揭秘知识图谱全生命周期技术
探索垂直领域知识图谱构建方法与应用落地
促进人工智能从感知时代向认知时代跨越

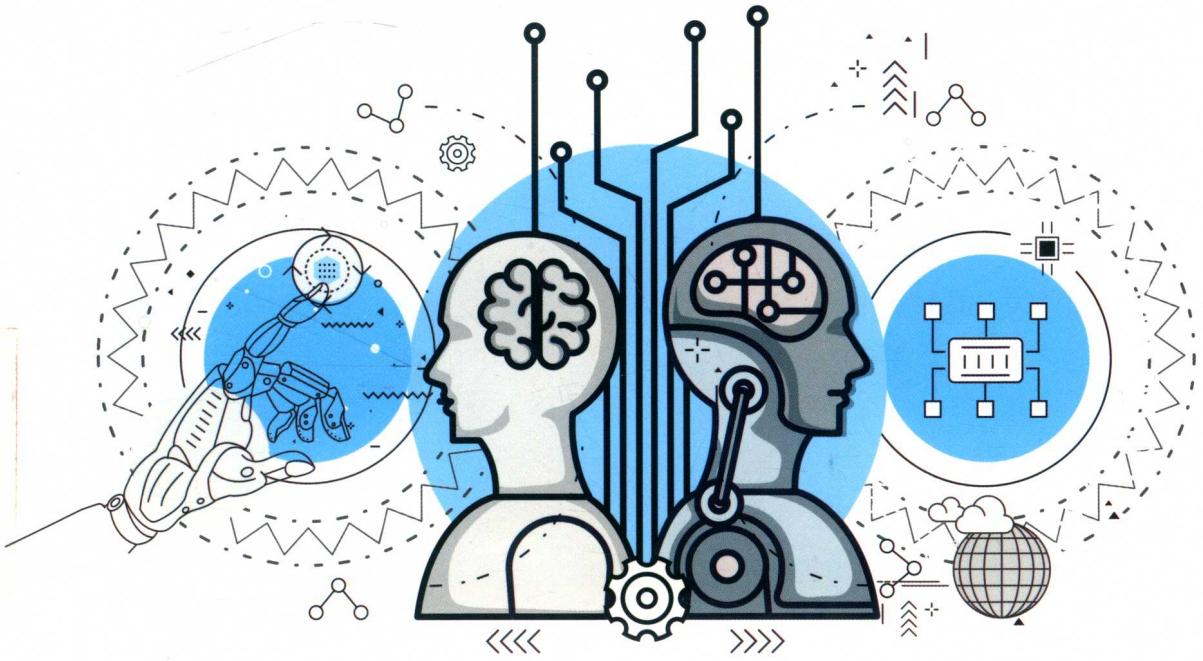
Broadview®
www.broadview.com.cn

知识图谱

方法、实践与应用

KNOWLEDGE GRAPH

王昊奋　漆桂林　陈华钧◎主编



中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

知识图谱

方法、实践与应用

KNOWLEDGE GRAPH

[主编] 王昊奋　漆桂林　陈华钧

[参编] Jeff Z. Pan　丁军　丁力　汪鹏　王萌　王鑫
王宇　王志春　肖国辉　杨成彪　张伟

電子工業出版社
Publishing House of Electronics Industry
北京•BEIJING

内 容 简 介

知识图谱是较为典型的多学科交叉领域，涉及知识工程、自然语言处理、机器学习、图数据库等多个领域。本书系统地介绍知识图谱涉及的关键技术，如知识建模、关系抽取、图存储、自动推理、图谱表示学习、语义搜索、知识问答、图挖掘分析等。此外，本书还尝试将学术前沿和实战结合，让读者在掌握实际应用能力的同时对前沿技术发展有所了解。

本书既适合计算机和人工智能相关的研究人员阅读，又适合在企业一线从事技术和应用开发的人员学习，还可作为高等院校计算机或人工智能专业师生的参考教材。

图书在版编目（CIP）数据

知识图谱：方法、实践与应用/王昊奋，漆桂林，陈华钧主编. —北京：电子工业出版社，2019.8
ISBN 978-7-121-36671-0

Ⅰ. ①知… Ⅱ. ①王… ②漆… ③陈… Ⅲ. ①知识管理 Ⅳ. ①G302

中国版本图书馆 CIP 数据核字（2019）第 100477 号

责任编辑：宋亚东
印　　刷：三河市良远印务有限公司
装　　订：三河市良远印务有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　邮编 100036
开　　本：787×980　1/16　印张：30　字数：546 千字
版　　次：2019 年 8 月第 1 版
印　　次：2019 年 8 月第 1 次印刷
定　　价：118.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：010-51260888-819，faq@phei.com.cn。

序

知识图谱是人工智能的一个分支，对可解释人工智能具有重要作用。近几年，随着知识表示和机器学习等技术的发展，知识图谱相关技术取得了突破性的进展，特别是知识图谱的构建、推理和计算技术以及知识服务技术，都得到了快速的发展。这些技术的进步使知识图谱在工业界受到了广泛关注，并取得了显著成果。谷歌、微软、百度等互联网公司率先构建了大规模通用知识图谱，提供基于实体和关系的语义搜索，可以更好地理解用户查询。知识图谱还在智能决策系统、推荐系统和智能问答系统中起到了重要作用。知识图谱不仅有巨大的应用价值，而且具有重要的理论价值。知识图谱使传统知识表示和推理技术有了落脚点，也为知识表示和推理带来了新的挑战。

本书系统介绍了知识图谱的理论、技术及应用。在理论方面，本书全面介绍了知识图谱的各种表示方法，以及知识图谱的推理方法，这些方法是知识图谱的根基。在技术方面，本书全面介绍了知识图谱的存储和查询技术、挖掘构建、知识融合技术，以及基于知识图谱的语义搜索和智能问答技术。在应用方面，本书全面地介绍了知识图谱在工业界的典型应用场景，为知识图谱的发展提供了养分。目前，关于知识图谱的专业书籍还比较缺乏，本书将给广大知识图谱研究人员和应用人员带来福音。

本书作者们都是在知识图谱的研究和产业应用方面有丰富经验的专家和学者，很好地融合了知识图谱的学术研究和产业化实践，相信本书的出版对于知识图谱技术的普及和发展会产生非常积极的作用。

清华大学教授

前　言

知识图谱的早期理念源于万维网之父 Tim Berners-Lee 关于语义网（The Semantic Web）的设想，旨在采用图结构（Graph Structure）来建模和记录世界万物之间的关联关系和知识，以便有效实现更加精准的对象级搜索。知识图谱的相关技术已经在搜索引擎、智能问答、语言理解、推荐计算、大数据决策分析等众多领域得到广泛的实际应用。近年来，随着自然语言处理、深度学习、图数据处理等众多领域的飞速发展，知识图谱在自动化知识获取、知识表示学习与推理、大规模图挖掘与分析等领域又取得了很多新进展。知识图谱已经成为实现认知层面的人工智能不可或缺的重要技术之一。

为什么写作本书

知识图谱是较为典型的交叉领域，涉及知识工程、自然语言处理、机器学习、图数据库等多个领域。而知识图谱的构建及应用涉及更多细分领域的一系列关键技术，包括：知识建模、关系抽取、图存储、自动推理、图谱表示学习、语义搜索、智能问答、图计算分析等。做好知识图谱需要系统掌握和应用这些分属多个领域的技术。

本书写作的第一个目的是尽可能地梳理和组织好这些知识点，帮助读者系统掌握相关技术，能够从整体、全局和系统的视角看待和应用知识图谱技术。早期的知识图谱应用主要是谷歌、百度等公司的通用域搜索引擎，以及基于搜索延续发展出来的基于知识图谱的智能问答应用，如天猫精灵、小米小爱等。这类应用主要依靠通用领域的知识图谱，如百科类知识图谱。近年来，知识图谱在医疗、金融、安全等垂直领域深入发展，知识图谱的应用也进一步从通用领域向越来越多的垂直领域扩展。对于刚刚进入该领域的从业人员，更需要能从应用入手，开展知识图谱的研究与开发。

本书写作的第二个目的是希望能够为这些知识图谱应用开发人员提供一本参考型的工具书。因此，本书在章节最后安排了一个小节介绍相关技术点的常用开源工具，并在与本书配套的网站上提供了完整的实际操作教程。

近几年，随着人工智能的进一步发展，知识图谱在深度知识抽取、表示学习与机器推理、基于知识的可解释性人工智能、图谱挖掘与图神经网络等领域取得了一系列新的进展。本书写作的第三个目的是希望梳理和整理这些与知识图谱相关领域的最新进展，帮助读者了解它们的技术发展前沿。

关于本书作者

本书邀请了国内从事相关领域研究和开发的一线专家。三位主编都在语义网和知识图谱领域有着十余年的研究和开发经验，同时也是中文领域开放知识图谱 OpenKG 的发起人。每个章节由各细分技术领域的专家主持撰写，参与编写的编者既有来自国内高校从事相关学术研究的教师，也有来自企业拥有丰富实际开发经验的技术专家。

本书主要内容

本书共包括 9 章，主要内容如下：

第 1 章主要介绍知识图谱的基本概念、历史渊源、典型的知识图谱项目、技术要素以及核心应用价值。

第 2 章围绕知识表示与建模，首先介绍传统人工智能领域的典型知识表示方法，如谓词逻辑、描述逻辑、框架系统等，接下来重点介绍 RDF、OWL 等互联网时代的知识表示框架，此外还介绍知识图谱的向量表示方法等。最后以 Protégé 为例介绍知识建模的具体实践过程。

第 3 章围绕知识存储，首先介绍知识图谱存储的主要特点和难点，然后介绍几种常用的知识图谱存储索引及存储技术，并对原生图数据库的技术原理进行简要介绍。此外，还概要介绍常用的图数据库，并以 Apache Jena 和 gStore 为例介绍知识图谱存储的具体实践过程。

第 4 章围绕知识抽取与知识挖掘，首先介绍从不同来源获取知识图谱数据的常用方法，然后重点围绕实体抽取、关系抽取和事件抽取等，对从文本中获取知识图谱数据的方法展开了较为具体的介绍。最后以 DeepDive 开源工具为例介绍关系抽取的具体实践过程。

第 5 章围绕知识图谱的融合，分别对概念层的融合和实体层的融合展开介绍，包括本

体映射、语义映射技术、实体对齐、实体链接等。最后以 LIMES 开源工具为例介绍实体融合的具体实践过程。

第 6 章围绕知识图谱推理，首先介绍推理的基本概念，然后分别从基于演绎逻辑的知识图谱推理和基于归纳的知识图谱推理，对常用的知识图谱推理技术进行介绍。最后以 Apache Jena 和 Drools 等开源工具为例介绍知识图谱推理的具体实践过程。

第 7 章和第 8 章分别围绕语义搜索和知识问答展开，介绍语义索引、基于知识图谱的问答等系列技术，并以 gAnswer 等开源工具为例，介绍基于知识图谱实现精准搜索和问答的具体实践过程。

第 9 章为应用案例章节，作者挑选了电商、图情、生活娱乐、企业商业、创投、中医临床领域和金融证券行业 7 个应用案例，对知识图谱技术在不同领域的实现过程和应用方法展开介绍。

如何阅读本书

这是一本大厚书，读者应该怎样利用这本书呢？

在阅读此书前，读者应当学过数据库、机器学习及自然语言处理的基本知识。这本书的章节是依据知识图谱的相关技术点进行安排的。由于知识图谱涉及的技术面较多，我们建议刚进入知识图谱领域的读者分几遍阅读本书。

- 第一遍先通读全书，主要厘清基本概念，对涉及学术前沿的内容以及开源工具实践部分的内容可以只简单浏览。
- 第二遍重点针对每个章节后面的开源工具进行实践学习，通过上手操作加深对各技术点的理解。
- 第三遍针对各章中介绍的算法进行学习，并结合相关论文的阅读加深对算法的理解。在这个阶段可以挑选自己感兴趣的技术点进行深入研究。

在撰写本书时，编者考虑了各章节技术点的独立性，对知识图谱的某些技术已经有些了解的读者，可以不用严格按照书的章节顺序阅读，而是挑选自己感兴趣的章节进行学习。

致谢

本书是很多人共同努力的成果，在此感谢各位编者的共同努力。同时，在本书写作过程中，北京大学的邹磊，湖南大学的彭鹏，海知智能的袁熙昊、韩庐山、王燚鹏、孙胜男、郭玉婷，东南大学的吴桐桐、谭亦鸣、花云程、胡森，浙江大学的张文、王冠颖、王若旭、陈名杨、王梁、叶志权等人也提供了非常有价值的调研结果和修改意见，在此表示衷心的感谢。

在电子工业出版社博文视点宋亚东编辑的热情推动下，最终促成了我们与电子工业出版社的合作。在审稿过程中，他多次邀请专家对此书提出有益意见，对书稿的修改完善起到了重要作用。在此感谢电子工业出版社博文视点和宋亚东编辑对本书的重视，以及为本书出版所做的一切。

为推动中文领域开放知识图谱的发展，本书的作者们一致同意将部分稿酬捐赠给OpenKG。在此，也对参与本书的所有作者的无私奉献表示感谢。

由于作者水平有限，书中不足及错误之处在所难免。此外，由于知识图谱技术涉及面广，本书难免有所遗漏，敬请专家和读者给予批评指正。

作者

2019 年 7 月

读者服务

轻松注册成为博文视点社区用户（www.broadview.com.cn），扫码直达本书页面。

- 提交勘误：您对书中内容的修改意见可在“提交勘误”处提交，若被采纳，将获赠博文视点社区积分（在您购买电子书时，积分可用来抵扣相应金额）。
- 交流互动：在页面下方“读者评论”处留下您的疑问或观点，与我们和其他读者一同学习交流。

目　录

第1章 知识图谱概述

陈华钧　浙江大学，漆桂林　东南大学，王昊奋　乐言科技，王鑫　天津大学

1.1 什么是知识图谱

知识图谱是一种用图模型来描述知识和建模世界万物之间的关联关系的技术方法[1]。知识图谱由节点和边组成。节点可以是实体，如一个人、一本书等，或是抽象的概念，如人工智能、知识图谱等。边可以是实体的属性，如姓名、书名，或是实体之间的关系，如朋友、配偶。知识图谱的早期理念来自 Semantic Web[2,3]（语义网），其最初理想是把基于文本链接的万维网转化成基于实体链接的语义网。

1989 年，Tim Berners-Lee 提出构建一个全球化的以“链接”为中心的信息系统（Linked Information System）。任何人都可以通过添加链接把自己的文档链入其中。他认为，相比基于树的层次化组织方式，以链接为中心和基于图的组织方式更加适合互联网这种开放的系统。这一思想逐步被人们实现，并演化发展成为今天的 World Wide Web。

1994 年，Tim Berners-Lee 又提出 Web 不应该仅仅只是网页之间的互相链接。实际上，网页中描述的都是现实世界中的实体和人脑中的概念。网页之间的链接实际包含语义，即这些实体或概念之间的关系；然而，机器却无法有效地从网页中识别出其中蕴含的语义。他于 1998 年提出了 Semantic Web 的概念[4]。Semantic Web 仍然基于图和链接的组织方式，只是图中的节点代表的不只是网页，而是客观世界中的实体（如人、机构、地点等），而超链接也被增加了语义描述，具体标明实体之间的关系（如出生地是、创办人是等）。相对于传统的网页互联网，Semantic Web 的本质是数据的互联网（Web of Data）或

事物的互联网（Web of Things）。

在 Semantic Web 被提出之后，出现了一大批新兴的语义知识库。如作为谷歌知识图谱后端的 Freebase[5]，作为 IBM Waston 后端的 DBpedia[6]和 Yago[7]，作为 Amazon Alexa 后端的 True Knowledge，作为苹果 Siri 后端的 Wolfram Alpha，以及开放的 Semantic Web Schema——Schema.ORG[8]，目标成为世界最大开放知识库的 Wikidata[9]等。尤其值得一提的是，2010 年谷歌收购了早期语义网公司 MetaWeb，并以其开发的 Freebase 作为数据基础之一，于 2012 年正式推出了称为知识图谱的搜索引擎服务。随后，知识图谱逐步在语义搜索[10,11]、智能问答[12-14]、辅助语言理解[15,16]、辅助大数据分析[17-19]、增强机器学习的可解释性[20]、结合图卷积辅助图像分类[21,22]等多个领域发挥出越来越重要的作用。

如图 1-1 所示，知识图谱旨在从数据中识别、发现和推断事物与概念之间的复杂关系，是事物关系的可计算模型。知识图谱的构建涉及知识建模、关系抽取、图存储、关系推理、实体融合等多方面的技术，而知识图谱的应用则涉及语义搜索、智能问答、语言理解、决策分析等多个领域。构建并利用好知识图谱需要系统性地利用包括知识表示（Knowledge Representation）、图数据库、自然语言处理、机器学习等多方面的技术。

图 1-1 知识图谱：事物关系的可计算模型

1.2 知识图谱的发展历史

知识图谱并非突然出现的新技术，而是历史上很多相关技术相互影响和继承发展的结