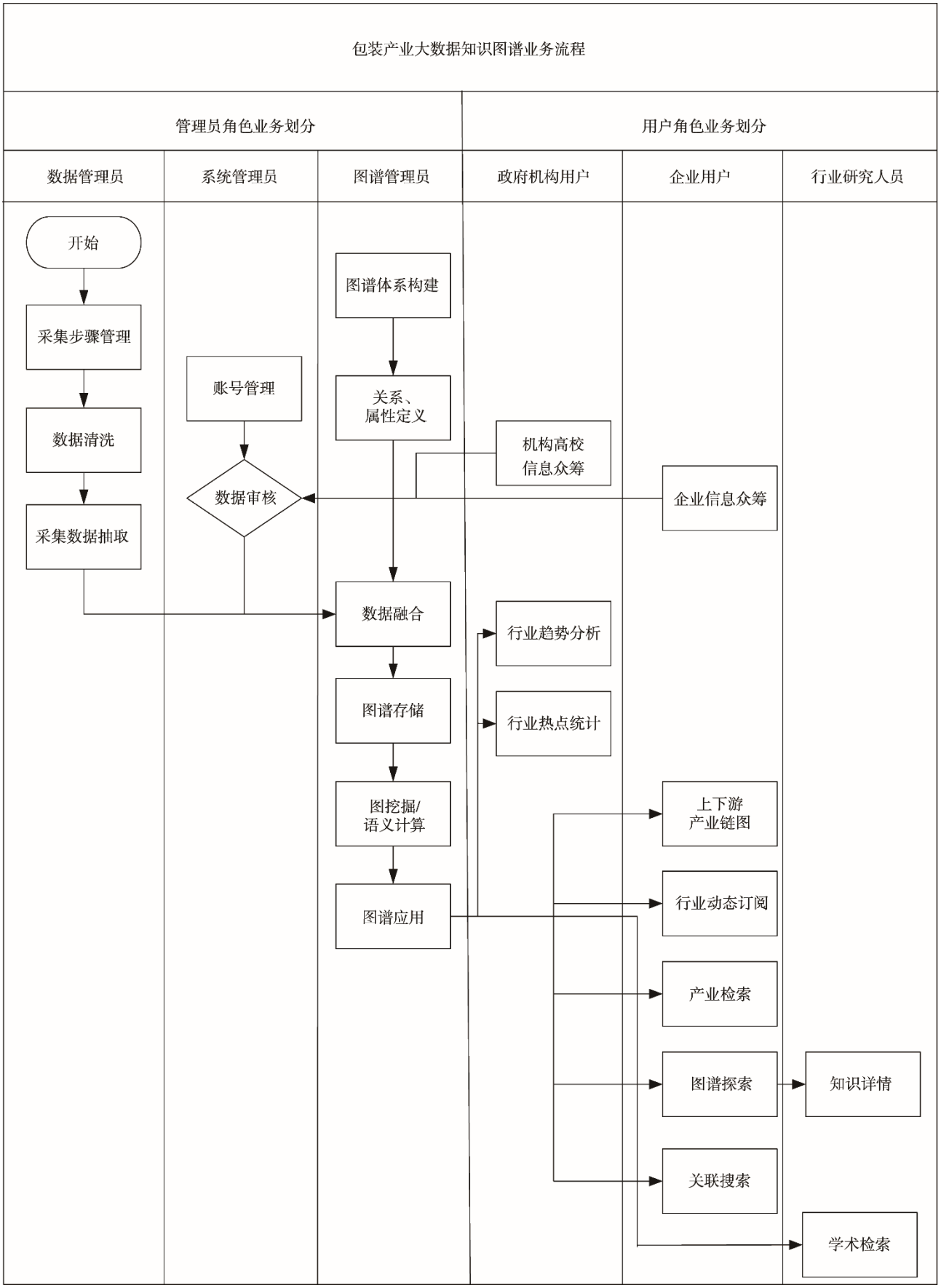


图 3 系统业务流程

Fig.3 Business process of system

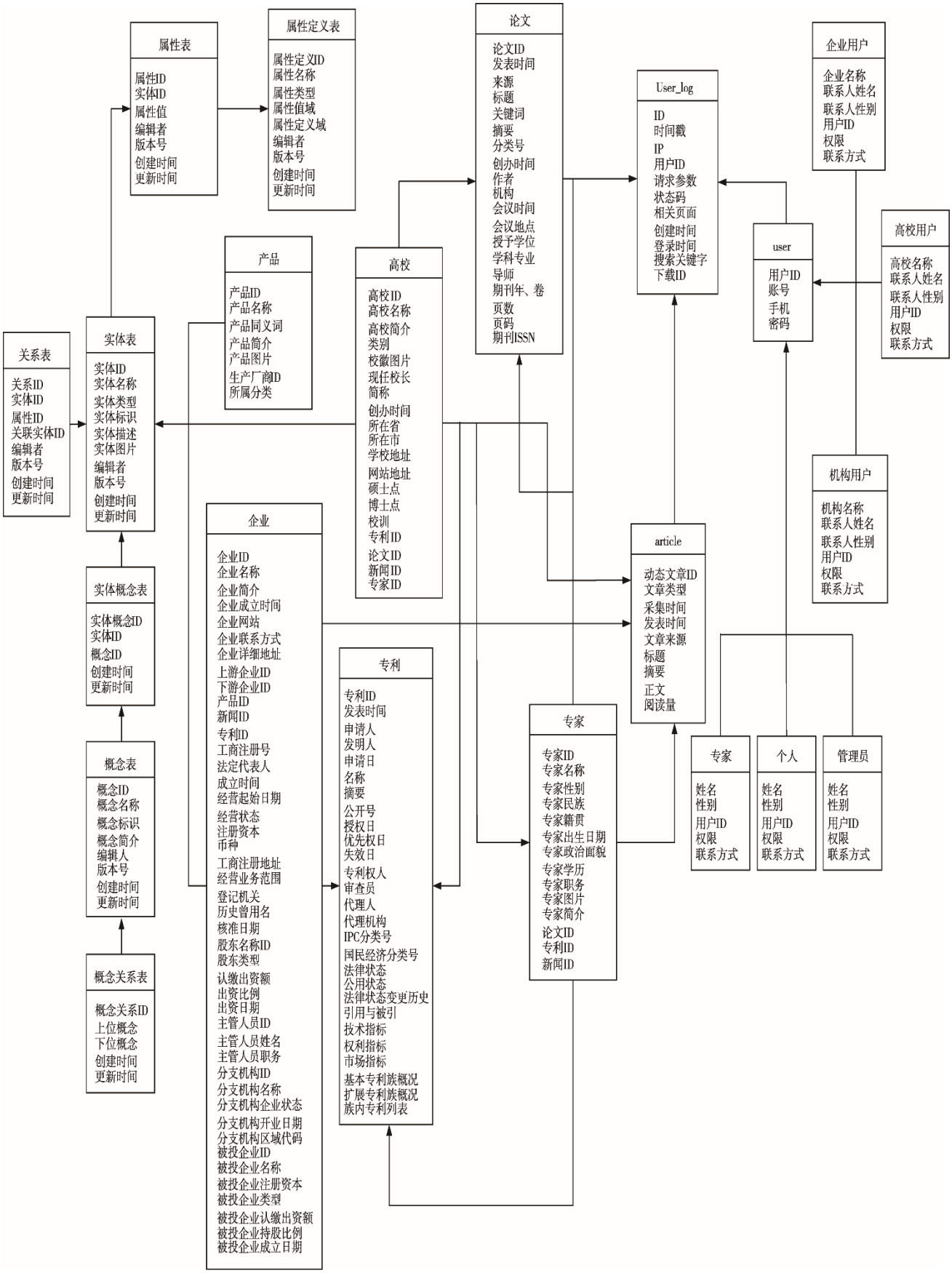


图 4　系统数据模型

Fig.4 System data model

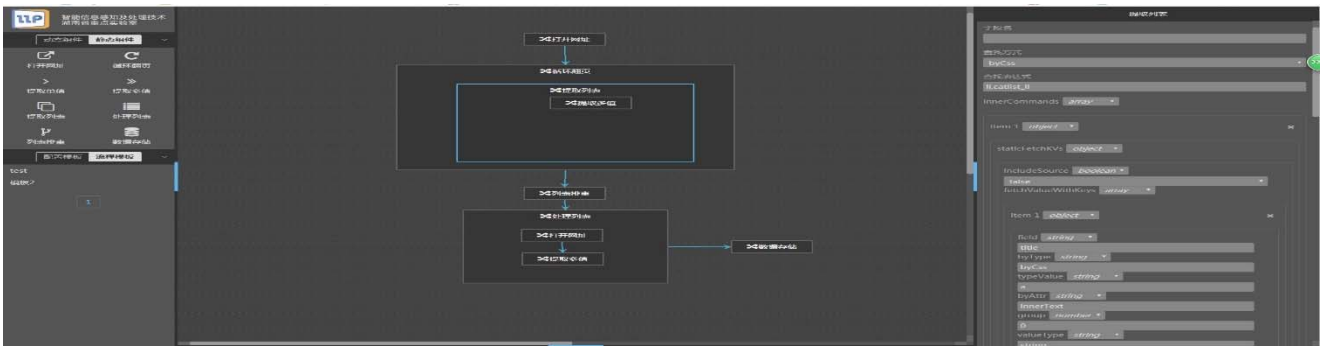


图 5　自动采集配置

Fig.5 Automatic acquisition configuration

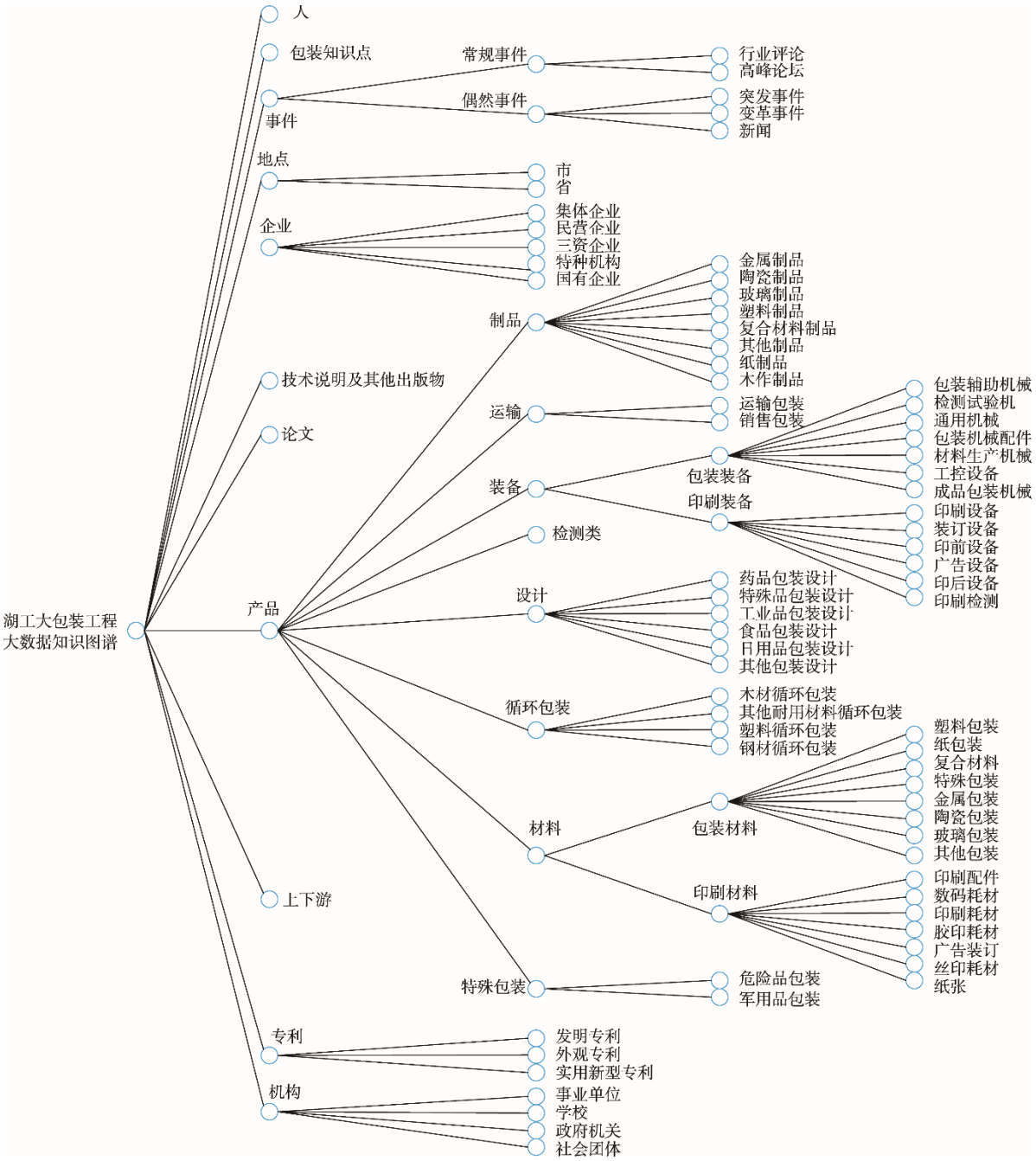


图 6　包装产业知识图谱

Fig.6 Knowledge graph in packaging industry

2.3　图谱应用

2018 年 5 月 17 日，中国包装联合会发文《关于推广使用“中国包装产业大数据知识图谱”平台的通知》（中国包联综字[2018]18 号），向所有中国包装联合会会员企业免费推广使用，网址为 http://58.20.192.198:8090。

1）系统部署。系统的具体部署见表 1。

表 1　系统部署
Tab.1 System deployment

分类	选型
技术选型	服务器开发技术：JDK1.8 运行环境：JDK1.8 数据访问：JDBC
开发平台	使用Bootstrap开发框架， 结合成熟的开源框架组件
中间件	应用服务器：tomcat 7 以上版本
数据库	知识图谱数据库： MongoDB 3.4以上版本 索引数据库：Elasticsearch 5.4以上
开源软件	Cent OS 7.0

2）图谱探索。图谱探索提供包装企业、专家、产品之间网络关系的图谱化展示，用户可以直接在网页中浏览包装产业大数据知识图谱，轻松直观地查看包装相关知识，也可改变筛选过滤的层次（1—3 层），进行拖动查看，双击穿梭展示另一家公司的图谱信息。文中以中国包装联合会副会长单位“湖南千山制药机械股份有限公司”为例，见图 7。

3）产业链图。产业链图是以产品链为中心的产品上下游产业链图，其可以从宏观整体的角度提供对包装行业的概览，通过产业链图定义的关系，发现企业和产品之间的供给与需求关系，辅助实现包装产业价值的最大化利用。文中以搜索产品“三片罐”为例，显示关联的产品生产厂家及其上下游产品信息，可点击查看具体详情，见图 8。

4）数据报告。数据报告利用统计算法对系统内包装数据进行多维的分析计算，进而提供产业热点分析、产业技术热词、创新企业、高校学习能力排名等可视化图表，纵观行业趋势及发展状况。文中可视化数据分析见图 9a，展示了专利热点分析、包装产业技术词云图和包装创新企业排名等方面的数据统计情况；个性化数据分析以企业“湖南千山制药机械股份有限公司”为例，展示了该企业基本信息、相关技术词云图、专利、论文、专家等方面的数据统计情况，见图 9b。

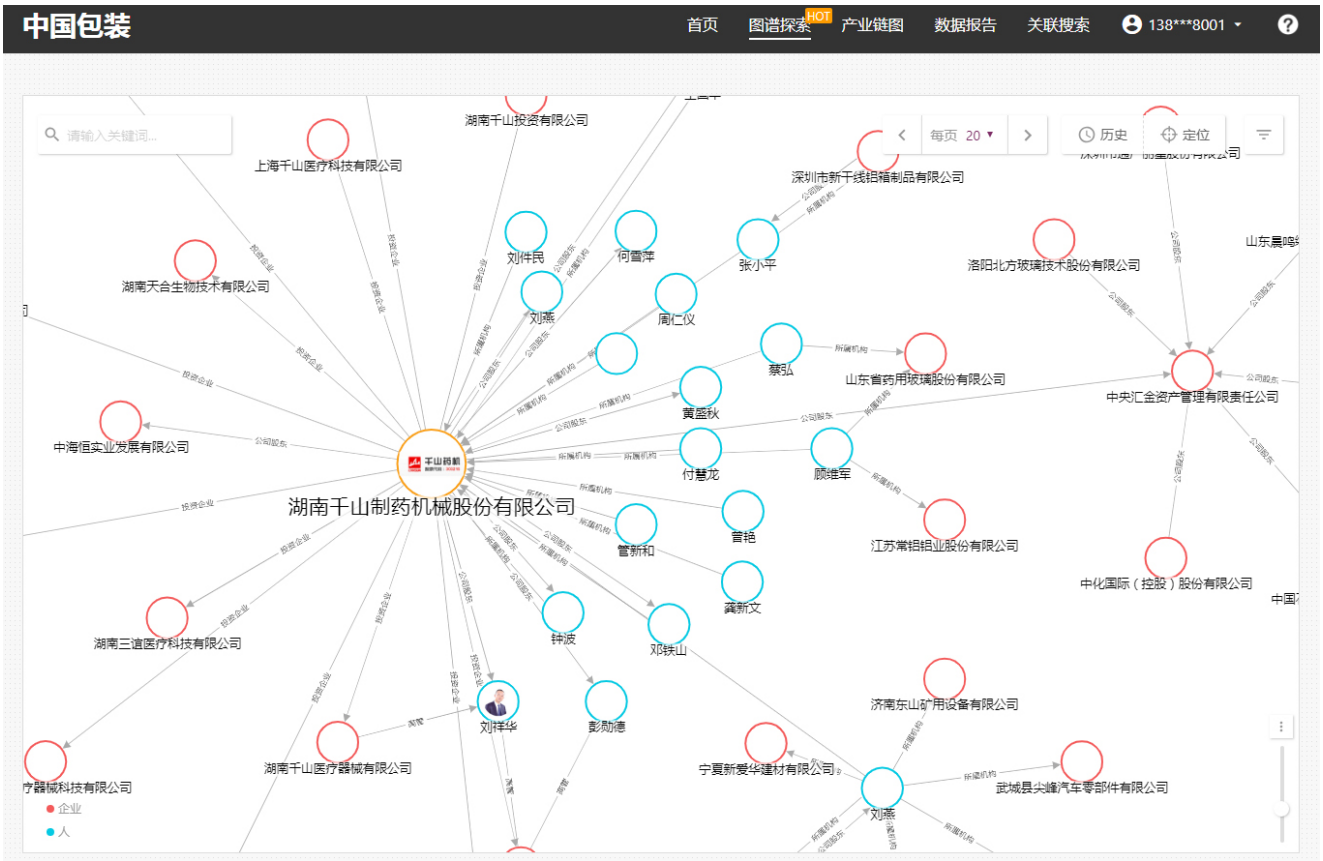
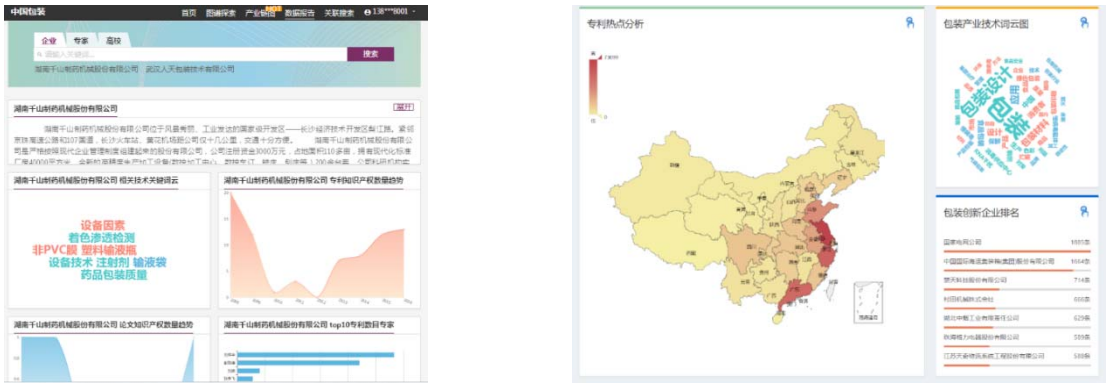


图 7　图谱探索
Fig.7 Graph exploration

5）关联搜索。基于六度空间理论，发现了企业、产品、高校、机构、专家多种实体中 2 个或多个实体间 6 步以内指定条件下的关系搜索；利用图挖掘算法，支持图统计和图解读，帮助用户厘清复杂关系中的有效信息。文中以 3 家包装企业“湖南千山制药机械股份有限公司”、“深圳市通产丽星股份有限公司”和“湖南乐福地医药包材科技有限公司”关联关系为例，见图 10。

图 8　产业链

Fig.8 Industrial chain



a 可视化数据分析　　b 个性化数据分析

图 9　数据报告

Fig.9 Data report function

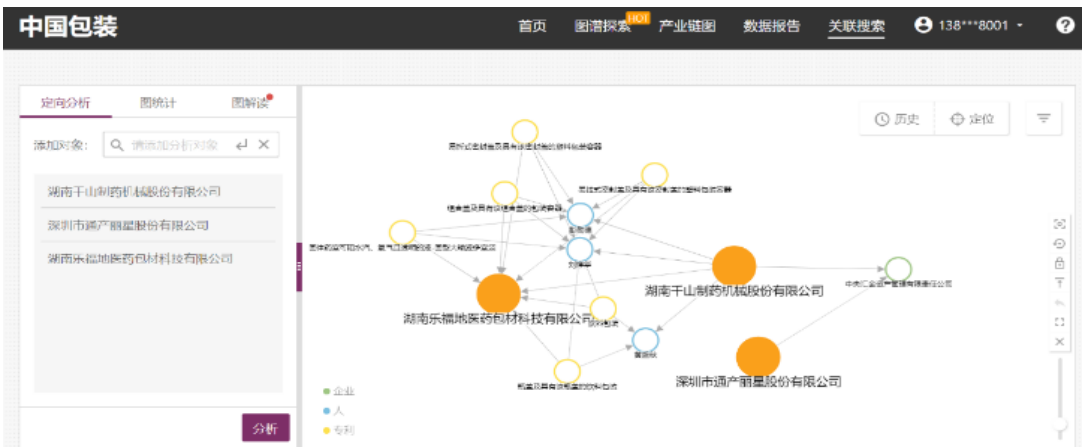


图 10　关联搜索

Fig.10 Association search

3　结语

文中简述了中国包装产业大数据知识图谱应用系统的设计及实现方法。该系统解决了中国包装产业大数据知识图谱的关键技术，包括包装产业大数据分类体系、知识图谱模式、体系架构、数据抽取与挖掘等关键技术问题。针对包装产业知识和数据的建模问题，开发了一种基于本体的包装知识库编辑平台；为解决从不同来源、不同结构的包装产业数据中进行知识提取问题，研制了一套支持智能化自动采集和信息众筹机制的自动学习引擎；针对包装行业存在产业链较长、数据较为分散和结构迥异等特点，提出了基于包装全生态分类体系的多端、多设备图谱探索技术；分析和挖掘包装数据，利用统计算法，进行多维的分析计算，提供细粒度、多视图、多维度可视化包装产业大数据分析技术；充分利用包装知识体系与相关标注体系，形成语义搜索机制，实现了基于语义的智能问答引擎和知识自动推荐引擎技术。系统在十余家包装企业单位得到了成功应用，包括湖南千山制药机械股份有限公司、深圳市通产丽星股份有限公司、济南兰光机电技术有限公司等，累计产生经济效益 3000 余万元。该系统对进一步促进大数据和知识图谱在包装行业的应用，提升行业的数字化、智能化水平起到了积极作用。

参考文献：

[1] 中国包装联合会. 关于印发《中国包装工业发展规划(2016-2020 年)》的通知[EB/OL]. (2016-12-20). http://www.cpta.org.cn/articleDetail.html?id=6821.
China Packaging Federation. Notice on Printing and Distributing the "China Packaging Industry Development Plan (2016-2020)"[EB/OL]. (2016-12-20). http://www.cpta.org.cn/articleDetail.html?id=6821.

[2] 国务院. 国务院关于印发新一代人工智能发展规划的通知[EB/OL]. (2017-07-08). http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm.
State Council. Notice of the State Council on Printing and Distributing a New Generation of Artificial Intelligence Development Plan. [EB/OL]. (2017-07-08). http://www.gov.cn/ zhengce/content/2017-07/20/ content_5211996.htm.

[3] 王志伟. 智能包装技术及应用[J]. 包装学报, 2018, 10(1): 27—33.
WANG Zhi-wei. Intelligent Packaging Technology and Its Application[J]. Packaging Journal, 2018, 10(1): 27—33.

[4] 马爽, 王迪功, 和克智, 等. 专家系统技术在产品包装要求智能决策系统中的应用[J]. 包装世界, 2002(2): 19—20.
MA Shuang, WANG Di-gong, HE Ke-zhi, et al. Application of Expert System Technology in Intelligent Decision System of Packaged Product Requirements[J]. Packaging World, 2002(2): 19—20.

[5] KLOSE M F. Cross-system Storage Management for Transferring Data Across Autonomous Information Management Systems: US, US9648100[P]. 2017-09-19.

[6] 赵瑞可, 胡德敏, 朱娟. 产品包装数据管理信息系统[J]. 计算机系统应用, 2012, 21(8): 15—18.
ZHAO Rui-ke, HU De-min, ZHU Juan. Product Packaging Design Data Management System[J]. Computer Systems & Applications, 2012, 21(8): 15—18.

[7] 王昕兵. 不同情境下常态产品包装交互系统设计[J]. 现代电子技术, 2019, 42(8): 140—144.
WANG Xin-bing. Design of Normal Product Packing Interactive System in Different Situations[J]. Modern Electronics Technique, 2019, 42(8): 140—144.

[8] 邓礼全. 饮料企业产品包装管理信息系统分析[J]. 物流技术, 2015, 34(9): 275—277.
DENG Li-quan. Analysis on Product Packaging Management Information System of Beverage Enterprises[J]. Logistics Technology, 2015, 34(9): 275—277.

[9] 李同英, 朱洪波. 分布式包装实时数据库 ARS 算法应用[J]. 包装工程, 2017, 38(11): 88—91.
LI Tong-ying, ZHU Hong-bo. ARS Algorithm Application of Real-time Database of Distributed Packaging[J]. Packaging Engineering, 2017, 38(11): 88—91.

[10] 金颖磊, 潘伟杰, 吕健, 等. 基于可拓语义分析的文化创意产品设计方法研究[J]. 工程设计学报, 2017, 24(1): 27—33.
JIN Ying-lei, PAN Wei-jie, LYU Jian, et al. Study on Cultural and Creative Product Design Method Based on Extension Semantics Analysis[J]. Chinese Journal of Engineering Design, 2017, 24(1): 27—33.

[11] 吴隔格, 石宇强, 姜辉, 等. 基于本体的包装设计辅助决策支持方法[J]. 包装工程, 2015, 36(3): 59—64.
WU Ge-ge, SHI Yu-qiang, JIANG Hui, et al. Support Method Used to Assist Decision Making in Packaging Design via Ontology[J]. Packaging Engineering, 2015, 36(3): 59—64.

[12] LI C Y, WU Y Z, HU F H. Establishment of Packaging Knowledge Graph Based on Multiple Data Sources[J]. Revista De La Facultad De Ingeniería, 2017, 32(14): 231—236.

[13] LI C Y, WU Y Z, HU F H, et al. Packaging Domain-based Named Entity Recognition with Multi-layer Neural Networks[J]. Neuro Quantology, 2018, 16(6): 564—569.

[14] 朱文球, 司元. 基于知识图谱的包装产业信息查询技术架构[J]. 计算机科学与应用, 2017, 7(9): 858—868.
ZHU Wen-qiu, SI Yuan. Packaging Industry Information Inquiry Technology Architecture Based on Knowledge Graph[J]. Computer Science and Application, 2017, 7(9): 858—868.

[15] 杨芳权. 基于包装产业大数据知识图谱的智能问答系统设计[J]. 现代电子技术, 2018, 41(4): 143—146.
YANG Fang-quan. Design of Intelligent Q-A System Based on Big Data Knowledge Map of Packing Industry[J]. Modern Electronics Technique, 2018, 41(4): 143—146.

[16] 张华, 吴岳忠. 信息设计背景下的包装大数据可视化研究[J]. 湖南工业大学学报(社会科学版), 2018, 23(5): 7—14.
ZHANG Hua, WU Yue-zhong. On Visualization of Packaging Big Data Under the Background of Information Design[J]. Journal of Hunan University of Technology (Social Science Edition), 2018, 23(5): 7—14.