Knowledge Graphs: A Guided Tour

Aidan Hogan
DCC, Universidad de Chile; IMFD Chile

Abstract

Much has been written about knowledge graphs in the past years by authors coming from diverse communities. The goal of these lecture notes is to provide a guided tour to the secondary and tertiary literature concerning knowledge graphs where the reader can learn more about particular topics. In particular, we collate together brief summaries of relevant books, book collections, book chapters, journal articles and other publications that provide introductions, primers, surveys and perspectives regarding: knowledge graphs in general; graph data models and query languages; semantics in the form of graph schemata, ontologies and rules; graph theory, algorithms and analytics; graph learning, in the form of knowledge graph embeddings and graph neural networks; and the knowledge graph life-cycle, which incorporates works on constructing, refining and publishing knowledge graphs. Where available, we highlight and provide direct links to open access literature.

2012 ACM Subject Classification Information systems → Graph-based database models; Information systems → Information integration; Computing methodologies → Artificial intelligence

Keywords and phrases knowledge graphs

Digital Object Identifier 10.4230/OASIcs.AIB.2022.1

Category Invited Paper

Funding *Aidan Hogan*: Supported by Fondecyt Grant No. 1181896 and by ANID – Millennium Science Initiative Program – Code ICN17_002.

Acknowledgements I wish to thank Camille Bourgaux, Ana Ozaki and Rafael Peñaloza for providing the idea for these lecture notes.

1 Introduction

Knowledge graphs have gained significant attention in recent years, both in industry and academia, as a way to integrate and leverage data and knowledge from diverse sources at large scale. Though the term "knowledge graph" had been used as far back as the 70's [60], it was the announcement of the Google Knowledge Graph [65] in 2012 around which a community began to crystallise [11]. The Google Knowledge Graph (in uppercase) was intended to be a proper name: the name, specifically, of a large, internal, graph-structured knowledge base housed within Google, which aimed to enhance Google's semantic search capabilities. Per the original blog post, Google envisaged using the knowledge graph as a way to shift from search based on "strings" to search based on "things" (or *entities*). Rather than performing string matching on searches like "`taj mahal`", the goal was to understand that this search likely refers to a thing—an *entity*: a mausoleum in Agra, India, on the bank of the river Yamuna (represented, in turn, by a node in the Google Knowledge Graph).

Arguably the greatest impact of the Google Knowledge Graph, however, was unanticipated. Other companies began to adopt the phrase "knowledge graph" in order to describe similar initiatives. Announcements of knowledge graphs by other well-known companies, such as Airbnb [16], Amazon [43], eBay [54], Facebook [51], IBM [18], LinkedIn [28], Microsoft [63], Uber [25], to name but a few, began to emerge. Meanwhile, "knowledge graphs" began to gain more and more traction in the academic literature [55, 70, 46]. Questions began to arise regarding how knowledge graphs relate to existing concepts such as graph databases, RDF graphs, ontologies, semantic networks, etc.; questions also began to arise about what,

International Research School in Artificial Intelligence in Bergen (AIB 2022).
Editors: Camille Bourgaux, Ana Ozaki, and Rafael Peñaloza; Article No. 1; pp. 1:1–1:21
OpenAccess Series in Informatics
OASICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

specifically, a "knowledge graph" even was [19, 14, 10, 11]. Ambivalent to such concerns, knowledge graphs continued to firmly establishing themselves in both industry and academia.

Some authors prefer to define a knowledge graph as "*a graph-structured knowledge base*" [50, 62]. Other authors go further, and assert additional necessary conditions that a knowledge graph must satisfy to be called as such, including the presence of a schema [53], the description of multiple domains [53], the use of an ontology and a reasoner to derive new knowledge [19], the presence of inference rules [10], etc. But perhaps the most accepted modern definition of a knowledge graph is as a "*graph whose nodes represent entities and whose edges represent relationships between those entities*" [70, 46, 51], sometimes adding the qualifier that the "*graph intends to collect and convey knowledge*" [33]. This latter definition is inclusive enough to embrace the diverse ways in which knowledge graphs have been used and studied, and is also the definition we adopt for the purposes of these lecture notes.

Why, then, have knowledge graphs gained so much traction and attention in practice, and in the literature? In practice, knowledge graphs have proven to be a useful abstraction for representing, integrating, managing and exploiting diverse data at large scale. Typical use-cases for knowledge graphs involve the extraction and accumulation of data from diverse sources within a unified graph-based representation. As Noy et al. [51] put it, "*knowledge graphs and similar structures usually provide a shared substrate of knowledge within an organization, allowing different products and applications to use similar vocabulary and to reuse definitions and descriptions that others create. Furthermore, they usually provide a compact formal representation that developers can use to infer new facts and build up the knowledge*" [51]. Such use-cases have given rise to the announcement not only of a broad range of enterprise knowledge graphs [65, 16, 43, 54, 51, 18, 28, 63, 25], internal to particular companies, but also open knowledge graphs [45, 13, 68, 31] available to the public.

In the academic setting, the use of graphs to represent both data and knowledge is far from a novel idea [24]; however, Knowledge Graphs, as an area, is remarkable in the degree to which it has brought together researchers from distinct communities that had previously been exploring graph-structured data, knowledge and analytics in isolation. In particular, the Database community has long explored *graph query languages and databases* as a way to query data stored as graphs; the Semantic Web community has long explored *graph data* and *ontologies* as a way to represent, integrate, interlink and reason over data on the Web; the Graph Theory and Algorithms community has long explored *fundamental properties of graphs*, and how different types of *graph analytics* can be used to understand and extract patterns from graphs (or networks); the Natural Language Processing and Information Extraction communities have long used *text graphs*, *semantic networks*, and other graph-based representations to induce structure from natural language. Notably, however, the area of Knowledge Graphs has also gained attention in communities where graphs have not traditionally been a mainstream topic; for example, within the Machine Learning community, the topics of *graph embeddings* and *graph neural networks* have recently gained significant traction, yielding novel machine learning techniques that can be applied natively over graphs. All these communities are now working on interrelated topics under the common umbrella of "knowledge graphs", which gives rise to a number of open questions regarding how such diverse techniques relate to or can complement each other [32].

Much has been written about knowledge graphs in the past decade, and in particular in the last five years, including a variety of books (e.g., [52, 56, 20, 40, 33]) and surveys of specific aspects of knowledge graphs (e.g., [53, 69]). Indeed the author of these lecture notes has recently co-authored a book [33] (among other contributions [32, 34]) on the topic. The volume of literature published in recent years has become somewhat overwhelming, in fact. Rather than providing yet another introduction to knowledge graphs, the goal of these lecture notes is to provide a roadmap of the existing literature on knowledge graphs and related topics, guiding the reader towards literature where they can learn more about the area in general, or about specific sub-topics of interest to them.

Thus our goal herein is to collate and summarise the secondary and tertiary literature that provides introductions, primers, surveys, perspectives, etc., regarding knowledge graphs and related topics. Specifically, we consider the following types of literature:

Books with global authorship (excludes edited collections).
Collections of edited book chapters on a specific topic (excludes proceedings).
Book chapters published in a collection not otherwise mentioned.
Articles published in journals.
Miscellaneous publications in conferences, workshops, or in preprint or online form.

Within each topic and category, we order papers by year and subsequently by author. We use the icon 🔓 to indicate an open access (OA) publication, and 🔓 to indicate a publication with an open access preprint or alternative version in a persistent repository (e.g., on arXiv). In the digital version of these notes, one can click on these icons to directly access the OA version. For the purposes of formatting, we may also abbreviate long titles, where we refer the reader to the bibliographic citation for the full details of the publication.

In terms of topics, we divide the literature into:

General: covering the broader topic of knowledge graphs.
Data: covering graph data models, databases and query languages.
Semantics: covering graph schemata, rules and ontologies.
Analytics: covering graph theory, algorithms, measures and frameworks.
Learning: covering knowledge graph embeddings, graph neural networks, etc.
Lifecycle: covering knowledge graph completion, refinement and publication.

The topics tend to become more technical as the lecture notes progress, so we recommend newcomers to check out some of the literature in the "General" category before proceeding. In the case of publications that cover multiple topics, we assign them to the topic that it places the most focus on. We include a range of different types of literature to suit different tastes, ranging from hands-on walkthroughs to theoretical treatments, from industry use-cases to academic conceptualisations, from high-level discussion to detailed definitions, etc.

Let's begin.

2 General: Knowledge Graphs

Here we cover literature providing a broad overview of the emerging area of Knowledge Graphs, including discussion of data, semantics, analytics, learning, and the knowledge graph lifecycle. The literature here provides a good starting point for newcomers.

Books

The Knowledge Graph Cook Book: Recipes That Work [12] (Blumauer & Nagy, 2020) introduces knowledge graphs in terms of data models, core concepts, and application scenarios. The book delves into organisational issues regarding enterprise knowledge graphs, before discussing specific aspects of (RDF) graph data, taxonomies and ontologies, and how knowledge graphs can be constructed. It concludes with discussion of enterprise knowledge graph architectures and services, set of interviews from industry leaders, and discussion of the future for knowledge graphs. The book is largely enterprise focussed.

Knowledge Graphs: Methodology, Tools & Selected Use Cases [20] (Fensel et al., 2020) first introduces knowledge graphs from a conceptual and practical viewpoint (contrasting open and enterprise/proprietary knowledge graphs), and then discusses how they can be constructed (knowledge graph creation, hosting, curation, and deployment) and used (AI use-cases, semantics, and dialogue systems). The final part of the book is dedicated to applications, and discussion of domains in which knowledge graphs can – and have been – deployed. The focus of the book is on building and maintaining knowledge graphs.

Knowledge Graphs: Data in Context for Responsive Businesses [8] (Barrasa et al., 2021) provides a practical introduction to knowledge graphs written by authors from Neo4j, targeted towards chief data officers. After motivating and defining knowledge graphs, the authors discuss how they can be modelled, and the role that taxonomies and ontologies play. The authors then compare actioning knowledge graphs vs. decisioning knowledge graphs, where the key focus is on data management and data analytics, respectively. The importance of context (for AI) is discussed, before the book concludes with an outlook. The book thus introduces and motivates knowledge graphs from an enterprise perspective.

Knowledge Graphs [33] (Hogan et al., 2021) provides a broad, conceptual introduction to knowledge graphs, covering graph data models and query languages, different forms of graph schema, contextual representations, ontologies and rules, and graph learning; it further discusses the creation, enrichment, quality, refinement, and publication of knowledge graphs. A brief survey of specific open and enterprise knowledge graphs, and their applications, is also provided. An appendix delves into the historical setting that gives rise to knowledge graphs. The presentation is mostly example-based. Formal definitions are provided, but can be skipped by the uninterested reader.

Knowledge Graphs: Fundamentals, Techniques & Applications [40] (Kejriwal et al., 2021) provides a broad introduction to knowledge graphs, covering popular graph-based data models, knowledge graph construction, knowledge graph completion, methods for querying and reasoning over knowledge graphs, and areas in which knowledge graphs have been successfully deployed. Aside from more general aspects, a major focus of the book is on creating knowledge graphs from text and other semi-structured sources, and subsequently refining the knowledge graph using (in particular) learning techniques.

Knowledge Graph [57] (Qi et al., 2022) is – at the time of writing – an upcoming book that is not yet available for download. According to the book's synopsis, it provides a systematic and comprehensive overview of knowledge graphs, their theoretical foundations, key techniques, methodologies, and applications. A key focus of the book is on the construction and management of knowledge graphs, including information extraction from text, as well as key techniques for knowledge fusion and reasoning.

Collections

Exploiting Linked Data and Knowledge Graphs in Large Organizations [52] (Pan et al., eds., 2017) is a collection of chapters, starting with an introduction to knowledge graphs, covering related standards, architectures for enterprise knowledge graphs, knowledge graph construction, ways in which knowledge graphs can be summarised and explored, question answering, and more besides. The book concludes with a discussion of applications and future directions. Overall the book is largely focussed on the Semantic Web.

Knowledge Graphs: New Directions for KR on the Semantic Web [14] (Bonatti et al., eds., 2018) collects together a number of short reports resulting from the discussions at

Dagstuhl Seminar 18371. These reports summarise discussions among participants on a wide range of topics relating to knowledge graphs, including evolution and dynamics, scholarly knowledge graphs, logic and learning, knowledge graph creation and management, symbolic reasoning, multilingualism, privacy and access control, graph analytics, etc.

Knowledge Graphs and Big Data Processing [35] (Janev et al., eds., 2020) is an open access collection of chapters that provide a general introduction to knowledge graphs, and their application for processing and managing data at large scale. Various authors have contributed chapters on both Knowledge Graphs and Big Data, covering, in the former case, an introduction to graph-based knowledge representation, the creation of knowledge graphs, data exchange using knowledge graphs, knowledge graph embeddings and their applications, and more. A key focus of the book is on how knowledge graphs can be leveraged for the purposes of Big Data applications.

Knowledge Graphs for eXplainable Artificial Intelligence [67] (Tiddi et al., eds., 2020) is a collection of chapters that introduce knowledge graphs more from an AI perspective. Earlier chapters cover knowledge graphs on the Web, embeddings, explainability in the context of knowledge graphs, and benchmarks. Chapters on applications include recommender systems, natural language processing, context understanding, explanations, transfer learning, and predictive analytics. The book concludes with an outlook to the future, as well as ethical and social issues surrounding knowledge graphs and explainable AI. The book thus largely focuses on knowledge graphs in the context of learning and AI.

Book chapters

Knowledge Graphs: Research Directions [32] (Hogan, 2020) is a book chapter providing a more technical introduction to knowledge graphs, including formal definitions of concepts such as graph models, queries, ontologies, rules, context, embeddings, and graph neural networks. A key aim of the chapter is to synthesise a set of research problems that arise from how these concepts potentially relate and complement each other, presenting a list of nine research topics that intersect different areas.

Articles

Knowledge Graphs [24] (Gutierrez and Sequeda, 2021) is an article discussing the history of knowledge graphs, and the phenomena that influenced them and led to their popularisation. The paper traces knowledge graphs back to ancient traditions of representing knowledge in diagrammatic form, taking a journey through the advent of logic, information retrieval, semantic networks, knowledge representation, the Web, and finally knowledge graphs. The focus of the article is on setting knowledge graphs in a broader historical perspective relating to data and knowledge.

Knowledge Graphs [34] (Hogan et al., 2021) is an article providing a tutorial on some of the key concepts and techniques underlying knowledge graphs. The article first motivates knowledge graphs, and then discusses graph data models, graph query languages, shapes for validation, and contextual representations. An introduction to ontologies and reasoning is then followed by discussion of graph analytics, knowledge graph embeddings, graph neural networks, and symbolic learning. The article concludes with future directions.[1]

Knowledge Graphs 2021: A Data Odyssey [71] (Weikum, 2021) is a position paper that begins with a brief overview of open and enterprise knowledge graphs, their applications,

[1] This article was extended into the book of the same name, mentioned previously [33].

and the challenges they pose. The paper then poses a number of positions, namely that knowledge graphs are more than simple graphs (often they consider higher-arity relations, provenance, constraints, etc.), and that knowledge graphs should prioritise precision (correctness) over recall (coverage), and thus should be constructed from select sources and incremental processes. The paper concludes with a list of open research challenges and problems that could be addressed with techniques from the area of Databases.

Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases [72] (Weikum et al., 2021) is a comprehensive article (comprising 380 pages) discussing the creation and curation, more generally, of knowledge bases. However, much of the content relates both directly and indirectly to knowledge graphs, which are also explicitly mentioned in various sections. The article covers key concepts relating to knowledge bases, methods for knowledge integration, techniques for knowledge base construction and curation, for schema construction, and for resolving entities. The article concludes with discussion of open and enterprise knowledge graphs, and an outlook to the future.

Miscellaneous

Towards a Definition of Knowledge Graphs [19] (Ehrlinger and Wöß, 2016) investigates the concept of knowledge graphs and the various – and sometimes incompatible or even contradictory – definitions attributed to them. They highlight that the differences between knowledge bases, knowledge graphs and ontologies remain unclear. Applying a terminological analysis, they arrive at the definition that "*A knowledge graph acquires and integrates information into an ontology and applies a reasoner to derive new knowledge*" [19].[2]

3 Data: Models and Query Languages

This section pertains to literature that discusses graph data models and graph query languages. Much of the literature comes from the (Graph) Database and Semantic Web communities, which have traditionally been hot spots for research on graph data management techniques.

Books

Graph Databases [58] (Robinson et al., 2015) provides a practical introduction to graph data models and graph databases. The book walks through various examples using the Neo4j graph database, and the Cypher query language for graphs. The authors discuss the creation of applications on top of a graph database, and real-world use-cases where such databases are deployed. Techniques implemented by graph databases (specifically Neo4j) are also discussed. The final chapter concludes with a discussion of graph algorithms, focusing particularly on graph search (depth- and breadth-first search, Dijkstra, A*, etc.). The book has a practical focus using concrete examples from Neo4j/Cypher.

Querying Graphs [15] (Bonifati et al., 2018) provides a technical introduction and overview of the state-of-art with respect to graph databases and graph query languages. The book focuses on the property graph model and its variants. Different formal fragments of query languages are introduced, including regular path queries, unions of conjunctive

[2] This definition was contested by later authors [34], given that it would exclude many initiatives surrounding knowledge graphs not involving ontologies, such as works in the machine learning community on graph representation, and (seemingly) the Google Knowledge Graph itself, which gave origin to the modern use of the phrase; in these settings, ontologies are not (always) used.

queries, relational queries, regular queries, etc. Graph constraints are introduced in terms of functional and entity dependencies. The book then discusses query specification, which incorporates the paradigms of query-by-example and reverse-engineering of queries. The latter part of the book turns to implementation issues, including representation, compression, indexing, query processing, physical operators, and cardinality estimation. It concludes with some open research challenges. The book is targeted at a more academic audience, providing formal definitions and theoretical discussion throughout.

Graph Databases in Action [9] (Bechberger and Perryman, 2020) gives a practical guide to graph databases with concrete examples provided in the Gremlin graph traversal language. The book first discusses graph data modelling from a practical viewpoint. It then discusses how to run graph traversals, insertions, deletions, paths, filters, subgraph extraction and graph analytics over graphs using Gremlin. Practical issues, such as application development and performance pitfalls, are further discussed. The book has a practical focus and is principally targeted at software developers.

Collections

Graph Data Management: Fundamental Issues & Recent Developments [21] (Fletcher et al., eds., 2018) collates six chapters relating to graph data management. The first chapter provides a general introduction to graph data management. Next, techniques for graph visualisation are discussed, including planarisation, energy-based approaches, and approaches for large graphs. The third chapter discusses methods for discovering motifs (frequently-occurring subgraph patterns) in graphs. The following chapter describes query relaxation and approximation in the context of flexible query processing for graphs. The fifth chapter provides an overview of parallel processing frameworks for graphs. The final chapter concludes with a survey of benchmarks for graph processing systems.

Book chapters

Storing and Querying Semantic Data in the Cloud [36] (Janke and Staab, 2018) offers lecture notes on techniques for storing and querying RDF graphs in local, distributed and cloud environments. After some preliminary definitions on RDF graphs and SPARQL queries, the lecture notes describe different architectures for RDF stores in local, cloud-based, distributed, peer-to-peer and federated settings. The notes describe partitioning and replication strategies for RDF graphs, popular indexing techniques, distributed query processing strategies, fault tolerance, and available (RDF/SPARQL) benchmarks.

Articles

Survey of Graph Database Models [5] (Angles et al., 2008) is a survey of a variety of graph-based data models that have been proposed in the literature down through the years. After a general introduction to the history, main concepts and applications driving graph-based data modelling, the paper enumerates different graph database models, ranging from straightforward models with simple nodes and edges, to more complex models that support hypernodes (with nested elements), attributes on nodes and edges, relations viewed as entities, derivation and inheritance, nested relations, constraints, schema, etc. The survey also covers query languages proposed for such models.

Query Languages for Graph Databases [73] (Wood, 2012) provides an overview and formal definition of the different primitives – specifically conjunctive queries, regular path queries, conjunctive regular path queries, and extended conjunctive regular path queries – underlying graph query languages. A brief survey is provided of graph query languages and the features they support, including also path comparisons, aggregation operators, node creation, approximate matching, ranking, etc. The article concludes with a discussion of the expressive power of different fragments of graph query languages.

Foundations of Modern Query Languages for Graph Databases [4] (Angles et al., 2017) covers popular data models and graph query languages from a conceptual point of view. The article specifically covers the directed edge-labelled graph (e.g., RDF) and property graph models. It then delves into the core primitives and semantics underlying graph query languages, including basic graph patterns, path queries, and relational algebra. Examples in concrete query languages – specifically SPARQL, Cypher and Gremlin – are presented. More advanced querying primitives, featuring recursion, are also covered.

RDF Data Storage and Query Processing Schemes: A Survey [75] (Wylot et al., 2018) surveys data management techniques for RDF graphs, including techniques for indexing and processing queries over large-scale RDF graphs. The survey covers the storage and indexing of RDF graphs both on individual machines, as well as in distributed settings over multiple machines (using NoSQL, Hadoop, Spark, etc.). Federated query processing is also discussed. The paper concludes with a survey of different SPARQL-based benchmarks.

A Survey of RDF Stores & SPARQL Engines for Querying Knowledge Graphs [2] (Ali et al., 2021) is a detailed survey on systems for querying RDF graphs. After some preliminaries on RDF/SPARQL, the survey discusses different storage schemes and indexing techniques that enable efficient access over RDF graphs. Thereafter, both traditional and more modern join processing techniques are described, followed by query processing techniques involving other relational operators and paths. Different partitioning strategies are studied and compared. An appendix includes a survey of over one hundred distributed RDF stores and SPARQL query engines, and the techniques they use.

Miscellaneous

Querying in the Age of Graph Databases & Knowledge Graphs [6] (Arenas et al., 2021) provides notes for a tutorial on querying graphs in the era of knowledge graphs. The paper begins by studying the growth in popularity surrounding knowledge graphs, finding fewer than one hundred papers on the topic in DBLP in 2015, which grows to almost one thousand in 2020. The paper asks: what is new about knowledge graphs and how do they relate to graph databases? A brief discussion is provided on data vs. information vs. knowledge, graph data and querying, and traditions in which graphs have been used to model knowledge (semantic networks, graph databases, semantic web). The authors outline their perspective on knowledge graphs, characterised by knowledge representation, integration and production using graphs. The paper then formalises different graph data models and query language features, and concludes that the popularity of graphs for representing knowledge is due to their being a "*simple, flexible and extensible data structure*", while also being a "*a deep-rooted form of representing human knowledge*".

4 Semantics: Schemata, Rules and Ontologies

Explicit representations of semantics play an important role in many knowledge graphs. Here we provide pointers to literature covering the topics of graph schemata (in various forms), rules, and ontologies. Much of this literature comes from the Knowledge Representation and Semantic Web communities, who have long studied and debated these topics.

Books

Foundations of Semantic Web Technologies [30] (Hitzler et al., 2010) introduces key Semantic Web standards and concepts using a mix of formal definitions and examples in concrete syntax. The book begins with a general motivation and historical perspective with respect to semantics, knowledge and reasoning. The book then describes the RDF data model, the RDFS schema language, and their formal semantics. The book continues with an in-depth treatment of OWL, covering its syntax, features, and formal semantics. Rules are introduced and their relation with ontologies discussed. Query languages for RDF (including SPARQL) are presented. The book concludes with discussion on ontology engineering topics and applications of Semantic Web concepts and standards. The book uses a mix of formal definitions and examples in concrete syntax, separating introductory material from more advanced concepts. It includes exercises for students.

Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL [3] (Allemang and Hendler, 2011) provides a pragmatic introduction to the Semantic Web; the importance of semantics; various key standards including RDF, RDFS, OWL, SKOS, and SPARQL, and how they can be used on the Web. A particular focus of the book is on how these standards can be used to model data and semantics, and how they enable reasoning and querying, both in local and decentralised (Web) settings. The book also touches on the importance of semantic modelling, in terms of human communication, explanation, prediction, and integrating heterogeneous data. The focus is on the use of lightweight ontology features for semantic modelling. The discussion, though often conceptual, is largely example-based, using concrete syntaxes. Thus the book is suitable both for an academic audience, as well as practitioners interested in semantic modelling.

An Introduction to Description Logic [7] (Baader et al., 2017) provides a comprehensive introduction to the area of Description Logics, which studies decidable fragments of first order logic that form the basis of ontology languages like OWL, and can be used to define the semantics of knowledge graphs. The book first introduces basic description logics, and then describes how to define their semantics with model theory. Reasoning algorithms for expressive description logics based on tableau are introduced, and the complexity of reasoning problems is studied. The latter part of the book is dedicated to more practical aspects of description logics, including tractable profiles, query answering, concrete ontology languages, and applications. The book is primarily aimed at an academic audience, and uses formal definitions alongside examples to introduce concepts.

Validating RDF Data [22] (Labra Gayo et al., 2017) discusses the use of shapes and shape languages (including SHACL and ShEx) for validating RDF graphs. Such languages allow for specifying constraints over a graph, for encapsulating and combining multiple constraints as a “shape”, for targeting specific nodes of the graph with particular shapes, and for validating graphs with respect to the shapes defined. Thus, shape languages can be seen as a form of validating schema for graphs. The book begins by introducing RDF graphs, issues relating to data quality, and the key concepts underlying shape languages. It then covers the ShEx and SHACL languages. The book concludes by discussing applications for shapes, and comparing the ShEx and SHACL languages. The discussion is example driven, using concrete syntax, and suitable for a broad audience.

An Introduction to Ontology Engineering [38] (Keet, 2018) provides a general introduction to the titular area. The book begins by discussing the notion of an ontology in Computer Science, and how ontologies are used. It then describes the logical foundation of ontologies, covering first-order logic, description logics and the OWL 2 ontology language. The next section of the book is devoted to ontology development, covering methodologies, tools, top-down approaches and bottom-up approaches. The book concludes with discussion of modern ontology engineering topics, including ontology-based data access, natural language in ontologies, contextual frameworks for representing uncertainty and temporal validity, and finally ontology modularisation. The book uses a mix of formal and example-driven presentation, with exercises provided for students.

Ontology Engineering [41] (Kendall and McGuinness, 2019) provides a pragmatic introduction to ontology engineering, with an emphasis on modelling. After an introduction to foundational aspects of ontologies, the book provides an overview of some key concepts, such as domain analysis, levels of abstraction, ontology evaluation, ontology design patterns, etc. The book then describes how to collect requirements and analyse use-cases. The following chapter introduces the importance of terminology for ontology engineering, and how to collect and curate domain terms. The book wraps up with discussion on conceptual modelling, including ontology reuse, naming conventions, metadata, etc. The book has a practical focus, relying on a more informal, didactic presentation.

The Web of Data [32] (Hogan, 2020) offers a comprehensive discussion of the Semantic Web standard, the Linked Data principles, and how they come together to realise a Web of Data. The book offers a general motivation for the Web of Data, before providing a detailed discussion of how graphs (RDF), semantic schemata (RDFS), ontologies (OWL), query languages (SPARQL), shapes (SHACL), and publishing principles (Linked Data) combine to enable data, and not just documents, to be interlinked on the Web. Each chapter provides a general motivation, examples in concrete syntax, as well as formal definitions, such that it can serve as both an introductory text book, and a reference book. Examples are included throughout for students to test their learning.

Collections

Ontology Engineering in a Networked World [66] (Suárez-Figueroa et al., eds., 2012) is a collection of book chapters relating to the topic of ontology engineering. The chapters broach a wide range of topics, including concrete methodologies for ontology engineering, ontology design patterns, requirements specification, ontology localisation, modularisation of ontologies, ontology evolution, and ontology matching. The book concludes with some more practice-oriented chapters based on the NeOn ontology engineering toolkit.

Applications and Practices in Ontology Design, Extraction, and Reasoning [17] (Cota et al., eds., 2020) collects together thirteen chapters relating to ontologies and rules. The diverse topics covered include ontology modularisation and reuse, FAIR principles, knowledge creation through mapping languages, probabilistic and preferential description logics, axiom pinpointing, defeasible reasoning, querying and reasoning via rules, as well as applications of ontologies within the humanities, the scholarly domain, and music.

Book Chapters

Foundations of Description Logics [59] (Rudolph, 2011) provides lecture notes on the theoretical foundations of description logics. The notes begin with a general introduction to description logics and the Semantic Web. The semantics of description logics are formally defined through model theory. Different description logics, and the features they support, are introduced. A number of semantic relations are introduced, namely concept equivalence, ontology equivalence and emulation. A primer is provided on modelling with description logics, showing how transitivity, cardinality constraints, etc., can be axiomatised; the open and closed world assumptions are also discussed. Reasoning tasks relating to satisfiability and entailment are described, and reasoning algorithms are sketched. Finally the concrete OWL language is introduced and mapped to description logics. The book offers a mostly formal treatment (with examples) of the topic.

5 Analytics: Theory, Algorithms, Measures & Frameworks

Graphs have long been used to conceptualise networks and interactions in a variety of domains. Analytics such as centrality, community detection, spectral analysis, etc., can then glean important insights about the respective domains from such graphs. Here we discuss literature regarding the graph theory, algorithms and measures that underlie such analytics, and the frameworks used to compute them. The literature here mainly stems from the Graph Theory and (various) Network Analysis communities, with graph processing frameworks studied by the Big Data, Database, Distributed Systems, and other related communities.

Books

Graphs, Algorithms and Optimization [42] (Kocay and Kreher, 2005) provides a general introduction to graphs, graph algorithms, and optimisation techniques. After some preliminaries on graphs, the book introduces paths, walks, and algorithms to find shortest (weighted) paths. Special classes of graphs are introduced, including bipartite graphs, line graphs, Moore graphs, and Euler tours. The book discusses trees and cycles, including algorithms for spanning trees, tree encodings, etc. Next, connectivity is discussed, including the notion of blocks, and algorithms to detect them. The book continues by introducing concepts relating to alternating paths, matchings, network flows, Hamilton cycles, digraphs, graph colourings, planar graphs, and graph embeddings on surfaces. Optimisation techniques are introduced in the form of linear programming, discrete linear programming, and related algorithms. The book thus provides a quite formal and technical introduction to graph theory, algorithms and optimisation.

Systems for Big Graph Analytics [76] (Yan et al., 2017) discusses parallel frameworks for distributed processing and analytics over large-scale graphs. The book introduces three main computational models for processing graphs in a distributed or parallel setting: vertex-based, block-based, subgraph-based and matrix-based. Vertex-based computation involves message passing between vertices of the graph, which then perform computations on the messages received. Block-based computation partitions vertices into densely-connected blocks, and defines separate communication and computation within and between blocks. In subgraph-based computation, (possibly overlapping) subgraphs are grown dynamically from seed vertices during the computation. Matrix-based computation views the graph as an adjacency or incidence matrix, and applies computation in terms of linear algebra operations. The book discusses technical concepts relating to communication and load balancing, out-of-core computation, fault tolerance, on-demand querying, shared memory abstractions, partitioning, and more besides. Systems such as Pregel, Giraph, GraphX, GraphLab, BigGraph@CUHK, Blogel, G-Thinker, PEGASUS, GBASE, SystemML, etc., are discussed throughout in relation to these concepts.

Graph Algorithms [49] (Needham and Hodler, 2019) features concrete examples of how to implement a variety of graph algorithms in Apache Spark and Neo4j. The book first contrasts transactional (OLTP) and analytical (OLAP) workloads, and describes use-cases for graph analytics. A primer on graph theory is provided, including properties of graphs (random, small-world, scale-free, etc.), different graph models, etc. Graph processing frameworks are introduced. The book then delves into specific graph algorithms and measures (and their implementation) for graph search/path finding, centrality measures and community detection. Practical use-cases are presented. The book concludes by discussing the relation between graph analytics and machine learning techniques. The book adopts a hands-on approach to introducing different graph algorithms.

The Practitioner's Guide to Graph Data [23] (Gosnell and Broecheler, 2020) provides a hands-on guide to modelling data as graphs and implementing various types of analytics and algorithms over those graphs using Gremlin, Cassandra, among other tools. After an introduction to graph data modelling, and an example use-case, the book delves into traversals using Gremlin, the use of Cassandra to index graphs, as well as techniques for navigating hierarchical data modelled as trees, graph search and path finding algorithms, collaborative filtering for recommendations, and entity resolution. The book concludes with discussion of graph algorithms, distributed graphs, graph theory and network theory. The book thus blends practical examples with more conceptual discussion.

Collections

Managing and Mining Graph Data [1] (Aggarwal and Wang, eds., 2010) is a collection of book chapters on graph data management and mining. The chapters of the book cover a variety of topics including graph data models, formal properties of graphs, graph generators, graph query languages, graph indexing, path queries and graph pattern matching, graph matching, keyword search on graphs, clustering algorithms, dense subgraph discovery, graph classification through kernels and boosting, frequent subgraph mining, streaming graphs, privacy for graphs, etc. The latter part of the book focuses on graph mining for use-cases in specific domains including the Web, social networks, software, biology, and chemistry. The collection largely targets an academic audience, with formal definitions and technical discussion found throughout its different chapters.

6 Learning: Embeddings and Architectures

One of the key sub-areas of Knowledge Graphs has been representation learning for graphs. Key techniques here include graph embeddings, which involve learning numerical representations for nodes (entities), edge labels (relations), and graphs themselves; and graph neural networks, which layer learned functions and message passing over the topology of the graph. The literature here is predominantly from the Machine Learning community.

Books

Graph Representation Learning [26] (Hamilton, 2020) is a book covering a broad range of topics relating to learning over graphs. The book begins by introducing different graphs models, and different abstract machine learning tasks one can consider over graphs. Next, different graph measures and algorithms are presented, along with spectral graph theory. The book then delves into the technical details of node embeddings and graph neural networks. It concludes with discussion of generative graph models.

Deep Learning on Graphs [47] (Ma and Tang, 2021) focuses on deep learning techniques for graphs. After a general introduction to and motivation for the topic, the authors discuss graph models, metrics, computational frameworks, and spectral graph theory. A similar introduction is provided for deep learning, including feedforward networks, convolutional and recurrent neural networks, autoencoders, and training methodologies. The book then focuses on graph embeddings, graph neural networks, and other deep learning models applied over graphs. It concludes with discussion of applications of graph neural networks for natural language processing, computer vision, data mining, and biochemistry. The book concludes with advanced topics, such as the expressivity of graph neural networks, combinatorial optimisation on graphs, etc.

Articles

Representation Learning on Graphs: Methods and Applications [27] (Hamilton et al., 2017) is an article reviewing techniques and applications relating to learning over graphs. The review primarily focuses on node and subgraph embeddings. The survey covers techniques for node embeddings based on factorisation, random walks, encoder–decoder architectures, and more besides; applications for node embeddings, such as pattern discovery, community detection, node classification and link prediction, are discussed. Techniques surveyed for subgraph embeddings include convolutional approaches, graph-coarsening, and graph neural networks; applications discussed for subgraph embeddings include subgraph classification, drug discovery, molecular classification, image classification, computer programme verification, and logical reasoning.

Knowledge Graph Embedding: A Survey of Approaches and Applications [69] (Wang et al., 2017) provides a comprehensive survey on knowledge graph embeddings, divided into translational models (TransE and related embeddings, Gaussian embeddings, etc.), and semantic matching models (RESCAL and related embeddings, neural-based embeddings, etc.). Within each category, specific embeddings are defined in terms of their entity embeddings, relation embeddings, plausibility scoring function, and constraints. The survey then discusses model training, and how negative examples can be extracted from knowledge graphs for training (under an open/closed world assumption). Embeddings are compared in terms of time and space complexity. The survey discusses how models can encode additional information, such as entity types, paths, text, rules, attributes, temporal metadata, and graph structures. The paper wraps-up with discussion of applications internal (link prediction, triple/entity classification, entity resolution) and external (relation extraction, question answering, recommendations) to the knowledge graph.

A Comprehensive Survey on Graph Neural Networks [74] (Wu et al., 2021) provides a detailed survey of advances on graph neural networks. After a general introduction, brief history, and preliminary definitions, the authors present a taxonomy of different graph neural networks that includes recurrent graph neural networks, convolutional graph neural networks, graph autoencoders, and spatio-temporal graph neural networks. They also identify tasks at different levels – node-level, edge-level and graph-level – and different learning paradigms – semi-supervised, supervised and unsupervised (embeddings). Within each category, the survey lists the specific graph neural networks that have been proposed, their inputs, their pooling and readout operations, and their time complexity. Theoretical aspects, such as VC dimension, expressivity with respect to graph isomorphism tests, universal approximation, etc., are also briefly discussed. The survey concludes with discussion of datasets, tasks, applications, and future directions.

A Survey on Knowledge Graphs: Representation, Acquisition, and Applications [37] (Ji et al., 2022) is a survey article that focuses primarily on representation learning and knowledge acquisition in the context of knowledge graphs. The survey first provides a technical introduction to representational concepts such as knowledge graph embeddings, plausibility scoring functions, encoding models, and contextual embeddings. In terms of knowledge acquisition, techniques for knowledge graph completion, entity discovery, and relation extraction are discussed. Temporal aspects are considered, along with applications relating to language models, question answering, and recommender systems.

Miscellaneous

Graph Neural Networks Meet Neural-Symbolic Computing: Survey & Perspective [44] (Lamb et al., 2020) provides a concise overview of graph neural networks in the broader context of neural-symbolic computing, and how the two interrelate. The paper first recaps different classifications of neural-symbolic computing frameworks, classifying graph neural networks as a TYPE 1 system, i.e., based on standard deep learning techniques (arguably not even neural-symbolic given the lack of symbolic representations). The paper discusses developments leading up to graph convolutional networks, graph neural networks, message-passing neural networks, and graph attention networks. Thereafter, the paper highlights the promise of combining graph neural networks with neural-symbolic reasoning through applications relating to relational learning and reasoning, combinatorial optimisation, and constraint satisfaction problems.

7 Lifecycle: Construction, Refinement, Publication

The final collection of literature that we consider relates to what we broadly call the "lifecycle" of knowledge graphs, encompassing knowledge graph construction, refinement, and publication. The literature discussed here comes primarily from the Database, Information Extraction, Machine Learning, and Semantic Web communities.

Books

Linked Data: Evolving the Web into a Global Data Space [29] (Heath and Bizer, 2011) describes how RDF graphs and Linked Data principles can combine to form a Web of Data that can be conceived of as a decentralised knowledge graph that spans the Web. The core concept is to use RDF as a graph-structured data model in which Web identifiers – URIs or IRIs – form the nodes and edge labels of the graph. Performing a HTTP lookup on these identifiers resolves – or *dereferences* – to a potentially remote RDF graph providing further data about the entity or relation (type) identified. Though they pre-date the modern popularisation of knowledge graphs, these principles have been used to publish a variety of important open knowledge graphs on the Web, including DBpedia, Freebase, Wikidata, YAGO, amongst (many) others. The book thus introduces principles and best practices for publishing graph-structured data on the Web.

Domain-Specific Knowledge Graph Construction [39] (Kejriwal, 2019) provides a brief general introduction to knowledge graphs along with some domain-specific examples. The book then focuses on a number of specific topics relating to the construction and refinement of knowledge graphs, specifically information extraction techniques, entity resolution, and knowledge graph completion. The book concludes with a discussion of ecosystems in which knowledge graphs have been broadly deployed, such as Linked Data, Google Knowledge Graphs, and Schema.org.

Designing and Building Enterprise Knowledge Graphs [61] (Sequeda & Lassila, 2021) gives a general introduction to different types of data that can be found in an enterprise setting, and the benefits of mapping data to knowledge graphs. A core focus of the book is on mapping relational databases to knowledge graphs, where it covers relevant mapping languages and mapping design patterns. The book further discusses the people, processes and tools involved in building and maintaining an enterprise knowledge graph.

Web Data APIs for Knowledge Graphs [48] (Meroño-Peñuela et al., 2021) discusses techniques and standards by which the content of (open) knowledge graphs can be published on the Web and accessed through APIs. After introducing knowledge graphs, the Linked Data principles, RDF, SPARQL and GraphQL, the book discusses how knowledge graphs can be accessed using HTTP requests, REST APIs, and SPARQL services, further discussing a number of tools to define REST APIs based on SPARQL queries, and to transform the response for SPARQL queries. The book is thus of particular interest to readers interested in publishing knowledge graphs on the Web.

Articles

Quality Assessment for Linked Data: A Survey [77] (Zaveri et al., 2016) collates a set of quality dimensions for Linked Datasets collected from the literature. Given that Linked Data refers to a set of principles for publishing graph-structured (RDF) data on the Web, many (though not all) dimensions apply for knowledge graphs [34]. The dimensions are organised into four high-level categorisations. Accessibility encompasses dimensions relating to how data can be accessed such as availability, licensing, interlinking, security, and performance. Intrinsic dimensions refer to potential issues within the data, such as syntactic validity, semantic accuracy, consistency, conciseness, and completeness. Contextual dimensions depend on a particular purpose, and include relevancy, trustworthiness, understandability, and timeliness. Representational dimensions deal with how data are represented, and include conciseness, interoperability, interpretability, and versatility. The paper includes discussion of measures for dimensions, as well as tools to detect issues.

Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods [53] (Paulheim, 2017) provides a comprehensive overview of techniques for refining knowledge graphs. The article first introduces the notion of a knowledge graph in the context of the Semantic Web, surveying prominent open and enterprise knowledge graphs. A categorisation of knowledge graph refinement approaches are then presented, based on whether they aim to complete or detect errors in the knowledge graph, what elements of the knowledge graph they target, and whether or not they depend on external sources. A categorisation of evaluation methods is also presented, including partial goal standards, using the knowledge graph itself as a silver standard, retrospective evaluation, and performance evaluation. Under techniques for knowledge graph completion, the survey considers methods that complete type assertions, and that predict relations. Techniques considered for error detection include binary classification (correct/incorrect), statistical techniques, reasoning, etc. The survey concludes with a summary of its findings.

Miscellaneous

Knowledge Graph Lifecycle: Building & Maintaining Knowledge Graphs [64] (Simsek et al., 2021) defines a lifecycle for knowledge graphs involving four stages: knowledge creation, knowledge hosting, knowledge curation (incorporating assessment, cleaning and enrichment), and knowledge deployment. Knowledge creation can involve manual processes or mappings from legacy sources. Knowledge hosting primarily refers to storing the graph in a database, enabling it to be queried. Knowledge curation involves three sub-phases: assessment aims to gather information about the quality of the knowledge graph; cleaning aims to pinpoint and address specific quality issues; enrichment aims to improve the completeness of the knowledge graph. Finally, knowledge deployment refers to the use of the knowledge graph for (e.g., end-user) applications. The paper concludes with lessons learnt from practical deployments of knowledge graphs.

8 Conclusions

Knowledge graphs have been the subject of a vast amount of publications in recent years. Our goal in these lecture notes has been to provide the reader with some orientation on how to approach this literature: on where to learn more about knowledge graphs in general, or about topics relating to data, semantics, analytics, learning, or the knowledge graph lifecycle. Our hope is that by the time you have read this far, you will have deviated from these notes in order to go out and explore some of the literature mentioned.

Much of the literature discussed herein deals with state-of-the-art techniques, tools, languages, etc., relating to general or specific aspects of knowledge graphs. But what about the future of Knowledge Graphs as a research area? As discussed in the introduction to these lecture notes, we expect Knowledge Graphs to establish itself as a research area in its own right, complete with its own conferences, journals, etc. This area will ideally serve as a confluence point for researchers coming from different constituent communities to pursue novel ideas relating to the use of a graph abstraction to collate and deploy knowledge at large scale. In this sense, the future of Knowledge Graphs should hopefully see combinations of techniques from different areas, such as graph representation learning that can consider semantics expressed as ontologies or rules, or combinations of graph querying and analytics in the form of hybrid languages and query processing tools, etc. For more discussion on possible future research directions for knowledge graphs, we refer the interested reader to the lecture notes "*Knowledge Graphs: Research Directions*" [32].

References

1 Charu C. Aggarwal and Haixun Wang, editors. *Managing and Mining Graph Data*, volume 40 of *Advances in Database Systems*. Springer, 2010. doi:10.1007/978-1-4419-6045-0.

2 Waqas Ali, Muhammad Saleem, Bin Yao, Aidan Hogan, and Axel-Cyrille Ngonga Ngomo. A Survey of RDF Stores & SPARQL Engines for Querying Knowledge Graphs. *VLDB Journal*, 2021. doi:10.1007/s00778-021-00711-3.

3 Dean Allemang and James A. Hendler. *Semantic Web for the Working Ontologist - Effective Modeling in RDFS and OWL, Second Edition*. Morgan Kaufmann, 2011. URL: http://www.elsevierdirect.com/product.jsp?isbn=9780123859655.

4 Renzo Angles, Marcelo Arenas, Pablo Barceló, Aidan Hogan, Juan L. Reutter, and Domagoj Vrgoc. Foundations of Modern Query Languages for Graph Databases. *ACM Computing Surveys*, 50(5):68:1–68:40, 2017. doi:10.1145/3104031.

5 Renzo Angles and Claudio Gutiérrez. Survey of graph database models. *ACM Computing Surveys*, 40(1):1:1–1:39, 2008. doi:10.1145/1322432.1322433.

6 Marcelo Arenas, Claudio Gutiérrez, and Juan F. Sequeda. Querying in the Age of Graph Databases and Knowledge Graphs. In Guoliang Li, Zhanhuai Li, Stratos Idreos, and Divesh Srivastava, editors, *SIGMOD '21: International Conference on Management of Data, Virtual Event, China, June 20-25, 2021*, pages 2821–2828. ACM, 2021. doi:10.1145/3448016.3457545.

7 Franz Baader, Ian Horrocks, Carsten Lutz, and Ulrike Sattler. *An Introduction to Description Logic*. Cambridge University Press, Cambridge, United Kingdom, 2017. URL: https://doi.org/10.1017/9781139025355.

8 Jesus Barrasa, Amy E. Hodler, and Jim Webber. *Knowledge Graphs: Data in Context for Responsive Businesses*. O'Reilly Media, 2021.

9 Dave Bechberger and Josh Perryman. *Graph Databases in Action*. Manning, 2020.

10 Luigi Bellomarini, Daniele Fakhoury, Georg Gottlob, and Emanuel Sallinger. Knowledge Graphs and Enterprise AI: The Promise of an Enabling Technology. In *35th IEEE International*

Conference on Data Engineering, ICDE 2019, Macao, China, April 8-11, 2019, pages 26–37. IEEE Computer Society, 2019. URL: `https://doi.org/10.1109/icde.2019.00011`.

11 Michael K. Bergman. A Common Sense View of Knowledge Graphs. Adaptive Information, Adaptive Innovation, Adaptive Infrastructure Blog, July 2019. `http://www.mkbergman.com/2244/a-common-sense-view-of-knowledge-graphs/`.

12 Andreas Blumauer and Helmut Nagy. *The Knowledge Graph Cook Book: Recipes That Work*. monochrom, 2020.

13 Kurt Bollacker, Patrick Tufts, Tomi Pierce, and Robert Cook. A platform for scalable, collaborative, structured information integration. In Ullas Nambiar and Zaiqing Nie, editors, *Intl. Workshop on Information Integration on the Web (IIWeb'07)*, 2007. URL: `https://www.aaai.org/Papers/Workshops/2007/WS-07-14/WS07-14-004.pdf`.

14 Piero Andrea Bonatti, Stefan Decker, Axel Polleres, and Valentina Presutti. Knowledge Graphs: New Directions for Knowledge Representation on the Semantic Web (Dagstuhl Seminar 18371). *Dagstuhl Reports*, 8(9):29–111, 2018. URL: `https://drops.dagstuhl.de/opus/volltexte/2019/10328/pdf/dagrep_v008_i009_p029_18371.pdf`.

15 Angela Bonifati, George H. L. Fletcher, Hannes Voigt, and Nikolay Yakovets. *Querying Graphs*. Synthesis Lectures on Data Management. Morgan & Claypool Publishers, 2018. `doi:10.2200/S00873ED1V01Y201808DTM051`.

16 Spencer Chang. Scaling Knowledge Access and Retrieval at Airbnb. AirBnB Medium Blog, September 2018. `https://medium.com/airbnb-engineering/scaling-knowledge-access-and-retrieval-at-airbnb-665b6ba21e95`.

17 Giuseppe Cota, Marilena Daquino, and Gian Luca Pozzato, editors. *Applications and Practices in Ontology Design, Extraction, and Reasoning*, volume 49 of *Studies on the Semantic Web*. IOS Press, 2020. `doi:10.3233/SSW49`.

18 Deepika Devarajan. Happy Birthday Watson Discovery. IBM Cloud Blog, December 2017. `https://www.ibm.com/blogs/bluemix/2017/12/happy-birthday-watson-discovery/`.

19 Lisa Ehrlinger and Wolfram Wöß. Towards a Definition of Knowledge Graphs. In Michael Martin, Martí Cuquet, and Erwin Folmer, editors, *Joint Proceedings of the Posters and Demos Track of the 12th International Conference on Semantic Systems - SEMANTiCS2016 and the 1st International Workshop on Semantic Change & Evolving Semantics (SuCCESS'16) co-located with the 12th International Conference on Semantic Systems (SEMANTiCS 2016), Leipzig, Germany, September 12-15, 2016*, volume 1695 of *CEUR Workshop Proceedings*. Sun SITE Central Europe (CEUR), September 2016. URL: `http://ceur-ws.org/Vol-1695/paper4.pdf`.

20 Dieter Fensel, Umutcan Simsek, Kevin Angele, Elwin Huaman, Elias Kärle, Oleksandra Panasiuk, Ioan Toma, Jürgen Umbrich, and Alexander Wahler. *Knowledge Graphs: Methodology, Tools and Selected Use Cases*. Springer, 2020. `doi:10.1007/978-3-030-37439-6`.

21 George H. L. Fletcher, Jan Hidders, and Josep Lluís Larriba-Pey, editors. *Graph Data Management: Fundamental Issues and Recent Developments*. Data-Centric Systems and Applications. Springer, 2018. `doi:10.1007/978-3-319-96193-4`.

22 José Emilio Labra Gayo, Eric Prud'hommeaux, Iovka Boneva, and Dimitris Kontokostas. *Validating RDF Data*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2017. `doi:10.2200/S00786ED1V01Y201707WBE016`.

23 Denise Gosnell and Matthias Broecheler. *The Practitioner's Guide to Graph Data*. O'Reilly Media, 2020.

24 Claudio Gutiérrez and Juan F. Sequeda. Knowledge graphs. *Commun. ACM*, 64(3):96–104, 2021. `doi:10.1145/3418294`.

25 Ferras Hamad, Isaac Liu, and Xian Xing Zhang. Food Discovery with Uber Eats: Building a Query Understanding Engine. Uber Engineering Blog, June 2018. `https://eng.uber.com/uber-eats-query-understanding/`.

26 William L. Hamilton. *Graph Representation Learning*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan & Claypool Publishers, 2020. `doi:10.2200/S01045ED1V01Y202009AIM046`.

27 William L. Hamilton, Rex Ying, and Jure Leskovec. Representation Learning on Graphs: Methods and Applications. *IEEE Data Eng. Bull.*, 40(3):52–74, 2017. URL: `http://sites.computer.org/debull/A17sept/p52.pdf`.

28 Qi He, Bee-Chung Chen, and Deepak Agarwal. Building The LinkedIn Knowledge Graph. LinkedIn Blog, October 2016. `https://engineering.linkedin.com/blog/2016/10/building-the-linkedin-knowledge-graph`.

29 Tom Heath and Christian Bizer. *Linked Data: Evolving the Web into a Global Data Space (1st Edition)*, volume 1 of *Synthesis Lectures on the Semantic Web: Theory and Technology*. Morgan & Claypool, 2011.

30 Pascal Hitzler, Markus Krötzsch, and Sebastian Rudolph. *Foundations of Semantic Web Technologies*. Chapman and Hall/CRC Press, 2010.

31 Johannes Hoffart, Fabian M. Suchanek, Klaus Berberich, Edwin Lewis-Kelham, Gerard de Melo, and Gerhard Weikum. YAGO2: Exploring and querying world knowledge in time, space, context, and many languages. In Sadagopan Srinivasan, Krithi Ramamritham, Arun Kumar, M. P. Ravindra, Elisa Bertino, and Ravi Kumar, editors, *Proceedings of the 20th International Conference on World Wide Web, WWW 2011, Hyderabad, India, March 28 - April 1, 2011 (Companion Volume)*, pages 229–232. ACM Press, March 2011.

32 Aidan Hogan. Knowledge Graphs: Research Directions. In Marco Manna and Andreas Pieris, editors, *Reasoning Web. Declarative Artificial Intelligence - 16th International Summer School 2020, Oslo, Norway, June 24-26, 2020, Tutorial Lectures*, volume 12258 of *Lecture Notes in Computer Science*, pages 223–253. Springer, 2020.

33 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, 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. *Knowledge Graphs*. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers, 2021. `doi:10.2200/S01125ED1V01Y202109DSK022`.

34 Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutiérrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab, and Antoine Zimmermann. Knowledge Graphs. *ACM Computing Surveys*, 54(4):71:1–71:37, 2021. `doi:10.1145/3447772`.

35 Valentina Janev, Damien Graux, Hajira Jabeen, and Emanuel Sallinger, editors. *Knowledge Graphs and Big Data Processing*, volume 12072 of *Lecture Notes in Computer Science*. Springer, 2020. `doi:10.1007/978-3-030-53199-7`.

36 Daniel Janke and Steffen Staab. Storing and Querying Semantic Data in the Cloud. In Claudia d'Amato and Martin Theobald, editors, *Reasoning Web. Learning, Uncertainty, Streaming, and Scalability - 14th International Summer School 2018, Esch-sur-Alzette, Luxembourg, September 22-26, 2018, Tutorial Lectures*, volume 11078 of *Lecture Notes in Computer Science*, pages 173–222. Springer, 2018. `doi:10.1007/978-3-030-00338-8`.

37 Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S. Yu. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2021. `doi:10.1109/TNNLS.2021.3070843`.

38 C. Maria Keet. *An Introduction to Ontology Engineering*. College Publications, 2018. URL: `https://open.umn.edu/opentextbooks/textbooks/590`.

39 Mayank Kejriwal. *Domain-Specific Knowledge Graph Construction*. Springer Briefs in Computer Science. Springer, 2019. `doi:10.1007/978-3-030-12375-8`.

40 Mayank Kejriwal, Craig A. Knoblock, and Pedro Szekely, editors. *Knowledge Graphs: Fundamentals, Techniques, and Applications*. The MIT Press, 2021.

41 Elisa F. Kendall and Deborah L. McGuinness. *Ontology Engineering*. Synthesis Lectures on the Semantic Web: Theory and Technology. Morgan & Claypool Publishers, 2019. `doi:10.2200/S00834ED1V01Y201802WBE018`.

42 William L. Kocay and Donald L. Kreher. *Graphs, algorithms and optimization*. Chapman&Hall/CRC Press, 2005.

43 Arun Krishnan. Making search easier: How Amazon's Product Graph is helping customers find products more easily. Amazon Blog, August 2018. `https://blog.aboutamazon.com/innovation/making-search-easier`.

44 Luís C. Lamb, Artur S. d'Avila Garcez, Marco Gori, Marcelo O. R. Prates, Pedro H. C. Avelar, and Moshe Y. Vardi. Graph Neural Networks Meet Neural-Symbolic Computing: A Survey and Perspective. In Christian Bessiere, editor, *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, pages 4877–4884. ijcai.org, 2020. `doi:10.24963/ijcai.2020/679`.

45 Jens Lehmann, Robert Isele, Max Jakob, Anja Jentzsch, Dimitris Kontokostas, Pablo N. Mendes, Sebastian Hellmann, Mohamed Morsey, Patrick van Kleef, Sören Auer, and Christian Bizer. DBpedia - A large-scale, multilingual knowledge base extracted from Wikipedia. *Semantic Web Journal*, 6(2):167–195, 2015.

46 Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In Blai Bonet and Sven Koenig, editors, *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, January 25-30, 2015, Austin, Texas, USA*, pages 2181–2187. AAAI Press, August 2015.

47 Yao Ma and Jiliang Tang. *Deep Learning on Graphs*. Cambridge University Press, 2021. URL: `https://web.njit.edu/~ym329/dlg_book/`.

48 Albert Meroño-Peñuela, Pasquale Lisena, and Carlos Martinez-Ortiz. *Web Data APIs for Knowledge Graphs: Easing Access to Semantic Data for Application Developers*. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers, 2021. `doi:10.2200/S01114ED1V01Y202107DSK021`.

49 Mark Needham and Amy E. Hodler. *Graph Algorithms*. O'Reilly Media, 2019.

50 Maximilian Nickel, Kevin Murphy, Volker Tresp, and Evgeniy Gabrilovich. A Review of Relational Machine Learning for Knowledge Graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.

51 Natasha F. Noy, Yuqing Gao, Anshu Jain, Anant Narayanan, Alan Patterson, and Jamie Taylor. Industry-scale Knowledge Graphs: Lessons and Challenges. *ACM Queue*, 17(2):20, 2019.

52 Jeff Z. Pan, Guido Vetere, José Manuél Gómez-Pérez, and Honghan Wu, editors. *Exploiting Linked Data and Knowledge Graphs in Large Organisations*. Springer, 2017. `doi:10.1007/978-3-319-45654-6`.

53 Heiko Paulheim. Knowledge graph refinement: A survey of approaches and evaluation methods. *Semantic Web Journal*, 8(3):489–508, 2017. `doi:10.3233/SW-160218`.

54 R. J. Pittman, Amit Srivastava, Sanjika Hewavitharana, Ajinkya Kale, and Saab Mansour. Cracking the Code on Conversational Commerce. eBay Blog, April 2017. `https://www.ebayinc.com/stories/news/cracking-the-code-on-conversational-commerce/`.

55 Jay Pujara, Hui Miao, Lise Getoor, and William W. Cohen. Knowledge graph identification. In Harith Alani, Lalana Kagal, Achille Fokoue, Paul T. Groth, , Josian Xavier Parreira, Lora Aroyo, Natasha Fridman Noy, Christopher A. Welty, and Krzysztof Janowicz, editors, *The Semantic Web - ISWC 2013 - 12th International Semantic Web Conference, Sydney, NSW, Australia, October 21-25, 2013, Proceedings, Part I*, volume 8218 of *Lecture Notes in Computer Science*, pages 542–557. Springer, October 2013. `doi:10.1007/978-3-642-41335-3_34`.

56 Guilin Qi, Huajun Chen, Kang Liu, Haofen Wang, Qiu Ji, and Tianxing Wu. *Knowledge Graph*. Springer, 2020. (to appear).

57 Guilin Qi, Huajun Chen, Kang Liu, Haofen Wang, Qiu Ji, and Tianxing Wu. *Knowledge Graph*. Springer Singapore, 2022.

58 Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases, 2nd Edition*. O'Reilly Media, 2015.

59 Sebastian Rudolph. Foundations of Description Logics. In Axel Polleres, Claudia d'Amato, Marcelo Arenas, Siegfried Handschuh, Paula Kroner, Sascha Ossowski, and Peter F. Patel-Schneider, editors, *Reasoning Web. Semantic Technologies for the Web of Data - 7th International Summer School 2011, Galway, Ireland, August 23-27, 2011, Tutorial Lectures*, volume 6848 of *Lecture Notes in Computer Science*, pages 76–136. Springer, August 2011.

60 Edward W. Schneider. Course Modularization Applied: The Interface System and Its Implications For Sequence Control and Data Analysis. In *Association for the Development of Instructional Systems (ADIS), Chicago, Illinois, April 1972*, 1973.

61 Juan Sequeda and Ora Lassila. *Designing and Building Enterprise Knowledge Graphs*. Synthesis Lectures on Data, Semantics, and Knowledge. Morgan & Claypool Publishers, 2021. doi:10.2200/S01105ED1V01Y202105DSK020.

62 Stephan Seufert, Patrick Ernst, Srikanta J. Bedathur, Sarath Kumar Kondreddi, Klaus Berberich, and Gerhard Weikum. Instant Espresso: Interactive Analysis of Relationships in Knowledge Graphs. In Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao, editors, *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11-15, 2016, Companion Volume*, pages 251–254. ACM Press, April 2016.

63 Saurabh Shrivastava. Bring rich knowledge of people, places, things and local businesses to your apps. Bing Blogs, July 2017. https://blogs.bing.com/search-quality-insights/2017-07/bring-rich-knowledge-of-people-places-things-and-local-businesses-to-your-apps.

64 Umutcan Simsek, Kevin Angele, Elias Kärle, Juliette Opdenplatz, Dennis Sommer, Jürgen Umbrich, and Dieter Fensel. Knowledge Graph Lifecycle: Building and Maintaining Knowledge Graphs. In David Chaves-Fraga, Anastasia Dimou, Pieter Heyvaert, Freddy Priyatna, and Juan F. Sequeda, editors, *Proceedings of the 2nd International Workshop on Knowledge Graph Construction co-located with 18th Extended Semantic Web Conference (ESWC 2021), Online, June 6, 2021*, volume 2873 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2021. URL: http://ceur-ws.org/Vol-2873/paper12.pdf.

65 Amit Singhal. Introducing the Knowledge Graph: things, not strings. Google Blog, May 2012. https://www.blog.google/products/search/introducing-knowledge-graph-things-not/.

66 Mari Carmen Suárez-Figueroa, Asunción Gómez-Pérez, Enrico Motta, and Aldo Gangemi, editors. *Ontology Engineering in a Networked World*. Springer, 2012. doi:10.1007/978-3-642-24794-1.

67 Ilaria Tiddi, Freddy Lécué, and Pascal Hitzler, editors. *Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges*, volume 47 of *Studies on the Semantic Web*. IOS Press, 2020. URL: http://ebooks.iospress.nl/volume/knowledge-graphs-for-explainable-artificial-intelligence-foundations-applications-and-challenges.

68 Denny Vrandečić and Markus Krötzsch. Wikidata: A Free Collaborative Knowledgebase. *Communications of the ACM*, 57(10):78–85, 2014.

69 Quan Wang, Zhendong Mao, Bin Wang, and Li Guo. Knowledge Graph Embedding: A Survey of Approaches and Applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743, December 2017. doi:10.1109/TKDE.2017.2754499.

70 Zhen Wang, Jianwen Zhang, Jianlin Feng, and Zheng Chen. Knowledge Graph Embedding by Translating on Hyperplanes. In Carla E. Brodley and Peter Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, July 27 -31, 2014, Québec City, Québec, Canada*, pages 1112–1119. AAAI Press, July 2014.

71 Gerhard Weikum. Knowledge Graphs 2021: A Data Odyssey. *Proc. VLDB Endow.*, 14(12):3233–3238, 2021. URL: http://www.vldb.org/pvldb/vol14/p3233-weikum.pdf.

72 Gerhard Weikum, Xin Luna Dong, Simon Razniewski, and Fabian M. Suchanek. Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases. *Found. Trends Databases*, 10(2-4):108–490, 2021. doi:10.1561/1900000064.

73 Peter T. Wood. Query languages for graph databases. *SIGMOD Rec.*, 41(1):50–60, 2012. doi:10.1145/2206869.2206879.

74 Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Networks Learn. Syst.*, 32(1):4–24, 2021. doi:10.1109/TNNLS.2020.2978386.

75 Marcin Wylot, Manfred Hauswirth, Philippe Cudré-Mauroux, and Sherif Sakr. RDF Data Storage and Query Processing Schemes: A Survey. *ACM Computing Surveys*, 51(4):84:1–84:36, 2018. doi:10.1145/3177850.

76 Da Yan, Yuanyuan Tian, and James Cheng. *Systems for Big Graph Analytics*. Springer Briefs in Computer Science. Springer, 2017. doi:10.1007/978-3-319-58217-7.

77 Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. Quality assessment for Linked Data: A Survey. *Semantic Web Journal*, 7(1):63–93, 2016.