Knowledge Graphs

Synthesis Lectures on Data, Semantics, and Knowledge

Synthesis Lectures on Data, Semantics, and Knowledge is edited by Ying Ding of the University of Texas at Austin and Paul Groth of the University of Amsterdam. The series focuses on the pivotal role that data on the web and the emergent technologies that surround it play both in the evolution of the World Wide Web as well as applications in domains requiring data integration and semantic analysis. The large-scale availability of both structured and unstructured data on the Web has enabled radically new technologies to develop. It has impacted developments in a variety of areas including machine learning, deep learning, semantic search, and natural language processing. Knowledge and semantics are a critical foundation for the sharing, utilization, and organization of this data. The series aims both to provide pathways into the field of research and an understanding of the principles underlying these technologies for an audience of scientists, engineers, and practitioners.

Topics to be included:

- Knowledge graphs, both public and private
- Linked Data
- Knowledge graph and automated knowledge base construction
- Knowledge engineering for large-scale data
- Machine reading
- Uses of Semantic Web technologies
- Information and knowledge integration, data fusion
- Various forms of semantics on the web (e.g., ontologies, language models, and distributional semantics)
- Terminology, Thesaurus, & Ontology Management
- Query languages

Knowledge Graphs
Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann
2021

Web Data APIs for Knowledge Graphs: Easing Access to Semantic Data for Application Developers
Alberto Meroño-Peñuela, Pasquale Lisena, and Carlos Martínez-Ortiz
2021

Designing and Building Enterprise Knowledge Graphs
Juan Sequeda and Ora Lassila
2021

Linked Data Visualization: Techniques, Tools, and Big Data
Laura Po, Nikos Bikakis, Federico Desimoni, and George Papastefanatos
2020

Ontology Engineering
Elisa F. Kendall and Deborah L. McGuinness
2019

Demystifying OWL for the Enterprise
Michael Uschold
2018

Validating RDF Data
José Emilio Labra Gayo, Eric Prud'hommeaux, Iovka Boneva, and Dimitris Kontokostas
2017

Natural Language Processing for the Semantic Web
Diana Maynard, Kalina Bontcheva, and Isabelle Augenstein
2016

The Epistemology of Intelligent Semantic Web Systems
Mathieu d'Aquin and Enrico Motta
2016

Entity Resolution in the Web of Data
Vassilis Christophides, Vasilis Efthymiou, and Kostas Stefanidis
2015

Library Linked Data in the Cloud: OCLC's Experiments with New Models of Resource Description
Carol Jean Godby, Shenghui Wang, and Jeffrey K. Mixter
2015

Semantic Mining of Social Networks
Jie Tang and Juanzi Li
2015

Social Semantic Web Mining
Tope Omitola, Sebastián A. Ríos, and John G. Breslin
2015

Semantic Breakthrough in Drug Discovery
Bin Chen, Huijun Wang, Ying Ding, and David Wild
2014

Semantics in Mobile Sensing
Zhixian Yan and Dipanjan Chakraborty
2014

Provenance: An Introduction to PROV
Luc Moreau and Paul Groth
2013

Resource-Oriented Architecture Patterns for Webs of Data
Brian Sletten
2013

Aaron Swartz's A Programmable Web: An Unfinished Work
Aaron Swartz
2013

Incentive-Centric Semantic Web Application Engineering
Elena Simperl, Roberta Cuel, and Martin Stein
2013

Publishing and Using Cultural Heritage Linked Data on the Semantic Web
Eero Hyvönen
2012

VIVO: A Semantic Approach to Scholarly Networking and Discovery
Katy Börner, Michael Conlon, Jon Corson-Rikert, and Ying Ding
2012

Linked Data: Evolving the Web into a Global Data Space
Tom Heath and Christian Bizer
2011

Knowledge Graphs

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann

ISBN: 978-3-031-00790-3 paperback
ISBN: 978-3-031-01918-0 PDF
ISBN: 978-3-031-00113-0 hardcover

DOI 10.1007/978-3-031-01918-0

A Publication in the Springer series
SYNTHESIS LECTURES ON DATA, SEMANTICS, AND KNOWLEDGE

Lecture #22
Series Editors: Ying Ding, *University of Texas at Austin*
Paul Groth, *University of Amsterdam*
Founding Editor Emeritus: James Hendler, *Rensselaer Polytechnic Institute*
Series ISSN
Print 2691-2023 Electronic 2691-2031

Knowledge Graphs

Aidan Hogan
DCC, Universidad de Chile; IMFD

Michael Cochez
Vrije Universiteit Amsterdam and Discovery Lab, Elsevier

Gerard de Melo
HPI, University of Potsdam and Rutgers University

Sabrina Kirrane
WU Vienna

Roberto Navigli
Sapienza University of Rome

Axel-Cyrille Ngonga Ngomo
DICE, Universität Paderborn

Sabbir M. Rashid
Tetherless World Constellation, Rensselaer Polytechnic Institute

Lukas Schmelzeisen
Universität Stuttgart

Steffen Staab
Universität Stuttgart and University of Southampton

Eva Blomqvist
Linköping University

Claudia d'Amato
University of Bari

Claudio Gutierrez
DCC, Universidad de Chile; IMFD

José Emilio Labra Gayo
Universidad de Oviedo

Sebastian Neumaier
St. Pölten University of Applied Sciences

Axel Polleres
WU Vienna

Anisa Rula
University of Brescia

Juan Sequeda
data.world

Antoine Zimmermann
École des mines de Saint-Étienne

SYNTHESIS LECTURES ON DATA, SEMANTICS, AND KNOWLEDGE #22

ABSTRACT

This book provides a comprehensive and accessible introduction to knowledge graphs, which have recently garnered notable attention from both industry and academia. Knowledge graphs are founded on the principle of applying a graph-based abstraction to data, and are now broadly deployed in scenarios that require integrating and extracting value from multiple, diverse sources of data at large scale.

The book defines knowledge graphs and provides a high-level overview of how they are used. It presents and contrasts popular graph models that are commonly used to represent data as graphs, and the languages by which they can be queried before describing how the resulting data graph can be enhanced with notions of schema, identity, and context. The book discusses how ontologies and rules can be used to encode knowledge as well as how inductive techniques—based on statistics, graph analytics, machine learning, etc.—can be used to encode and extract knowledge. It covers techniques for the creation, enrichment, assessment, and refinement of knowledge graphs and surveys recent open and enterprise knowledge graphs and the industries or applications within which they have been most widely adopted. The book closes by discussing the current limitations and future directions along which knowledge graphs are likely to evolve.

This book is aimed at students, researchers, and practitioners who wish to learn more about knowledge graphs and how they facilitate extracting value from diverse data at large scale. To make the book accessible for newcomers, running examples and graphical notation are used throughout. Formal definitions and extensive references are also provided for those who opt to delve more deeply into specific topics.

KEYWORDS

knowledge graphs, graph databases, knowledge graph embeddings, graph neural networks, ontologies, knowledge graph refinement, knowledge graph quality, knowledge bases, artificial intelligence, semantic web, machine learning

Contents

Preface

The origins of this book can be traced back to a Dagstuhl Seminar, held in 2018, on the topic of Knowledge Graphs. At the time of the seminar, the topic was quickly becoming mainstream in academia and industry, but there were conflicting messages as to what a "knowledge graph" was. Much of the discussion of the seminar centered on this question, and there were divergent opinions as to how knowledge graphs could (or should) be defined; how they relate to previous concepts such as graph databases, knowledge bases, ontologies, RDF graphs, property graphs, semantic networks, etc.; and how the emerging area of Knowledge Graphs should be positioned with respect to the established areas of Artificial Intelligence, Big Data, Databases, Graph Theory, Logic, Machine Learning, Knowledge Representation, Natural Language Processing, Networks (in their various forms), and the Semantic Web. As the discussion continued, a consensus began to emerge: Knowledge Graphs, as a topic, involves a novel confluence of techniques stemming from previously disparate scientific communities, with the unifying goal of developing novel graph-based techniques for better integrating and extracting value from diverse knowledge sources at large scale.

As a follow-up to the seminar, the attendees agreed that in order to foster this unifying view of Knowledge Graphs, there was a need for a manuscript that would serve as a general introduction to the area. This manuscript would:

- motivate knowledge graphs and the value of abstracting data as graphs;
- survey the historical context of knowledge graphs and the key initiatives leading to their popularization;
- draw together disparate views of knowledge graphs into a unifying definition;
- provide an introduction to the key techniques that knowledge graphs enable, relating to querying, validation, reasoning, learning, refinement, enrichment, quality assessment, and more besides;
- describe how knowledge graphs are used in practice, surveying the companies using knowledge graphs, the applications they are used for, the open knowledge graphs that have been published, etc.; and
- delineate future research directions for knowledge graphs.

The manuscript would then serve as an introductory text for students, practitioners and researchers new to the area, helping to form a consensus in terms of what is a knowledge graph, laying the foundations for future developments.

The goal of preparing this manuscript was an ambitious one, and involved drawing together and distilling down a vast amount of literature on a diverse range of topics into a set of key concepts described in an accessible way. For this reason, the manuscript has been prepared by many authors, who have lent their knowledge and expertise to the preparation of specific sections. A short version of the manuscript was first published as a tutorial paper [Hogan et al., 2021], consisting of an abridged version of the first five chapters of this book, along with a summary of how knowledge graphs are used in practice, and conclusions. However, there was not enough space to describe all of the important developments in the area. This led us to publish this book, which further includes topics relating to the creation, enrichment, quality assessment, refinement and publication of knowledge graphs, as well as formal definitions, a historical perspective, and extended discussion throughout.

The book is divided into ten chapters. Chapter 1 provides a general introduction to the area, defines the concept of a "knowledge graph", and provides a high-level overview of how knowledge graphs are currently being used. Chapter 2 presents and contrasts popular graph models that are commonly used to represent data as graphs, and the languages by which they can be queried. Chapter 3 describes how the resulting data graph can be enhanced with notions of schema, identity, and context. Chapter 4 discusses how ontologies and rules can be used to encode knowledge, and how they enable deductive forms of reasoning. Chapter 5 delves into how inductive techniques—based on statistics, graph analytics, machine learning, etc.—can be used to encode and extract knowledge. Chapter 6 is dedicated to techniques for the creation and enrichment of knowledge graphs from legacy sources of data. Chapter 7 enumerates a variety of quality measures that can be used to assess a knowledge graph in terms of its fitness for use in a variety of applications. Chapter 8 presents key methods for the refinement of knowledge graphs, with the goal of improving their completeness and correctness. Chapter 9 provides a survey of the open and enterprise knowledge graphs that have emerged in recent years, along with the industries within which, and the applications for which, they have been most widely adopted. Chapter 10 wraps up the book with discussion of the current limitations and future directions along which knowledge graphs are likely to evolve. An Appendix further covers knowledge graphs from an historical perspective, establishing their significance in the broader context of the academic study of data and knowledge, as well as surveying prior definitions of "knowledge graphs" from the literature.

A key aim of this book is to be accessible to a broader audience. While background knowledge of related topics such as Databases, Logic, Machine Learning, Semantic Web, etc., will help to understand some of the particular topics mentioned, such a background is not necessary to follow the general concepts described within. The book aims to motivate and illustrate the various concepts it introduces from a practical perspective, and in order to be as accessible as possible, relies heavily on an example-driven presentation using a graphical notation. For the reader wishing to dig more into the technical minutiae, we complement this discussion with formal definitions throughout; however, the reader more interested in understanding the general concepts and their rationale will find the discussion to be self-contained if they choose to skip the definitions presented in visually distinctive boxes.

The book serves as an entry point for those new to the topic, and may thus serve as a useful textbook for university courses, for researchers who are venturing into the topic for the first time, and for practitioners who wish to understand more about how knowledge graphs might be of use within their company or organization, or indeed, how to maximize the value of the knowledge graphs that they are currently developing. Readers who are already active within specific sub-areas of Knowledge Graphs may further appreciate the technical definitions included, the references to other literature provided, and the broader perspective that this book offers in terms of the other related sub-areas and how they complement each other.

By drawing together diverse techniques from disparate areas, Knowledge Graphs has become an exciting topic in terms of both research and applications. We expect to see growing interest on this topic as the years advance, and indeed hope that this book will help to more firmly establish the foundations of this topic, and to foster future developments upon these foundations, potentially by its readers.

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann
September 2021

Acknowledgments

We thank the organizers and attendees of the Dagstuhl Seminar on "Knowledge Graphs." We also thank those who provided feedback on this content.

Hogan was funded by Fondecyt Grant No. 1181896. Hogan & Gutierrez were funded by ANID Millennium Science Initiative Program, Code ICN17_002. Cochez did part of the work while employed at Fraunhofer FIT, Germany and was later partially funded by Elsevier's Discovery Lab. Kirrane, Ngonga Ngomo, Polleres & Staab received funding through the project "KnowGraphs" from the European Union's Horizon program under the Marie Skłodowska-Curie grant agreement No. 860801. Kirrane & Polleres were supported by the European Union's Horizon 2020 research and innovation programme under grant 731601. Labra was supported by the Spanish Ministry of Economy and Competitiveness (Society challenges: TIN2017-88877-R). Navigli was supported by the MOUSSE ERC Grant No. 726487 under the European Union's Horizon 2020 research and innovation programme. Rashid was supported by IBM Research AI through the AI Horizons Network. Schmelzeisen was supported by the German Research Foundation (DFG) grant STA 572/18-1.

Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan Sequeda, Steffen Staab, and Antoine Zimmermann
September 2021