

# 知识图谱

中文名：知识图谱

外文名：Knowledge Graph

学科：知识工程、大数据、自然语言处理、WEB

实质：图数据库与图计算推理

**由来** 在知识工程领域，知识库的构建与管理一直是研究热点和难点。早期的知识工程主要通过“自顶向下”的方法构建知识库，即首先构建知识本体（ontology）然后再添加知识实例（instance）。随着大数据时代的到来，知识工程再次受到学术界和产业界的广泛关注，并成为人工智能在知识工程领域发展的最新前沿<sup>[1]</sup>。

**定义** 知识图谱是通过符号化的形式描述物理世界中的概念及其相互关系的结构化语义数据库。其基本组成单位是实体、实体间的关系以及实体的属性；实体通过关系相互连接，进而组成网状的知识结构。在信息爆炸式增长的时代，知识图谱通过对数据的整合以及规范，向人们提供有价值的结构化信息<sup>[2]</sup>。

**发展** 区别于传统“自顶向下”的方式，自动化知识图谱构建通常采用“自底向上”的方式，借助自然语言处理等技术手段，从文本和图像等非结构化（或半结构化）数据以及存量结构化数据中获取结构化知识，经过审核后加入到知识图谱中。目前，已形成了诸如 YAGO<sup>1</sup>、DBpedia<sup>2</sup>、NELL<sup>[3]</sup> 等具有代表性的大规模通用知识图谱；此外，在医疗、金融、司法、政务、智能制造等行业领域，也涌现出了一批领域知识图谱<sup>[4,5]</sup>。

<sup>1</sup> <https://yago-knowledge.org/>

<sup>2</sup> <https://www.dbpedia.org/>

<sup>3</sup> <https://www.xlore.cn/>

<sup>4</sup> <http://openkg.cn/dataset/zhishi-me>

<sup>5</sup> <http://kw.fudan.edu.cn/>

<sup>6</sup> <http://www.openkg.cn/>

我国在中文知识图谱方面的研究积累了大量成果，如 HowNet<sup>[6]</sup>、大规模中英文跨语言知识图谱 XLore<sup>3</sup>、基于 OpenKN<sup>[7]</sup> 的“人立方、事立方、知立方”系统、中文知识图谱平台 zhishi.me<sup>4</sup>、CN-DBpedia<sup>5</sup> 系列中文知识图谱项目，以及百度知心和搜狗知立方，等等。

近年来，针对不同现实场景，一系列新型知识图谱被提出，如事件知识图谱<sup>[8]</sup>、事理知识图谱<sup>[9]</sup>、时态知识图谱<sup>[10]</sup> 和多模态知识图谱<sup>[11]</sup> 等。此外，还有研究工作指出，大规模预训练模型（pre-trained models）或将成为下一代知识图谱的组成形态之一<sup>[12,13]</sup>。

**工具及应用** 为了让初学者更好地理解知识图谱，文献[14~17]全面系统地总结了知识图谱的基本概念、构建技术、精化方法，及其实际应用。文献[18]介绍了知识图谱表示相关技术，文献[19]系统地总结了知识图谱融合对齐相关研究进展，而文献[20]综述了新一代知识图谱的关键技术的研究发展现状。OpenKG<sup>6</sup> 作为一个开放平台，集中提供了以中文为基础的知识图谱数据、算法和工具。

基于知识图谱，最基本的利用就是知识检索，同时还支持一系列融合知识的高级信息服务，如自动问答和信息推荐等；此外，知识图谱也被用于知识驱动的自然语言处理相关任务<sup>[21]</sup>，与嵌入（embedding）、图神经网络（graph neural networks）等深度学习技术紧密结合，互相渗透。

**挑战与展望** 虽针对广泛的现实需求，知识图谱赋能的知识工程尚未形成完整成熟的理论体系、技术方法与系统实践，仍有众多有待解决的挑战性前沿问题。例如，高效构建和管理事件、事理、多模态这些新类型知识图谱，需要后续更加深入的研究。此外，如何在知识图谱的表示和推理中实现“神经+符号”的结合<sup>[22]</sup>，值得我们深入思考。 ■

## 王 鑫

CCF 杰出会员、CCF 信息系统专委会秘书长、CCF 数据库专委会执行委员。天津大学教授。主要研究方向为知识图谱、图数据库。wangx@tju.edu.cn

## 赵 翔

CCF 高级会员，CCF 信息系统专委会、数据库专委会、大数据专家委执行委员。国防科技大学教授。主要研究方向为知识图谱、大图数据分析。xiangzhao@nudt.edu.cn

## 王昊奋

CCF 术语工委副主任，CCF TF-SIGKG 主席，CCF 上海秘书长，CCF 信息系统专委会、自然语言处理专委会执行委员。同济大学特聘研究员。主要研究方向为知识图谱、知识问答。carter.whfcarter@gmail.com

(本文由 CCF 术语工委执行委员林俊宇推荐)

## 参考文献

- [1] 漆桂林, 高桓, 吴天星. 知识图谱研究进展. 情报工程, 2017(01): 4-25.
- [2] 王鑫, 邹磊, 王朝坤, 彭鹏, 冯志勇. 知识图谱数据管理研究综述 [J]. 软件学报, 2019, 30(7): 2139-2174.
- [3] Carlson A, Betteridge J, Kisiel B, et al. Toward an architecture for never-ending language learning[C]// *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, 2010:1306-1313.
- [4] 中国中文信息学会语言与知识计算专业委员会. 知识图谱发展报告 [OL]. 2018. <http://cips-upload.bj.bcebos.com/KGDevReport2018.pdf>.
- [5] 中国电子技术标准化研究院. 知识图谱标准化白皮书 [OL]. 2019. <http://www.cesi.cn/201909/5589.html>.
- [6] 董振东, 董强, 郝长伶. 知网的理论发现 [J]. 中文信息学报, 2007(4): 3-9.
- [7] Jia Y, Wang Y, Cheng X, et al. OpenKN: An open knowledge computational engine for network big data[C]// *IEEE/ACM International Conference on Advances in Social Networks Analysis & Mining*. ACM, 2014:657-664.
- [8] Guan S, Cheng X, Bai L, et al. What is Event Knowledge Graph: A Survey[J]. *IEEE Trans. Knowl. Data Eng.*, in press, 2022.
- [9] Li Z, Ding X, Liu T. Constructing Narrative Event Evolutionary Graph for Script Event Prediction[C]// *Twenty-Seventh International Joint Conference on Artificial Intelligence*. 2018:4201-4207.
- [10] Zhu C, Chen M, Fan C, et al. Learning from History: Modeling Temporal Knowledge Graphs with Sequential Copy-Generation Networks[J]. *Proceedings of AAAI*, 2021:4732-4740.
- [11] Zhu X, Li Z, Wang X, et al. Multi-Modal Knowledge Graph Construction and Application: A Survey[J]. *arXiv preprint*, 2202.05786, 2022.
- [12] Thorne J, Yazdani M, Saeidi M, et al. From Natural Language Processing to Neural Databases[C]// *Proc. VLDB Endow*, 2021, 14(6): 1033-1039.
- [13] Petroni F, Rocktschel T, Lewis P, et al. Language Models as Knowledge Bases? [C]// *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2463-2473, 2019.
- [14] 王昊奋, 漆桂林, 陈华钧. 知识图谱: 方法, 实践与应用 [M]. 电子工业出版社, 2019.
- [15] 刘知远, 韩旭, 孙茂松. 知识图谱与深度学习 [M]. 清华大学出版社, 2020 年第 1 版.
- [16] Hogan A, Blomqvist E, Cochez M, et al. Knowledge Graphs. *Synthesis Lectures on Data, Semantics, and Knowledge*. Morgan & Claypool Publishers, 2021:1-257.
- [17] Weikum G, Dong X L, Razniewski S, et al. Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases[J]. *Found. Trends Databases*, 2021, 10(2-4): 108-490.
- [18] 刘知远, 孙茂松, 林衍凯, 等. 知识表示学习研究进展 [M]. 计算机研究与发展, 2016, 53(2):247-261.
- [19] Zhao X, Zeng W, Tang J, et al. An Experimental Study of State-of-the-Art Entity Alignment Approaches[J]. *IEEE Trans. Knowl. Data Eng.*, 2022, 34(6): 2610-2625.
- [20] 王萌, 王昊奋, 李博涵, 赵翔, 王鑫. 新一代知识图谱关键技术研究进展综述 [OL]. 计算机研究与发展, 2022. DOI:10.7544/issn1000-1239.20210829.
- [21] Wang H, Ma F, Wang Y, et al. Knowledge-Guided Paraphrase Identification. [C]// *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP) Findings*, 2021:843-853.
- [22] 王昊奋, 王萌. “神经+符号”: 从知识图谱角度看认知推理的发展 [J]. 中国计算机学会通讯, 2020,(8):52-55.