

装备-标准知识图谱的过程建模研究

尹 亮¹ 何明利¹ 谢文波² 陈端兵^{2,3,4}

(装甲兵工程学院 北京 100072)¹ (电子科技大学计算机科学与工程学院 成都 611731)²
(电子科技大学大数据研究中心 成都 611731)³ (电子科技大学数字文化与传媒研究中心 成都 611731)⁴

摘 要 为了清晰地描述装备、标准以及标准化要素之间的复杂联系,构建装备-标准知识图谱是一种重要的分析手段。利用装备-标准知识图谱,可实现标准化研究从型号跟随到体系引领、从定性分析到定量分析、从单项评审到系统验证的转变,而过程建模是构建装备-标准知识图谱的核心环节之一。文中采用 IDEF3 建模方法,对装备-标准知识图谱的整体架构以及图谱中涉及到的各个子过程进行了建模分析。通过过程建模,得到了装备-标准知识图谱的异质网络模型。

关键词 知识图谱,过程建模,异质网络模型

中图法分类号 TP391 **文献标识码** A

Process Modeling on Knowledge Graph of Equipment and Standard

YIN Liang¹ HE Ming-li¹ XIE Wen-bo² CHEN Duan-bing^{2,3,4}

(The Academy of Armored Forces Engineering, Beijing 100072, China)¹

(School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China)²

(Center for Big Data, University of Electronic Science and Technology of China, Chengdu 611731, China)³

(The Center for Digitized Culture and Media, University of Electronic Science and Technology of China, Chengdu 611731, China)⁴

Abstract In order to clearly describe the complex association between equipment, standards, and standardized elements, it is an importantly analytical tool to construct a knowledge graph of equipment-standard. Using the constructed knowledge graph of equipment-standard, the transformation of standardization research can be achieved from model following to system leading, from qualitative analysis to quantitative analysis, and from individual evaluation to system verification. The process modeling is a key step in the knowledge graph modeling. The IDEF3 method is applied to model the main structure of knowledge graph and the sub-processes involved. A heterogeneous network model of equipment-standard knowledge graph is obtained through process modeling.

Keywords Knowledge graph, Process modeling, Heterogeneous network model

1 引言

现实世界中各事物之间、数据之间均存在各种复杂的关系,采用知识图谱结构能够清楚、直观地描述这些复杂系统。近年来,知识谱图的研究得到了迅速发展^[1-5]。在搜索领域^[6],以 Google^[7]、百度^[8]、搜狐为首的搜索引擎公司,以此为基础构建出描述各种实体和实体之间关系丰富的知识图谱^[9];图书管理领域以作者和图书信息关联数据为基础,建立图书书目的知识图谱^[10],以达到信息重用和数据开放互联的目的^[11];在医学领域,构建了大量的疾病、基因以及蛋白质之间的知识图谱^[12]等。

装备-标准知识图谱作为新兴事物,目前其研究还较少。利用复杂网络模型将装备、标准化要素和标准三者进行有机结合,是实现标准化研究从型号跟随到体系引领、从定性分析到定量分析、从单项评审到系统验证转变的基本手段,也是标

准化工作在装备建设应用中的创新和探索。

装备-标准知识图谱过程建模^[16-18]是装备-标准知识图谱建模的核心,它反映了装备-标准知识图谱的运作特征。围绕装备-标准知识图谱开展的各项研究,均需要先对图谱的构建进行分析。装备-标准知识图谱过程建模的主要目的是根据装备-标准知识图谱的目标和约束条件,通过定义图谱内部标准、标准条目、装备以及装备模块等实体对象之间的逻辑关系,对图谱的构建过程进行刻画,实现对装备-标准知识图谱的构建模式分析、反馈优化、延伸应用等。

过程建模研究主要集中在 3 个方面:过程建模理论与建模方法,过程模型和其他模型的整合方法,模型验证和模型仿真。目前,最常见的建模方法有 IDEF3 (Integrated DEFinition for Process Description Capture Method)^[19]、Petri 网^[20]、状态图、语言行为理论以及基于 ARIS^[21]和面向对象的建模分析方法等。综合考虑装备-标准知识图谱的特点及其定性

本文受国家自然科学基金(61433014,61673085),中央高校基本科研业务费专项资金(ZYGX2014Z002)资助。

尹 亮(1982—),男,博士生,工程师,主要研究方向为指控通信与装甲车辆工程,E-mail:1804359156@qq.com;何明利(1963—),男,硕士,高级工程师,博士生导师,主要研究方向为指控通信与装甲车辆工程;谢文波(1990—),男,博士生,主要研究方向为数据挖掘、数据建模;陈端兵(1971—),男,博士,副教授,硕士生导师,主要研究方向为复杂网络分析、数据建模、大数据挖掘。

建模体系的规划,本文采用 IDEF3 方法进行装备-标准知识图谱的过程建模。

IDEF3 是一种与 IDEF0 互补的业务流程建模方法,是一种场景驱动的流程描述捕获方法,旨在捕获有关特定系统如何工作的知识,是 IDEF 系统和软件工程领域的建模语言系列的一部分。

IDEF3 方法从两个方面对过程模型进行表示:

(1)过程流程描述,捕获特定场景的上下文中的动作之间的关系;

(2)对象状态转换,捕获允许状态和条件的描述。

2 装备-标准知识图谱过程模型的总体设计

装备-标准知识图谱是标准化工作中的重要环节之一,为标准的应用提供成体系、成系统的方法与手段,为新装备的标准化工作评估和管控提供基本依据,为标准制修订或废止提供支撑和依据。

2.1 装备-标准知识图谱中的实体和实体之间的关系

为了厘清知识图谱中的对象,首先需要装备-标准知识图谱中的对象进行详细分析,主要分为实体和实体关系两类。实体对象的描述如表 1 所列,共有 5 类实体,包括标准文档、标准条目、标准化要素、组件/模块以及装备。

表 1 装备标准知识图谱中的实体

序号	实体	简写
1	标准文档(Document)	D
2	标准条目(Item)	I
3	标准化要素(Topic)	T
4	组件/模块(Component)	C
5	装备(Equipment)	EQ

装备-标准知识图谱中的实体关系较为复杂,主要有 6 类,如表 2 所列。

(1)标准文档之间的关系。标准文档之间具有一定的相似性,可能包含了相同的条目,或者使用了相同的标准化要素进行规范。另外,标准文档之间可能存在引用的情况,因此标准文档之间的关系包含了相似关系和引用关系两种情况。

(2)标准条目之间的关系。与标准文档之间的关系类似,标准条目之间也具有一定的相似性,还可能使用相同的标准化要素进行规范。标准条目之间也存在引用的情况,因此标准条目之间的关系也包含了相似关系和引用关系两种情况。

(3)标准文档与标准条目之间的关系。标准文档由多个不同的标准条目组成,而同一项标准条目可能被多个不同的标准文档引用,因此标准文档与标准条目之间的关系为多对多的引用关系。

(4)标准条目与标准化要素之间的关系。标准化要素作为整个装备-标准知识图谱中的纽带,在标准条目中作为具体的规范目标。因此,标准条目与标准化要素之间的关系为规范关系。

(5)标准化要素与组件/模块之间的关系。标准化要素在组件/模块中作为具体的约束条件,规范了组件或模块的技战术水平,因此标准化要素与组件/模块之间的关系为约束关系。

(6)组件/模块与装备之间的关系。装备由多个组件和模块组装而成,显然组件/模块与装备之间的关系为构成关系。

表 2 装备标准知识图谱中的实体关系

序号	实体关系	关系类型	简写
1	标准文档之间	相似性/引用	$R_{D,D}$
2	标准条目之间	相似性/引用	$R_{I,I}$
3	标准文档与标准条目之间	引用	$R_{D,I}$
4	标准条目与标准化要素之间	规范	$R_{I,T}$
5	标准化要素与组件/模块之间	约束	$R_{T,C}$
6	组件/模块与装备之间	构成	$R_{C,EQ}$

2.2 装备标准知识图谱过程模型整体框架

装备-标准知识图谱不同于其他领域的知识图谱,它不仅将标准中的定义、规范按照体系结构逐层分解并进行详细描述,以标准化要素为纽带用图形化的形式呈现标准、标准条目、装备模块之间的关联关系,为标准制修订、贯彻实施提供工作方法和手段,还将装备、模块、组件等与标准、标准条目进行关联,从而指导改进装备体系和研发新装备。根据 2.1 节中知识图谱中的实体和实体关系,装备-标准知识图谱过程模型的整体框架结构的设计如图 1 所示,共包括 9 个具有一定先后顺序的过程:

(1)通过对比标准文档之间的相似程度,来挖掘文档之间的相似性和引用关系,以确定标准文档之间的关联关系;

(2)将标准文档拆分为标准条目;

(3)通过构建标准文档与拆分后的标准条目之间的多对多引用关系,来确定标准文档与标准条目之间的关联关系;

(4)通过对比标准条目之间的相似程度,来挖掘条目之间的相似性,以确定标准条目之间的关联关系;

(5)确认标准条目中所规范的标准化要素;

(6)通过层次分析的方法,将装备划分为装备模块;

(7)通过确认装备与组件/模块之间的构成关系,来确定组件/模块与装备之间的关联关系;

(8)通过确认组件/模块战技指标中所对应的标准化要素,来确定标准化要素与标准化要素之间的关联关系;

(9)综合所有关联关系,完成装备-标准知识图谱的构建。

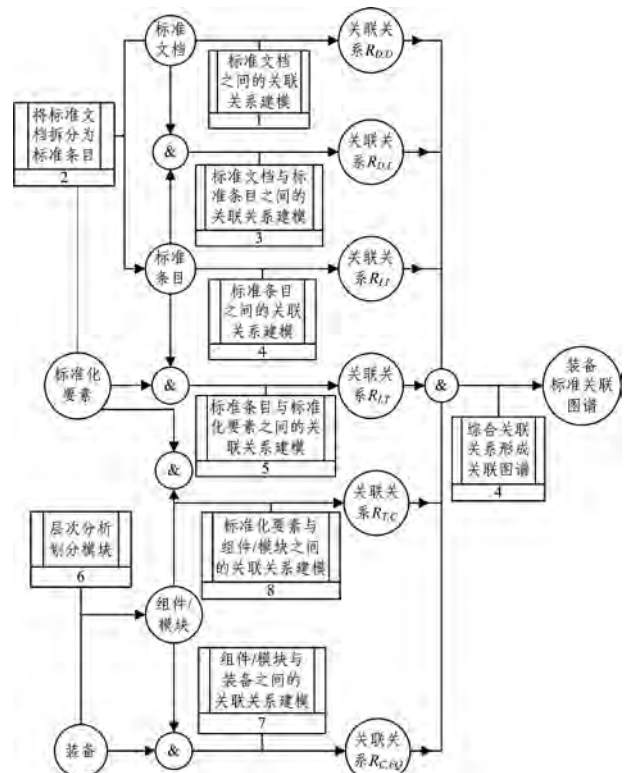


图 1 装备标准知识图谱过程模型的整体框架

3 装备-标准知识图谱子过程建模

3.1 标准文档之间的关联关系建模

现有装备的标准文档具有数量巨大、层次结构不明、交叉引用普遍存在的问题。为了更高效地完成标准文档之间的关系建模,给出了标准文档之间的关联关系建模过程,如图2所示。首先,根据标准文档的标题对标准文档归类,大幅度减少需要对比的文档数量;其次,对不同类别的文档进行逐一对比,根据标准文档内容的重复度、引用文字、发布时间等特征,从结构相似性和内容相似性两个角度找到每个文档的相似文档集合;然后,根据内容相似性挖掘文档中的直接引用关系,再联合文档之间的内容相似性和结构相似性判断文档之间的相似关系;最后,综合文档之间的引用关系和相似关系,构建标准文档之间的关联关系模型。

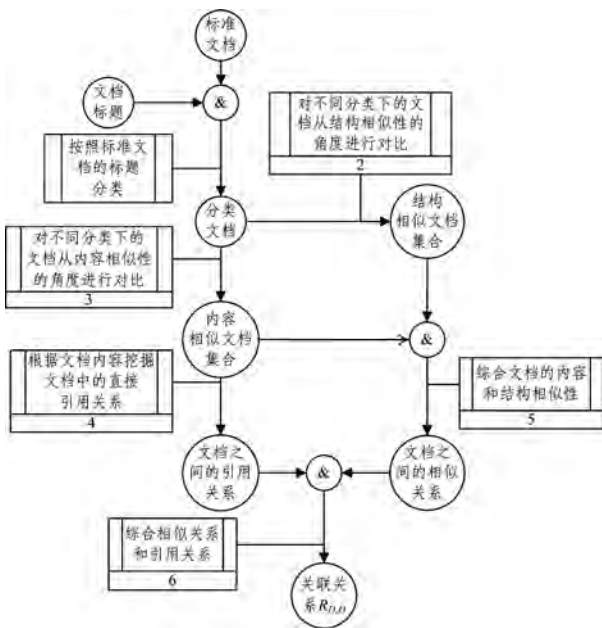


图2 标准文档之间的关联关系建模过程

3.2 标准文档与标准条目之间的关联关系建模

标准文档与标准条目之间是多对多的引用关系,因此将标准条目从标准文档中摘取后还不能直接建立标准文档与标准条目之间的关联关系,还需要比较、归并不同标准文档所提取的标准条目,确定引用了相同标准条目的文档。标准文档与标准条目之间的关联关系建模过程如图3所示。

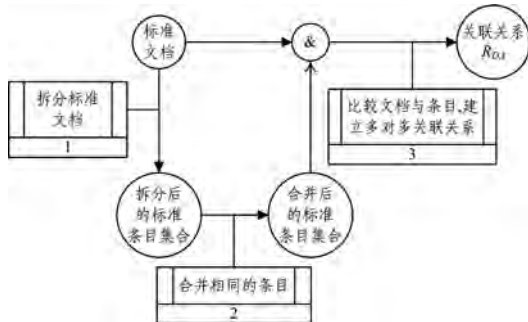


图3 标准文档与标准条目之间的关联关系建模过程

3.3 与标准化要素相关的关联关系建模

与标准化要素相关的关系建模主要包括两方面:1)建立标准条目到标准化要素之间的关联关系;2)建立标准化要素与组件/模块之间的关联关系。由此,以标准化要素为纽带,

将标准条目与组件或装备联系在一起,进而将标准文档、标准条目、标准化要素、组件/模块、装备之间的关联关系串联在一起。与标准化要素相关的关联关系建模过程如图4所示。

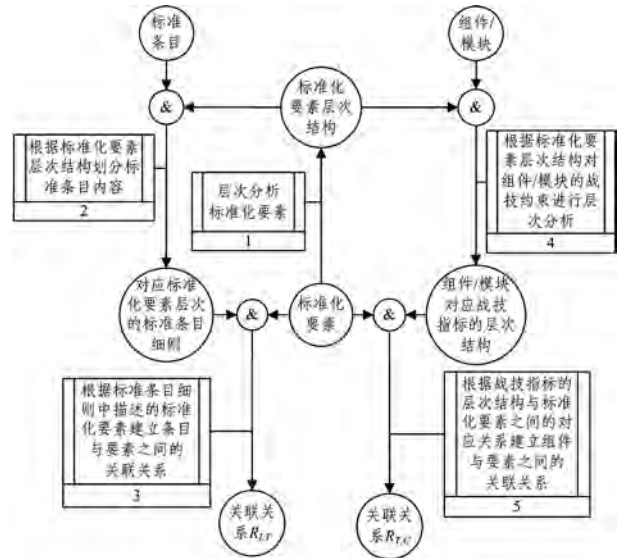


图4 与标准化要素相关的关联关系建模

为了建立与标准化要素相关的关联关系过程模型,首先需要厘清标准化要素的层次结构;然后,根据标准化要素的层次结构划分标准条目中的细则,并分析这些细则中描述的规范内容对应的标准化要素,建立标准条目与标准化要素之间的关联关系;最后,根据标准化要素的层次结构对组件/模块的战技指标进行层次分析,确定其对应的标准化要素,建立标准化要素与组件/模块之间的关联关系。

经过各子过程建模,用异质网络模型的方式对整个装备-标准知识图谱的全貌进行描述,图5给出了通过该过程模型构建的装备-标准知识图谱图例。从图中可以看出,该知识图谱能够清楚地描述标准文档或标准条目之间的引用关系、标准化要素对装备模块的约束关系、装备与装备模块之间的构成关系等。

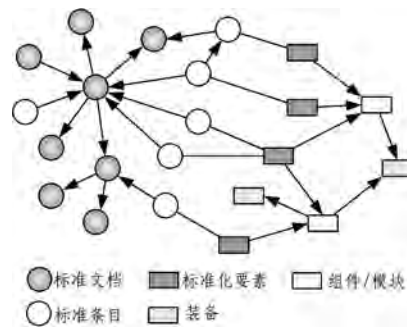


图5 装备标准知识图谱建模结果

结束语 在装备体系中,通过知识图谱能够清晰地反映装备、装备组件/模块、标准文档、标准条目以及标准化要素之间形成的复杂系统。本文利用过程建模方法对标准文档、标准条目、标准化要素、组件/模块、装备这几项实体之间的关联关系进行了细致的建模和分析。通过构建这些实体之间的关联关系,形成装备-标准知识图谱,利用此图谱能够从任一实体出发,快速检索和定位与此实体相关联的其他实体和隐含关系。但装备所对应的标准文档、标准条目的数量极其庞大,而本文的过程模型需要业务专家的大量人力劳动,效率低且

依赖于专家的个人知识水平。这使得通过人工的方式建立装备-标准知识图谱在短时间内无法完成。因此,通过知识图谱过程建模厘清各实体单元之间的逻辑联系,在此基础上,借助高性能计算机自动、准确地将标准文档、标准条目、标准化要素、组件/模块、装备等实体进行关联,从而构建装备-标准知识图谱的数学模型是装备标准化建设的关键,后续将在这方面开展深入的建模研究。

参考文献

- [1] LIN Y, LIU Z, SUN M, et al. Learning entity and relation embeddings for knowledge graph completion[C]// Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence. 2015:2181-2187.
- [2] JI G, LIU K, HE S, et al. Knowledge graph Completion with adaptive sparse transfer matrix[C]// Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. 2016:985-991.
- [3] YANG S, ZOU L, WANG Z, et al. Efficiently answering technical questions-A knowledge graph approach[C]// Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence. 2017:3111-3118.
- [4] HEIKO P. Knowledge graph refinement: A survey of approaches and evaluation methods[J]. *Semantic Web*, 2017, 8(3): 489-508.
- [5] NIKOLAEV F, KOTOV A, ZHILTSOV N. Parameterized fielded term dependence models for Ad-hoc entity retrieval from knowledge graph[C]// Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. Pisa, Italy, 2016:435-444.
- [6] BRACHMAN R J. What IS-A is and isn't: An analysis of taxonomic links in semantic networks [J]. *Computer*, 1983, 10: 30-36.
- [7] STEINER T, VERBORGH R, TRONCY R, et al. Adding real-time coverage to the Google knowledge graph[C]// Proceedings of the 2012th International Conference on Posters & Demonstrations Track-Volume 914. CEUR-WS. org, 2012:65-68.
- [8] WANG Z, WANG Z, LI J, et al. Knowledge extraction from Chinese wiki encyclopedias[J]. *Journal of Zhejiang University Science C*, 2012, 13(4): 268-280.
- [9] ZENG Y, WANG H, HAO H, et al. Statistical and structural analysis of web-based collaborative knowledge bases generated from Wiki Encyclopedia[C]// Proceedings of the 2012 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01. IEEE Computer Society, 2012:553-557.
- [10] HUANG Z, CHUANG W, ONG T H, et al. A graph-based recommender system for digital library[C]// Proceedings of the 2nd ACM/IEEE-CS Joint Conference on Digital Libraries. ACM, 2002:65-73.
- [11] SIEK J G, LEE L Q, LUMSDAINE A. The boost graph library: user guide and reference manual, portable documents[M]. Pearson Education, 2001.
- [12] DAI X, LI J, LIU T, et al. HRGRN: A graph search-empowered integrative database of Arabidopsis signaling transduction, metabolism and gene regulation networks[J]. *Plant and Cell Physiology*, 2016, 57(1): e12.
- [13] 刘峤, 李杨, 杨段宏, 等. 知识图谱构建技术综述[J]. *计算机研究与发展*, 2016, 53(3): 582-600.
- [14] 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.
- [15] 耿霞, 张继军, 李蔚妍. 知识图谱构建技术综述[J]. *计算机科学*, 2014, 41(7): 148-152.
- [16] 刘昶, 史海波, 于海斌. 基于多 Agent 制造过程的建模方法[J]. *控制工程*, 2005, 12(6): 515-519.
- [17] 高飞, 高阜乡, 王钰, 等. 基于实体的指挥控制过程仿真建模[J]. *指挥控制与仿真*, 2012, 34(3): 116-120.
- [18] 张树玲, 孙波, 田艳琴. 过程建模技术研究综述[C]// 2010 Third International Conference on Education Technology and Training. Wuhan, Hubei, 2010:542-547.
- [19] PLAIA A, CARRIE A. Application and assessment of IDEF3 - process flow description capture method[J]. *International Journal of Operations & Production Management*, 1995, 15(1): 63-73.
- [20] 肖磊, 金光, 周忠宝, 等. 基于 Petri 网的维修过程建模仿真方法研究[J]. *军械工程学院学报*, 2008, 20(5): 6-11.
- [21] SCHEER A W. Architecture of integrated information system-foundations of enterprise modelling[M]. Berlin: Springer-Verlag, 1992.
- (上接第 475 页)
- [10] LIN Y T, AGRAWAL D, CHEN C, et al. Llama: Leveraging Columnar Storage for Scalable Join Processing in the MapReduce Framework[C]// The ACM SIGMOD International Conference on Management of Data. Athens, Greece: ACM, 2011:961-972.
- [11] FLORATOU A, PATEL J M, SHEKITA E J, et al. Column-Oriented Storage Techniques for MapReduce[J]. *PVLDB*, 2011, 4(7): 419-429.
- [12] THUSOO A, SARMA J S, JAIN N, et al. Raghatham Murthy: Hive-A Warehousing Solution Over a Map-Reduce Framework. [C]// VLDB Conference. Lyon, France, VLDB Endowment, 2009:1626-1629.
- [13] HE Y Q, LEE R B, HUAI Y, et al. RCFile: A Fast and Space-efficient Data Placement Structure in MapReduce-based Warehouse Systems[C]// IEEE International Conference on Data Engineering. Hannover, Germany, 2011:1199-1208.
- [14] HSIAO H, CHEN M S, YU P S. Parallel execution of hash joins in parallel databases[J]. *IEEE Trans. on Parallel and Distributed Systems*, 1997, 8(8): 872-883.
- [15] BONCZ P, MANEGOLD S, KERSTEN M L. Database architecture optimized for the new bottleneck: memory access[C]// The 25th Int'l Conf. on Very Large Data Bases. ACM Press, 1999: 231-246.
- [16] O'NEIL P, O'NEIL B, CHEN X D. Star Schema Benchmark Revision[EB/OL]. [2010-2-9]. <http://www.cs.umb.edu/~poneil>.