

基于本体论的食品安全数据融合框架研究

李松函¹, 刘 杨², 生吉萍^{1*}

(1. 中国人民大学, 农业与农村发展学院, 北京 100872; 2. 贵州科学院, 贵州 550001)

摘 要: 庞杂的食品安全抽检数据由于分类编码问题不统一而使得数据融合难以实现, 并进一步阻碍了与其他监测系统的融合与数据分析。解决各种食品安全数据融合的问题将对国家的食品安全监管具有重要意义, 本文归纳总结了食品安全大数据在分类编码方面的主要存在问题, 引入本体论的思想, 并结合计算机工程, 提出了构建基于本体论的计算机工程与大数据分析相结合的总体框架, 以实现数据融合, 服务政府和企业的有效监管。

关键词: 食品安全; 分类编码; 本体论; 大数据; 数据融合

Research on food safety data fusion framework based on ontology

LI Song-Han¹, LIU Yang², SHENG Ji-Ping^{1*}

(1. School of Agricultural Economics and Rural Development, Renmin University of China, Beijing 100872, China;
2. Guizhou Academy of Sciences, Guizhou 550001, China)

ABSTRACT: The complicated food safety sampling data makes data fusion difficult to achieve due to the inconsistency of classification and coding problems. Addressing the integration of various food safety data will be of great importance to national food safety regulation. The paper summarized the main problems in the classification and coding of food safety big data, introduced the ontological idea and combining with computer engineering, proposed an overall framework for constructing an ontology-based computer engineering and big data analysis, so as to achieve data fusion and serve the effective supervision of the government and enterprises.

KEY WORDS: food safety; classification code; ontology; big data; data fusion

1 引 言

统一明确的食物分类和编码是我国实现食品安全智慧监管的基础, 每年庞杂的食品安全抽检数据由于分类编码问题不统一而使得数据融合难以实现, 并进一步阻碍了与其他监测系统的融合与数据分析, 解决各种食品安全数据融合的问题将对国家的食品安全监管具有重要意义。在信息化高速发展的背景下, 统一食品安全领域不同层级不同部门之间的数据, 需要构建基于计算机工程本体论的本体库, 从而助力实现食品安全大数据体系的建立与分析。

本文归纳总结了食品安全大数据在分类编码方面的主要存在问题, 引入本体论的思想, 并结合计算机工程, 提出了构建基于本体论的计算机工程与大数据分析相结合的总体框架, 为实现数据融合, 服务政府和企业的有效监管提供技术支持。

2 食品安全大数据面临的问题

2.1 食品分类编码现状

目前, 为提高我国食品安全的信息化水平, 全国已建立了多个与监测相关的信息系统, 包括食品微生物污染监

基金项目: 国家重点研发计划项目(2018YFC1603300)

Fund: Supported by National Key R&D Project (2018YFC1603300)

*通讯作者: 生吉萍, 教授, 博士生导师, 主要研究方向为食品安全管理与食品经济管理。E-mail: shengjiping@126.com

*Corresponding author: SHENG Ji-Ping, Professor, School of Agricultural Economics and Rural Development, Renmin University of China, Beijing 100872, China. E-mail: shengjiping@126.com

测系统、食品化学污染物监测系统、食源性疾病病例监测系统、食源性疾病暴发事件系统、食源性疾病分子溯源系统、食品安全舆情监测系统和膳食消费量调查系统等。但由于这些信息系统在建立时缺乏整体统一的标准,各部门根据各自情况采用不同的食品分类编码原则,导致这些系统数据难以实现融合共享,大大降低了数据的可利用性和拓展性^[1,2]。

2.1.1 分类名称不统一

我国食品种类极其丰富,各个数据监测系统对食品数据分类灵活且容易交叉,若没有合理清晰的分类原则统一指导,便易造成食品安全信息纷杂混乱、不同系统难以兼容的情况。以食品化学污染物分类树和微生物污染物分类树为例,两者一致的分类名称占极少数,大多数的情况是前者中的分类名称为“焙烤食品”而后的分类名称为

“焙烤及油炸类食品”、前者“坚果及籽类”而后者为“坚果与籽类及其加工制品类”等。这种相对较为混乱的分类名称命名在其他系统中也十分常见,对系统数据融合带来了很大阻力。

2.1.2 种类层级不统一

由于各个食品安全系统是根据不同的监测目标建立的,分类导向不同,因而系统之间分类层级差异巨大(如图 1 所示)。此外,不同分类层级下分别对应着大量的叶节点数和非叶节点数,信息系统十分庞大复杂。

2.1.3 分类逻辑不统一

即使在相同的种类层级下,不同系统的分类逻辑也存在差异。如图 2 所示,在水产品这一相同节点上,化学性污染监测系统将其分为“淡水产品”和“海水产品”,微生物监测系统中则将其分为“动物性”和“植物性”。这主要是

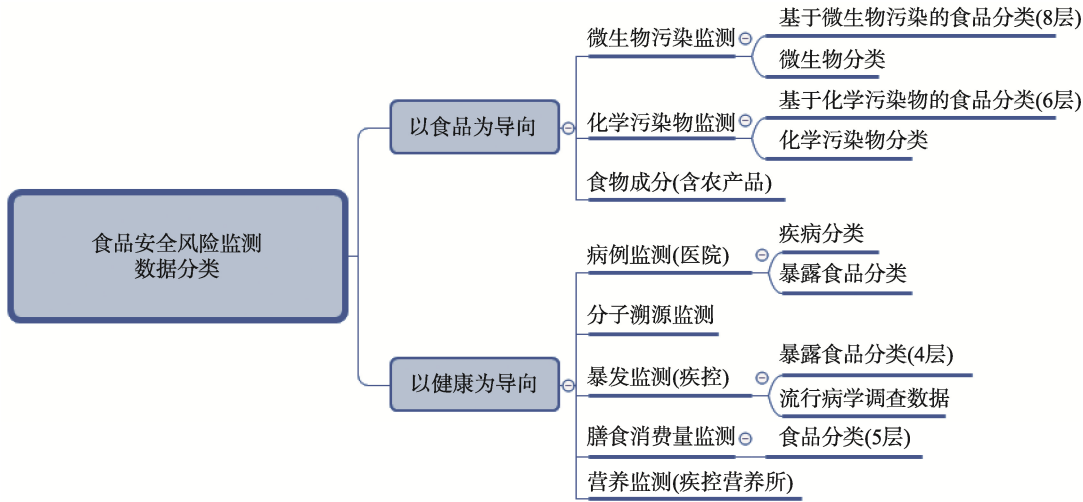


图 1 食品安全风险监测数据分类示意图

Fig.1 Schematic diagram of food safety risk monitoring data classification

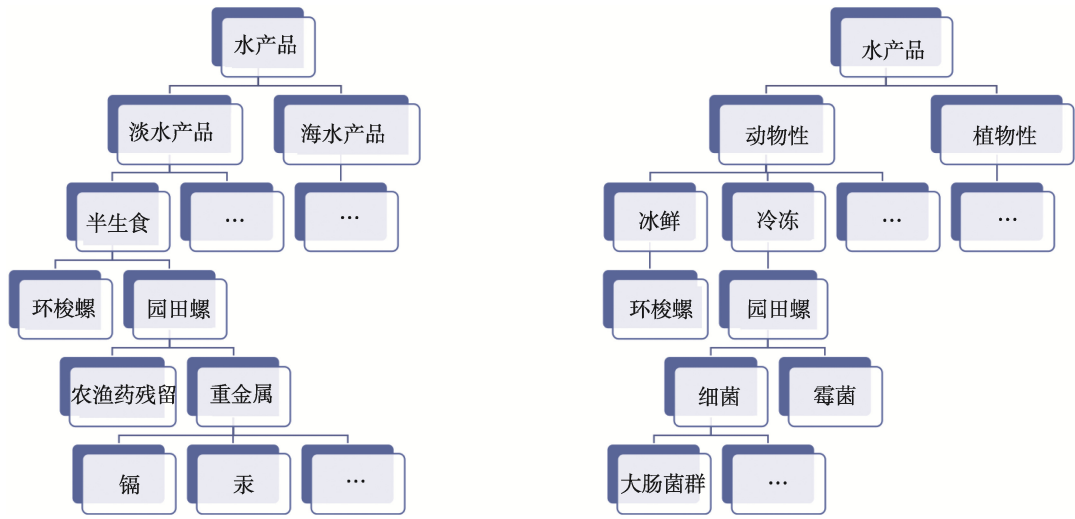


Fig.2 Classification logic of chemical and microbiological systems

因为,前者以化学性污染物为导向将食品按照污染物—食品—食用方式—生活环境路径进行分类,而后者则是以微生物污染为导向,按照微生物—食品—贮藏方式—生物属性的路径进行分类。最终,不同的分类系统中“环梭螺”和“园田螺”虽然为相同的交叉内容,却对应着完全不同的分类路径。

2.2 食品安全数据编码问题

食品分类编码规则的不统一,是当前数据难以融合共享的最大阻碍。国内各系统食品分类编码差异巨大,分类层级、编码位数和编码规则几乎完全不同。其最大的弊端就是无法对系统进行整合、综合利用和分析,形成“信息孤岛”。

2.3 食品属性字段问题

除了分类和编码的不统一,各食品安全监测系统数据记录也存在诸多问题。首先,数量指标单位不统一,例如在某系统中以“克/千克”为单位,而在其他系统中则以“毫克/千克”为单位;其次,检测结果表述方式不统一;另外,数据格式问题极大地影响了数据的准确性,数值型数据中常出现非数字的符号和空格等;最后,数据缺失现象十分严重,在必填项中数据部分缺失的比例达到 75%以上,而非必填项中数据的完全缺失情况超过 25%、部分缺失则占 50%左右。

由于中国饮食文化的多样性,同一种食物或食品在不同地区存在各种不同的称呼,一种食品在 3 个世界中存在不同的存在方式,唯一能将这些本质相同的不同名称或定义统一起来的方法就是运用“本体论”,构建食品的“本体库”,将一个食品个体所有可能的描述映射到同一本体并添加相应描述,在尽可能细化分类标准的基础上继续建立不同本体之间的联系,通过运用本体论建立统一食品分类编码对食品安全大数据的分析具有重要意义。

3 本体论的提出

3.1 本体论思想的发展:

本体论(Ontology)源于早期哲学关于世界的本源和存在研究的理论,是形而上学的一个分支。柏拉图把“存在”分为“现象”和“理念”,创立了“二元论”。亚里士多德在继承其思想的基础上,进行了部分批判,并予以发展,发明了“共相论”。“共相”的意思就是可以用于描述许多主体的东西,共相论要讨论的是泛指的、一类事物的名词。而与“共相”对应的是“实体”,它指个别的、唯一的、独立的个体。往往说“这类”指的是共相,而“这个”指的是实体^[3]。

在现代西方哲学研究中,“本体论”逐渐演变为对“存在”、“自然存在”和“社会存在”三者之间关系的讨论,对传统“我思故我在”的形而上学研究作出了深刻的反思和批判^[4]。

海德格尔通过对存在与存在者之间的差异的揭示,提出了“现象学本体论”的新理论,描述本体论的任务即是使存在从存在者中显露出来,并对存在本身进行解释^[5]。马克思的本体论思想启示我们,既要看到作为人之存在的“此在”与其他“存在者”之间的区别,又要看到在资本主义社会这一历史背景下它们之间的密切的、普遍的联系,马克思的社会存在本体论,就其实质而言,乃是社会生产关系本体论^[6]。

3.2 本体论的现代化应用原理

哲学领域本体论思想的建立与发展为本体工程提供了坚实的理论支撑。19 世纪 90 年代,亚里士多德“共相”的理论被引入到人工智能领域,就是所谓“本体工程”,用来定义领域内的一些概念,以达到知识共享的目的。越来越多的研究人员意识到本体并不仅仅是基于知识的系统,所有的软件开发都需要宏观的总体模型,因此可以在软件设计中运用“本体论”,包括自然语言处理,智能信息检索(尤其是来自互联网),虚拟组织以及模拟和建模等都可以较好应用本体论的程序类别^[7]。

本体工程包括在本体的概念化、设计、实现和部署期间进行的一系列活动。它还为我们提供了知识库的设计理论基础,有助于定义研究领域的基本概念,允许更加规范的知识库设计,并可实现知识的可持续扩充^[8]。

通常,本体是描述如何构建模型的元模型。它的组成部分(它定义的概念以及它们之间的关系)在对知识库的一部分建模时总能作为构建块利用或复用。没有本体,即使使用类似的知识模型,代表相同领域知识的知识库通常也是不兼容的。使用这种元模型的优点是我们不需要丢弃任何有用的模型,本体仅仅简单地为一知识库内的相应模型提供框架,在开发实用软件系统时,本体的元模型功能使原有工具变得智能化。

3.3 本体构造的方法

3.3.1 确定本体应用的研究目标和范围

建立本体的第一步就是确定所研究的领域或任务,建立相应的领域本体或过程本体。领域越大,所建本体就越大,因此需限制研究的范围。由于食品种类多、组合多样,对食品进行的本体库的构建,首先确定范围为食品可能涉及到的原材料,也就是常见的初级农产品及初级加工食品。

3.3.2 本体分析

定义本体所有术语的意义及其之间的关系,此环节需要领域专家的参与,以尽可能完善本体。如马铃薯,在不同的分类系统中名称不同,需要识别马铃薯的所有“身份”,收集其他名称如“土豆”、“洋芋”、“地蛋”等不同表示方法,再报送专家进行校验审阅。

3.3.3 本体建立

构建语义模型表示本体,遵循建立本体的基本标准

(清晰性、一致性、完整性、可扩展性),包括本体中的术语定义无歧义;术语之间逻辑关系应一致;本体中的概念及其关系也应是完整包含该领域内所有概念,但往往很难达到,需不断完善,这就要求本体具有可扩展性,即在该领域不断发展的过程中能加入新的概念。

3.3.4 本体的检验

对所建本体按以上标准进行检验,符合要求的可以文件形式存放,否则返回第二步重新分析^[9],过程如图 3 所示。

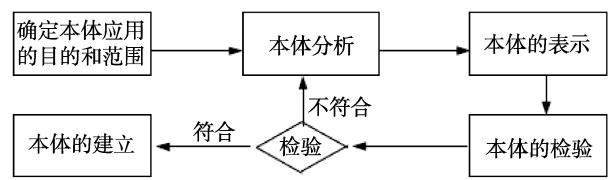


图 3 构造本体的方法
Fig.3 Method of constructing ontology

4 食品安全数据融合的框架

4.1 食品安全数据融合的框架建立

4.1.1 设计目标

基于较为完备的食品分类体系建立完善的本体库,实现全国每年大量的食品安全抽检数据中的各个项目和食品来源及其属性标签在不同数据结构下可以兼容,将数据融合扩展到食品相关监测系统的各个方面,在兼容原有各个食品安全相关监测系统的基础上融合相关数据,并在综

合系统中统一编码分析。

4.1.2 框架模型建立

以本体论的思想为中心,以本体工程理论为框架,对国内常见食品及农产品构造本体,形成食品安全数据本体库,数据一旦产生,就将成为客观物质世界和主观精神世界之外的客观知识及其存在物,具有客观性、自主性和实在性^[10-12]。如图 4 所示,在国家各领域涉及食品的监测系统中,通过检索可与本体库建立一一映射关系,从而形成食品为出发点的“安全数据链条”;如果引入其他领域,如食物营养数据库、食品生产流通数据库等则可以实现食品安全大数据的进一步扩充,为国家建立统一的食品安全大数据信息系统、分析食品安全数据提供框架基础。

4.2 基于本体论的食品安全数据融合应用

目前基于本体论的多源异构数据集成已应用在医疗、交通、公共安全等各个领域,这种多源异构数据集成平台,可以实现对多个数据源进行统一处理,屏蔽数据之间的物理和逻辑差异,实现统一的表示、存储和管理,将多源异构数据集成为互相理解、相互关联的有机整体,以解决数据来源广泛、结构异构问题^[13],从而更好地发挥数据价值。

基于本体论的食品安全数据库的建立,使得各种复杂多样的食品在不同的感性认知和实际状态下具有了标准化数据化的属性,以此达到各种与食品安全相关监测系统的互联互通^[14]。在进行实际监管过程中,执法人员无需掌握编码原则和相关知识,终端录入的一些信息可以被智能识别出其“本体”,进而按照本体库内部设置的标准和原则进行编码,从而实现食品安全数据的统一分析。

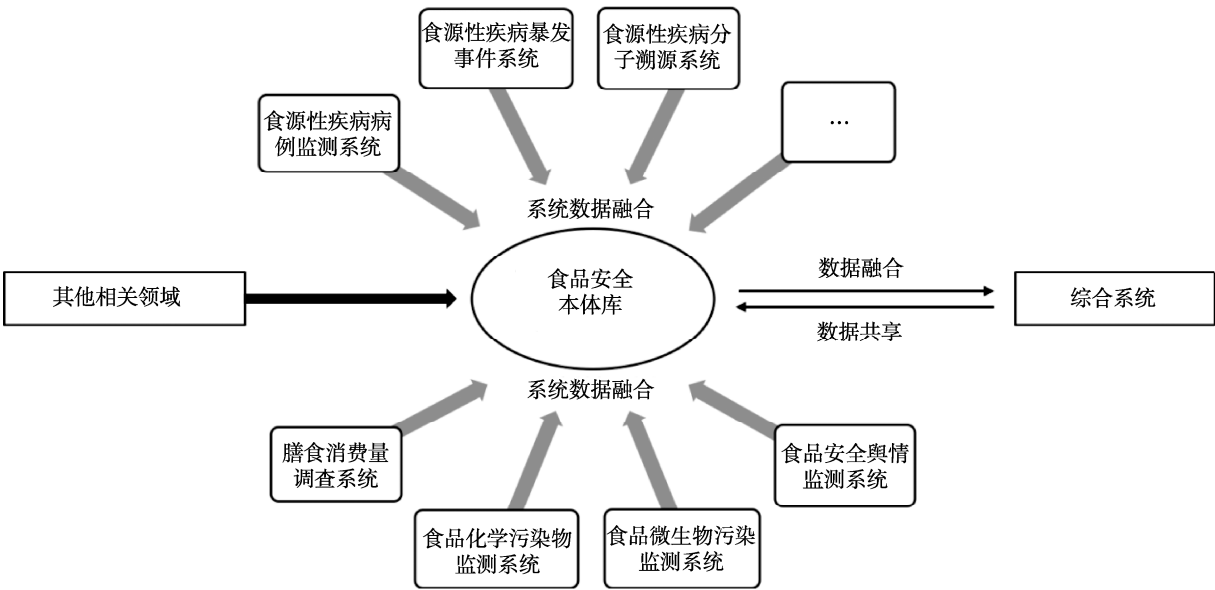


图 4 基于本体论的食品安全数据融合框架
Fig.4 Food safety data fusion framework based on ontology

5 前景展望

基于本体论的食品安全数据融合框架的构建是实现食品安全智慧监管的基石, 此框架的优点在于它能根据数据的增多而不断扩展变化。在未来的研究中, 通过继续加入食品安全风险评价的指标体系可以实现对食品安全风险的快速预警; 通过引入食品的营养组成成分数据, 可以实现科学配餐等消费者的营养目标, 更好地服务消费者; 拓展食品的生产工艺流程和关键控制点, 可以指导企业的风险管理与过程控制等。所以, 基于本体论的食品安全数据融合框架的构建是国家食品安全管理迈向信息化和智能化的第一步也是最重要的一步。

参考文献

- [1] 刘杨, 马东, 肖革新. 食品安全数据融合的实现路径——数据编码[J]. 食品安全导刊, 2018, (31): 42–43.
Liu Y, Ma D, Xiao GX. Data coding is the implementation path of food safety data fusion [J]. China Food Saf Magaz, 2018, (31): 42–43.
- [2] 刘杨, 马东, 肖革新. 基于食品健康链的大数据智能编码系统设计[J]. 中国卫生信息管理杂志, 2018, 15(5): 575–578.
Liu Y, Ma D, Xiao GX. The design of big data intelligent coding system based on food and health chain [J]. Chin J Health Inf Manag, 2018, 15(5): 575–578.
- [3] Nicola G. Formal ontology, conceptual analysis and knowledge representation [J]. Int J Human-Computer Stud, 1995, 43: 625–640.
- [4] 俞吾金. 存在、自然存在和社会存在——海德格尔、卢卡奇和马克思本体论思想的比较研究[J]. 中国社会科学, 2001, (2): 54–65, 205.
Yu WJ. Existence, natural existence and social existence—a comparative study of the ontological concepts of martin heidegger, gyorgy lukacs and karl marx [J]. Soc Sci China, 2001, (2): 54–65, 205.
- [5] 段伟文. 大数据知识发现的本体论追问[J]. 哲学研究, 2015, (11): 114–119.
Duan WW. The ontology of big data knowledge discovery [J]. Philosoph Res, 2015, (11): 114–119.
- [6] 刘红. 大数据的本体论探讨[J]. 自然辩证法通讯, 2014, 36(6): 115–121, 128.
Liu H. Ontology of big data [J]. J Dialect Nat, 2014, 36(6): 115–121, 128.
- [7] Fridman-Noy N, Hafner CD. The state of the art in ontology design: A survey and comparative review [Z].
- [8] Mizoguchi R. A step towards ontological engineering [R]. Proceedings of the 12th National Conference on AI of JSAL, 1998, 24–31.
- [9] 邢平平, 施鹏飞, 赵奕. 基于本体论的数据挖掘方法[J]. 计算机工程, 2001, (5): 15–16, 59.
Xing PP, Shi PF, Zhao Y. Ontology-based data mining [J]. Computer Eng, 2001, (5): 15–16, 59.
- [10] 胡志强. 基于“三个世界”理论的大数据指挥与控制本体论[J]. 指挥与控制学报, 2015, 1(2): 232–237.
Hu ZQ. Ontology of big data command and control based on theory of "three worlds" [J]. J Command Control, 2015, 1(2): 232–237.
- [11] Kim G, Lee D. Intelligent health diagnosis technique exploiting automatic ontology generation and web-based personal health record services in IEEE access [Z]. 2019.
- [12] Maleki E. Ontology-based framework enabling smart product-service systems: application of sensing systems for machine health monitoring in IEEE [Z]. 2018.
- [13] 苏立悦, 董明, 杨东. 大规模定制下基于本体的服务产品配置研究[J]. 计算机应用研究, 2010, (2): 483–484.
Su LY, Dong M, Yang D. Service product configuration of mass customization based on ontology [J]. Appl Res Comput, 2010, (2): 483–484.
- [14] Tome E, Gordana I, Doris P, et al. ISO-FOOD ontology: A formal representation of the knowledge within the domain of isotopes for food science [J]. Food Chem, 2019, 277: 382–390.

(责任编辑: 韩晓红)

作者简介



李松函, 主要研究方向为食品安全管理、食品经济管理。
E-mail: lish9399@ruc.edu.cn



生吉萍, 教授, 博士生导师, 主要研究方向为食品安全管理与食品经济管理。
E-mail: shengjiping@126.com