

我国包装产业大数据知识图谱的构建

doi:10.3969/j.issn.1674-7100.2018.04.013

罗学明¹ 陈 一²

1. 湖南工业大学
计算机学院
湖南 株洲 412007
2. 湖南工业大学
城市与环境学院
湖南 株洲 412007

摘要: 在包装逐渐成为我国支柱产业的背景下,依据实用与全面性原则、针对性原则、时效性原则、可追溯与数据关联性原则,构建以包装产业链为主线多层次立体数据模式,建立动态与静态相结合的核心数据库,设置多情景的可视化知识图谱模式,将极大地方便包装产业信息的传播和查询,从而推动包装产业的发展。

关键词: 包装产业;大数据;知识图谱;可视化

中图分类号: TP751.1

文献标志码: A

文章编号: 1674-7100(2018)04-0088-06

1 大数据与知识图谱

目前,大数据已经成为出现频率极高的热词。被誉为“数据仓库之父”的比尔·恩门(Bill Inmon)在20世纪90年代开始关注大数据,当时它被称作海量数据。近几年,互联网、移动设备、物联网和云计算等相关技术的迅猛发展,使得各种数据呈几何级数增长,传统的数据库、数据仓库管理系统等已不能适应数据管理的需要。因此,大数据的概念应运而生并得到产业界和学术界的广泛关注,目前相对成熟的Hadoop系列解决方案和哈佛大学2012年8月的一项研究成果——“DNA硬盘”都是人类对大数据问题积极探索的成果。大数据在学术界的正式提出始于2008年9月《自然》杂志发表的*Big Data—Science in the Petabyte Era*论文^[1]。该文章提出,数据洪流为大数据,要解决大数据的获取、存储、处理、检索和使用问题,首先必须正确理解大数据的概念及特性。维基百科对于大数据的解读:互联网企业日常运营所生成和积累用户网络行为数据量的增长已突

破传统计量单位,使用现有的数据库管理工具难以实现数据的获取、存储、搜索、共享、分析和可视化等,故称之为大数据。据此,人们将该定义概括为4V:更大的容量(volume)、更复杂的多样性(variety)、更快的生成速度(velocity)以及三者组合带来的第4个因素——价值(value)^[2]。

大数据现象最早出现在物理学、生态学、生物学、自动控制学等学术性领域,后来逐渐出现在军事、通讯、金融等行业领域。近年来,随着互联网、移动设备、物联网和云计算等相关技术的迅猛发展,互联网上的数据量开始呈现爆发式增长趋势。网络用户激增且在线时长延长导致用户行为数据飞速增长,网络服务数据量随着网络服务多元化而大幅扩展;同时,网络终端的扩展也为数据的产生带来了更多的渠道。世界网络业的领导型企业IBM、Facebook、谷歌、苹果和亚马逊都已在大数据的运用上处于领先地位。大数据在商业上的应用价值已体现在诸多经典案例中,如美国印地安那大学和英国曼彻斯特大学的学者将twitter上的公众情绪变化数据所产生的曲线与道琼斯工业

收稿日期:2018-03-11

基金项目:国家工商行政管理总局基金资助项目(2014GSZJW001KT012)

作者简介:罗学明(1981-),男,湖北荆州人,湖南工业大学讲师,硕士,主要研究方向为大数据,

E-mail: luoxueming0733@qq.com

通信作者:陈 一(1982-),男,湖南衡阳人,湖南工业大学副教授,博士,主要从事包装产业与包装材料方面的研究,

E-mail: yiyue514@aliyun.com

指数进行对照分析,发现可以提前3~4 d预测股市大盘走势,由此推出了欧洲第一只基于社交媒体的对冲基金。目前,有许多商业巨头在依靠大数据开发新的商业项目,如阿里巴巴旗下的阿里信贷就是依靠其拥有的海量客户经营数据全面掌握用户的信用水平,从而降低信用风险。商业零售业已成为大数据的主战场,大家在网上购物时应该会有这样的体验,今天买的东西或浏览过的网页第二天就会出现你浏览各网站的主推位置。京东商城、当当网、亚马逊、淘宝网等电商网站商品交易数据量庞大,电商数据分析部门通过数据挖掘,构建关联模型,可以更好地组织网站上的商品,根据顾客当前的购买意向向顾客推介商品,以满足顾客需求,并为顾客提供更好的体验。麦肯锡全球研究院详细研究了医疗、公共服务、零售、制造和个人定位等5个领域内大数据的发展趋势,认为大数据可以使任何一个领域创造价值^[3-5]。

知识图谱是一种新近出现的数据表现形式,在图书情报界也称为知识领域可视化或知识领域映射地图,是通过可视化技术描述知识资源及其载体,挖掘、分析、构建、绘制和显示知识及知识发展进程和结构关系的一系列图形化方法^[6]。该方法的核心在于其运用可视化技术,形象地展示学科的核心结构、发展历史、前沿领域及其整体知识架构,从而为学科研究提供切实的、有价值的参考^[7-8]。目前,知识图谱已被运用于多个学科领域,尤其是在学科文献关联、信息汇总方面发挥了巨大的作用。大数据结合知识图谱,既能够实现数据的丰富性和全面性,又能够提供可视化的表现模式,与传统的数据查询和知识图谱相比,大数据知识图谱具有明显的优势。因此,把大数据知识图谱运用于包装领域,将对包装产业的发展起到巨大的促进作用。

2 我国包装产业的特点及现状

2.1 特点

包装是知识密集型行业,是多学科交叉的科学,是门槛虽低,但天花板效应很高的行业。包装产业较其他产业具有鲜明的特点,表现在以下几个方面。

1) 产业规模庞大。根据2007年的统计数据,全世界每年仅包装材料与容器的消费总额就超过8 000亿美元,尤其是近年来电商的飞速发展,进一步带动了包装容器需求的增长^[9]。2016年我国包装工业总产值超过1.7万亿人民币^[10],继续巩固了世界第二包

装大国的地位,包装已经成为我国国民经济重要的支柱产业,这还不包括与商品关联而产生的包装隐形价值。

2) 产业链纵深延展广阔、涉及行业门类众多。包装是一个行业,也是一个大的产业,其上下游产业链纵深广,包含了多个小的行业门类,且互为上下游。以烟草包装为例,其制作工序非常繁锁,就纸包装成型环节而言已涉及许多行业,需要研制自动化纸包装黏合设备、开发专用黏合剂、制备外包双向拉伸聚丙烯膜以及研发防伪技术;纸包装成型后还需要进行多条烟的整体包裹和堆码。因此,每一种包装产品的开发,均涉及多个产业和门类。

3) 涉及学科知识多,交叉性强。包装看似平常,到达消费者手上的包装成品作为商品附属品往往容易让人忽略其中的技术含量。一个普通的商品包装所涉及的知识和技术至少包括材料学、设计学、机械学、数字媒体技术等,而每一门学科又涉及到许多门类的知识与技术,如材料学就涉及包装材料的开发及其机械加工问题。因此,每一个包装的过程都是一个多学科协同的过程。

2.2 现状

近10年来,我国包装产业的体量已经提高了近5倍,逐渐成为我国的支柱性产业,包装的概念也随时代的进步而变化。最早的时候包装是为了保护产品、方便储运、促进销售,到后来发展到品牌识别再到更深层次的集产品信息化、用户体验于一体的“商品外的商品”,包装在商品的销售中起到了越来越重要的作用。随着社会的发展,消费者对包装的功能性要求越来越高。特别是在国际贸易过程中,包装的防锈、抗菌、环保等功能已变成了最基本的要求。因而包装已不仅仅是商品的附属物,而是独立的、能够产生一定价值的商品。然而,我国的包装行业在包装品种、产品质量、新品研发能力以及经济效益等方面,均与发达国家存在较大的差距。包装产业发展与转型困难,尤其是中小企业的发展举步维艰。

1) 包装企业规模小,产品技术含量低且无力单独承担技术攻关。国内的大部分包装企业属于中小企业,技术含量不足,主要高档包装设备和原辅材料严重依赖进口,高水平研发人员少,技术研发经费投入严重不足,创新能力差,往往无法独立或小联合解决生产中的技术瓶颈,更无法进行技术创新。

2) 产业集中度低,信息渠道不通畅。产业集聚和产业集中度是区域产业竞争力的重要体现。若区域产

业高度集群,产业集中度高,形成完整的服务和产业链,即能大幅提升产业综合竞争力。然而,我国的包装企业往往紧跟制造业,分布零散,集中聚度低,至今没有形成具有一定规模的专业包装工业园,导致信息渠道不通畅,在一定程度上制约了包装工业的发展。

3) 制造企业普遍存在与包装相关的诉求,但对接成难题。中小制造型企业生产产品,尤其是生产消费产品,因其本身体量较小,在包装上均存在一定的个性化诉求。尤其是针对新产品的包装,在包装材料、工艺和设计上均遇到一些共性或个性问题,由于受到企业规模和专业性等方面的限制,其自身不可能实现包装的研发。如何寻找有能力的包装企业和相关专家对接,已成为一个普遍的问题。

4) 公共服务平台少,无法满足包装产业各层面的需要。我国包装行业受到本身特点的限制,目前国内虽然拥有一些针对小门类的数据库、服务和研发平台,也在单一的包装类别上有了一站式服务的企业和机构,但是仍然缺少提供全方位服务的包装专业平台,也没有真正意义上的相对完备的包装专业数据库。因此,包装企业有了问题不知道找谁,不知道哪些资源可以利用,甚至不知道问题出在哪里。这既耗费了资源,也无法快速解决问题,严重制约了包装企业和行业的发展。

综上,包装行业具有多学科交叉、宽口径融合、涉及巨量行业企业的特性,但存在单体体量小、技术能力匮乏、信息渠道不通畅的问题以及对信息通畅、产业融合的巨大需求。基于此,构建一个行业内的大数据知识图谱,使其囊括包装行业的主要信息,并发挥知识图谱简单、实用、可视化的优势,为包装行业服务无疑是非常有意义的。

3 包装产业大数据知识图谱的构建原则与模式

3.1 构建原则

基于包装产业的特点,我们认为,包装产业大数据知识图谱的构建应基于以下几个原则:

1) 实用与全面性原则。实用是产业知识图谱的必要条件,包装行业的知识图谱应有别于其他单独学科领域的知识图谱,服务于包装产业的知识图谱应具有鲜明的包装产业特征,不但要能用,而且要用好。同时,要做到产业全覆盖,即包装产业大数据知识图

谱必须实现实用性基础上的全面性。

2) 针对性原则。包装产业大数据知识图谱的使用群体包括全包装产业链的各个领域,其设计者应该充分了解处于产业链不同位置人群的需求,甄别不同数据的价值,并针对不同人群的需求提供不同的数据图谱呈现。

3) 时效性原则。时效性是所有信息利用领域必须遵循的原则,只有具有时效性的信息才可以发挥其价值。产业发展瞬息万变,大数据图谱的内容也应实时更新以保证数据的时效性。

4) 可追溯与数据关联性原则。数据的关联尤为重要,当使用者无法精确的描述其需求时,建立有逻辑的数据关联以呈现客户的潜在需要(也可表现为一定智能化),实现数据的时间联系即可追溯性。

3.2 构建模式

产业大数据知识图谱的构建在遵循以上原则的基础上需要有清晰的结构和方式,本课题组给出包装产业大数据知识图谱的构建模式以供相关各方参考。

1) 以包装产业链为主线的多层次立体数据模式

为了更好地囊括包装相关产业信息,本课题组以产业链为主线构建了多层次立体数据模式,并采用树状结构模型建立了包装大数据知识图谱,如图1所示。该模型以包装产业链为主线,包装产业链从下游到上游依次包括包装材料行业、包装加工行业、包装设计行业、包装整体开发行业、包装废弃物处理行业等。其中,材料行业包括金属材料、陶瓷材料、玻璃材料、塑料材料、可降解材料等行业;包装加工行业包括塑料等材料加工行业、成型加工装备行业、包装机械行业、自动化控制行业、模具加工行业等;包装设计行业包括结构设计、外观设计、防伪设计等内容;包装整体开发行业包括包装一体化解决服务、包装技术服务等内容;包装废弃物行业包括包装废弃物处理和包装废弃物再利用等。在此基础上梳理出其各自所包含的元素,即每个行业相关的企业、人员、高校、事件、知识点、数据信息等元素,然后提出每个元素上的信息细节属性,如与企业相关的细节属性包括企业生产规模、地域、生产产品情况、客户需求情况等;人员(如科研工作者)的细节属性包括其所属专业学科、所在单位、科研成果、合作企业、社会兼职、获得专利情况等;相关高校的细节属性包括其所属地域、办学类型、相关专业、学科水平、科研机构等;事件、知识点、数据信息等元素的信息细节属性也作相应的分析,

基于包装行业跨度大的特性, 应当运用针对性原则, 去除其中的不直接相关信息。如企业、人员的属性很多, 其信息细节属性的选择应严格贴近包装, 从而有效精简数据库, 以满足特定包装人士的需求。同时,

应当遵照时效性原则选取信息, 包装产业与其他任何产业一样具有变化的特征, 其数据、知识、产业的各种信息都在发生变化, 企业的有关信息一旦过时则失去了价值, 所以该部分实用数据必须做到实时更新。

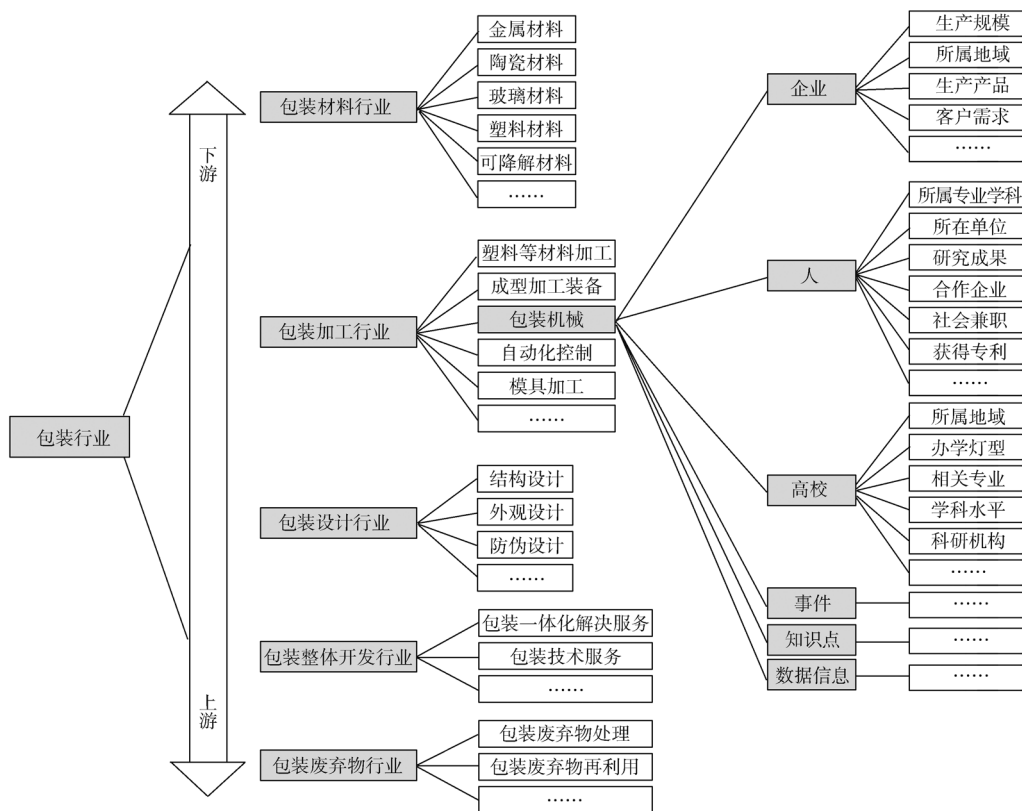


图1 包装大数据知识图谱
Fig. 1 Knowledge map of packing big data

2) 构建动态与静态相结合的核心数据库

在产业数据库的构建中, 基于应用的需要, 数据库应能够满足固定知识和数据信息查询, 也包含实时变化的信息查询。所以, 应构建动态数据库和静态数据库相结合的大数据库, 静态信息主要满足企业对技术和专业包装人士的需求, 而动态数据库则更加关注包装产业、行业大环境的变化, 两者结合才能构成真正实用性的包装大数据平台。

静态数据库应包含包装行业需要用到的不同的、不会变化的技术信息和机构数据, 初拟(但不局限于)以下几种。①设备数据库: 包装设备龙头企业、包装技术中心和包装相关企业的加工、研发、检测设备等信息; 包装高校的研发、检测、加工设备等信息。②标准数据库: 包装相关的国际标准、国家标准、行业标准、有效企业标准数据库。③材料数据库它常用包

装材料物性数据, 型材、板材、膜材等国内外常用材料牌号对照, 常用包装材料选择等。④零配件数据库: 包装机械、设备零配件编号、零配件名称、规格型号、类别、计量单位等。⑤材料加工工艺数据库: 材料挤出、注塑、流延、压制、吹塑等加工方式中的常用参数与实例等。⑥专家人才数据库: 高校、科研院所和国家级、省级企业技术中心与工程中心的教授、专家、高级技工等。⑦企业数据库: 包装企业相关数据, 如所属行业、企业规模、产品等。⑧成果数据库: 高校、科研院所、企业成熟待转化的科技成果与科技成果转化分析报告。动态数据库则应包括各种实时更新的信息, 如行业的数据更新, 包装大事件的记录, 包装企业的变化、供求信息、包装技术报道, 包装相关展会信息, 产业动态, 人员更替等。包装数据库类型及数据如表1所示。

表1 拟建包装数据库类型及数据

Table 1 Proposed type and data of packaging database

数据库类型	数 据
设备数据库	包装设备龙头企业、包装技术中心和包装相关企业的加工、研发、检测设备等信息
标准数据库	包装相关的国际标准、国家标准、行业标准、有效企业标准数据库
材料数据库	常用包装材料物性数据, 型材、板材、膜材等国内外常用材料牌号对照, 常用包装材料选择
零配件数据库	包装机械、设备零配件编号、零配件名称、规格型号、类别、计量单位
材料加工工艺数据库	材料挤出、注塑、流延、压制、吹塑等加工方式中的常用参数与实例
专家人才数据库	高校、科研院所和国家级、省级企业技术中心、工程中心的教授、专家、高级技工
企业数据库	包装企业相关数据, 如所属行业、企业规模、产品
成果数据库	高校、科研院所、企业成熟待转化的科技成果与科技成果转化分析报告

3) 多情景的可视化知识图谱模式

知识图谱的优势在于其具备可视性和关联性特征。在产业大数据知识图谱中, 由于信息的多样化及明显不同人群的需求差异, 其所呈现的可视化界面也应有所区别。如企业主需要了解的信息是企业的供求情况、行业的变化情况、原材料的价格趋势等; 企业技术人员所关注的是技术参数、可检测设备分布、材料加工的条件及相关企业、可提供产品的企业等; 作为新产品的构思、设计人员以及包装研发人员则更关注新技术、新产品和各种展会的信息。所以, 当界面无法全面展示时, 可根据不同的人群及其需求进行分类, 并进行个性化设置, 这将有助于提高可视化知识图谱模式的人性化程度。在实现途径上, 可设置智能问答或输入关键词的方式, 也可通过反向选择剔除不需要的信息种类的方式, 还可设置优选显示的图谱层级等方式, 以提高可视化界面的切合度。

4 结语

包装发展到现在, 时代已经赋予了它更多的含义, 它是商品的延展, 是商品附加值的载体, 是集功能化、艺术化、用户体验于一身的“商品中的商品”。好的包装能够吸引消费者的眼球并帮助企业增强市场竞争力。当然, 优秀的包装是许多要素组合而成的, 在创意、材料、技术等方面的要求极高, 单靠个人或单个企业的力量已经很难高效实现, 而在这个数据化的时代, 通过建立包装大数据平台并将其转化为可视化的知识图谱, 让对包装有个性化需求的人群可以像在淘宝购买商品一样可视化地查找到海量相关信息为自己服务, 了解资讯、搜索技术、查询供求、寻求合作、对接人才, 从而实现过去难以实现的目的。包装产业大数据知识图谱的构建无疑将极大地推动包装行业信息的传播, 提高行业的效率进而促进行业

整体的发展。目前, 此项目正在研发之中, 由湖南工业大学牵头, 多家包装企业参与。希望在不久的将来, 我们可以看到真正实用性的包装产业大数据知识图谱面世, 并为产业发展注入新的动力。

参考文献:

- [1] BUXTON B, GOLDSTON D, DOCTOROW C, et al. Big Data: Science in the Petabyte Era[J]. Nature, 2008, 455 (7209): 8-9.
- [2] 王新才, 丁家友. 大数据知识图谱: 概念、特征、应用与影响 [J]. 情报科学, 2013, 31(9): 10-14. WANG Xincan, DING Jiayou. Mapping Knowledge Domain of Big Data: Concept, Feature, Application and Impact[J]. Information Science, 2013, 31(9): 10-14.
- [3] BOLLIER D. The Promise and Peril of Big Data[M]. Washington: The Aspen Institute, 2010: 3-37.
- [4] 刘学枫. 浅谈大数据背景下的商业应用 [J]. 中国新通信, 2016(15): 109. LIU Xuefeng. A Brief Discussion About the Commercial Application Based on the Big Data[J]. China New Telecommunications, 2016(15): 109.
- [5] 韩梅. 浅谈大数据在经济领域中的应用 [J]. 网络安全技术与应用, 2015(1): 80-81. HAN Mei. A Brief Discussion on the Application of Big Data in Economic Field[J]. Network Security Technology & Application, 2015(1): 80-81.
- [6] 秦长江, 侯汉清. 知识图谱: 信息管理与知识管理的新领域 [J]. 大学图书馆学报, 2009, 27(1): 30-37. QING Changjiang, HOU Hanqing. Mapping Knowledge Domain: A New Field of Information Management and Knowledge Management[J]. Journal of Academic Libraries, 2009, 27(1): 30-37.
- [7] 胡泽文, 孙建军, 武夷山. 国内知识图谱应用研究综述 [J]. 图书情报工作, 2013, 57(3): 131-138. HU Zewen, SUN Jianjun, WU Yishan. Research

- Review on Application of Knowledge Mapping in China[J]. Library and Information Service, 2013, 57(3): 131-138.
- [8] 徐增林, 盛泳潘, 贺丽荣, 等. 知识图谱技术综述 [J]. 电子科技大学学报, 2016, 45(4): 589-606.
XU Zenglin, SHENG Yongpan, HE Lirong, et al. Review on Knowledge Graph Techniques[J]. Journal of University of Electronic Science and Technology of China, 2016, 45(4): 589-606.
- [9] 许文才. 包装工业的现状与发展趋势 [J]. 印刷工业, 2007(4): 56-58.
XU Wencai. Current Situation and Development Trend of Packaging Industry[J]. Print China, 2007(4): 56-58.
- [10] 义峰. 我国包装工业总产值达 1.7 万亿元 [J]. 广东印刷, 2016(6): 67.
YI Feng. China's Packaging Industry Has a Total Output Value of 1.7 Trillion Yuan[J]. Guangdong Print, 2016(6): 67.

Construction of Big Data Knowledge Map in China's Packaging Industry

LUO Xueming¹, CHEN Yi²

(1. School of Computer Science, Hunan University of Technology, Zhuzhou Hunan 412007, China;
2. School of Urban and Environmental, Hunan University of Technology, Zhuzhou Hunan 412007, China)

Abstract: Under the background of packaging becoming a pillar industry in China, according to the principles of practicality and comprehensiveness, pertinence, timeliness, traceability and data relevance, constructing the multi-level three-dimensional data model by using the packaging industry chain as the main line, establishing the kernel database with dynamic and static data and setting up the multi-scenarios visual knowledge mapping model would greatly facilitate the communication and query of packaging industry information, and thereby promote the development of the packaging industry.

Keywords: packaging industry; big data; knowledge graph; visualization