

Knowledge Graphs

Dieter Fensel • Umutcan Şimşek • Kevin Angele •
Elwin Huaman • Elias Kärle •
Oleksandra Panasiuk • Ioan Toma •
Jürgen Umbrich • Alexander Wahler

Knowledge Graphs

Methodology, Tools and Selected Use Cases

Dieter Fensel
Semantic Technology Institute Innsbruck,
Department of Computer Science
University of Innsbruck
Innsbruck, Tirol, Austria

Umutcan Şimşek
Semantic Technology Institute Innsbruck,
Department of Computer Science
University of Innsbruck
Innsbruck, Tirol, Austria

Kevin Angele
Semantic Technology Institute Innsbruck,
Department of Computer Science,
University of Innsbruck,
Innsbruck, Tirol, Austria

Elwin Huanan
Semantic Technology Institute Innsbruck,
Department of Computer Science
University of Innsbruck
Innsbruck, Tirol, Austria

Onlim GmbH
Telfs, Tirol, Austria

Elias Kärle
Semantic Technology Institute Innsbruck,
Department of Computer Science
University of Innsbruck
Innsbruck, Tirol, Austria

Oleksandra Panasiuk
Semantic Technology Institute Innsbruck,
Department of Computer Science
University of Innsbruck
Innsbruck, Tirol, Austria

Ioan Toma
Onlim GmbH
Telfs, Tirol, Austria

Jürgen Umbrich
Onlim GmbH
Telfs, Tirol, Austria

Alexander Wahler
Onlim GmbH
Telfs, Tirol, Austria

ISBN 978-3-030-37438-9 ISBN 978-3-030-37439-6 (eBook)
<https://doi.org/10.1007/978-3-030-37439-6>

© Springer Nature Switzerland AG 2020

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

In 2005, Manuel Sahli became an active contributor to Wikipedia. He was particularly interested in his hometown, the beautiful Swiss city of Winterthur. Over the years, he edited and meticulously maintained the article on Winterthur. One of the sections he worked on was the list of sister cities.

Around the same time, a contributor added Winterthur to the list of sister cities on the article on Ontario, California, a city about an hour east of Los Angeles. Even though sister cities are supposed to be reciprocal, the information never made it to the article on Winterthur.

Manuel, although interested in all things Winterthur, did not learn about this addition for almost a decade.

2012 saw the launch of Wikidata, a sister project of Wikipedia. It was a Knowledge Graph that anyone could edit and use. As time went by, a contributor added the list of sister cities of Ontario to Wikidata, using the Ontario article on Wikipedia.

Manuel Sahli had since been elected to the cantonal parliament of Winterthur. Using Wikidata, he wanted to make a map of Winterthur's sister cities and Wikidata has a powerful query interface that allows creating ad hoc visualizations of query results over its Knowledge Graph. To Manuel's big surprise, the map contained one place he was not familiar with yet: Ontario.

At first, Manuel thought that this was an act of vandalism that has remained undiscovered for a while, but the fact came with a source: the official government website of Ontario. When he went to that site, there he found Winterthur! As a member of the cantonal parliament, he asked the Winterthur city administration and archives if they knew anything about that claim. Both denied.

Manuel wrote an official letter to the city of Ontario to clarify the situation, and lo and behold, they delivered documents, signed by both sides, establishing the sister city relation in 1982. Given these documents and the precise dates, the Winterthur city archives were now able to find the relevant documents. Thanks to Wikidata, Winterthur discovered its long lost fifth sister city, and the two cities have since rekindled their relationship.

No one was trying to hide that fact from Manuel or the world. The city archives had the documents (but could not find them at first), Wikipedia was publishing the fact for almost a decade, and it was on the official Ontario website, and yet Manuel, although deeply invested in his hometown and Wikipedia, did not learn about it. It was only when it was added to a Knowledge Graph, one that Manuel had access to and knew how to query, that the fact was surfaced.

It is far too easy to bury knowledge in documents and heaps of natural language and very hard to surface it at the right time. A Knowledge Graph makes facts easier to index, process, and find.

However, that is only half the story. The other half is that it was sheer luck that Manuel was not only interested enough to discover this fact but also capable of querying Wikidata using the SPARQL query language. Most people do not know how to do that.

We are witnessing a paradigm shift in how we use computers, as momentous as the introduction of time-sharing systems, graphical user interfaces, the web, and smartphones before: the nascent ability to interact with computers through natural language. Conversational interfaces have been around for a long time just as any novel technology has been preceded by various adoptions in niche areas. However, so-called intelligent assistants are quickly spreading and evolving, and more and more devices, from wristwatches to cars, from televisions to earbuds, are being equipped with the capability to listen to your commands and questions and to answer them.

The number of people who will gain access to knowledge through intelligent assistants for their daily tasks is growing remarkably fast. In order to find answers to burning questions, whether trivial or life-changing, we will not need the traditional computer in the office or school or library anymore. No one will need to learn how to use the newest smartphone features, how to use a mouse, or even how to read and type, thus widely opening the door to a much more inclusive world.

This new paradigm needs new ways to allow people to join the new data space, to give access to their data and their services to their prospective clients and consumers, and to allow people to discover and enjoy what they have to offer. In a world where more and more people do not use apps and websites to book a hotel room, rent a car, buy a flight ticket, order their lunch, and ask for information, we need to make sure to understand the users' intents and offer our data and services in ways that allow their integration and composition. This is not a vision of a future to come but the world we already built. Tens of thousands of capabilities and services are already integrated into the Google Assistant, Apple Siri, Amazon Alexa, the open source Mycroft, Microsoft Cortana, Samsung Bixby, and others.

Bringing together the natural language technology required to understand users' queries and the power of structuring facts in large Knowledge Graphs will allow us to bring knowledge to the end user when they need it. People will become empowered to explore and ask about the world, to follow their curiosity, to learn. Businesses will be able to expand the reach of their services and products to many new customers.

This book offers a hands-on overview of the architecture and the steps to join the world of intelligent assistants. It guides us through building, growing, and maintaining our Knowledge Graph. It discusses how Knowledge Graphs represent the facts that are important in your organization and how to make them available to other providers.

Dieter Fensel was one of the people who have very early recognized the potential of semantic technologies and the need to allow for the composition of services into novel interfaces. He and his research groups were leading the field in the semantic description of services, in linked data applications, and Ontology engineering and learning methods and tools. This book by him and the researchers of his team is distilling the experience of many years.

Enjoy going through the book! It will help you to better understand the novel field of Knowledge Graphs and how to use them. It will allow you and your organization to move to a world where the knowledge you already have in your archives and systems does not remain so stubbornly hidden and where you and your customers both will be enabled to work with the knowledge you already have, so that we can empower users, discover new sources of revenue, and never lose a sister city again.

San Francisco, CA, USA
June 2019

Denny Vrandečić

Preface

Smart speakers such as Alexa and Google Home introduced Artificial Intelligence (AI) in millions soon billions of households, making AI an everyday experience. We can now look for information, order products, and services without leaving the house or touching a computer. We just talk to a box and this thing will conveniently perform the desired tasks for us. These new communication channels define a new challenge for successful e-marketing and e-commerce. Just running a traditional website with many colorful pictures is no longer state of the art. Even the web is currently reinventing itself by applying schema.org. Data, content, and services become semantically annotated, allowing software agents, so-called bots, to search through the web understanding its content. The times where humans were browsing through large number of websites and manually extracting and interpreting their information are passing by. The users nowadays consult their personal bot to find, aggregate, and personalize information and to reserve, book, or buy products and services. In consequence, it becomes increasingly important for providers of information, products, and services to be highly hearable and visible in these new online channels to ensure their future economic sustainability. In this book, we discuss methods and tools helping to achieve these visibility goals. The core is the development and application of machine-processable (semantic) annotations of content, data, and services as well as their aggregation in large Knowledge Graphs. Only this enables bots to not only understand a question but being able to answer a question in a knowledgeable way.

These Knowledge Graphs, especially when based on schema.org, play an increasingly important role in Internet-based information search. They become a significant key technology for successful e-commerce and e-marketing and their influence on the value distribution in economic sectors that interact with their customers online. It is yet another approach for scalable data integration on a large scale and most likely it will not be the last approach tackling this very hard problem. However, it is also the first time that we approach this problem on a worldwide scale. In this book, we

describe methods and tools that empower information providers to build and maintain such a Knowledge Graph. In particular, the following aspects will be introduced:

- Methods and tools for manual, semi-automatic, and automatic construction and validation plus verification of semantic annotations and their integration into Knowledge Graphs.
- Methods and tools for the implementation of Knowledge Graphs.
- Lifecycle-based approaches for semi-automatic and automatic curation of such graphs. This includes approaches for assessment, error correction, and enrichment of knowledge with other static and dynamic resources.

Knowledge is half the way. It must be made applicable as potential answers for questions and as guidance for a dialog.

- e-marketing: with reasoning methods and tools, we can derive dialog-based bots for specific tasks and domains from a Knowledge Graph.
- e-commerce: based on the semantic descriptions of services and products, a goal-oriented dialog can be designed improving the process of reserving, renting, booking, or buying goods and services.

To illustrate the practical usage of these approaches, we discuss several pilots with a focus on e-tourism domain. Tourism is one of the largest verticals worldwide with significant growth potential. Also, it is one of the verticals where Europe may have a future, and the distribution of value critically depends on proper competence in e-marketing and e-commerce. Potential customers are distributed worldwide, and service providers are fragmented, mostly small business units (e.g., tens of thousands of small family hotels in Tyrol). In general, we focus on the following aspects:

- Integration of content, data, and service descriptions from open, proprietary, heterogeneous, and distributed sources.
- Efficient and effective maintenance of context of data (e.g., provenance, geographical and temporal validity).
- Using Knowledge Graphs for guiding dialogs.
- Integration of static and dynamic sources.
- Integration of Semantic Web services to facilitate actions and automatic service invocation.

The structure of this book follows these arguments. Chapter 1 provides a definition for Knowledge Graphs. We do not aim for mathematical precision but rather try to cover the various approaches regarding their impact. Chapter 2 details how Knowledge Graphs are built, implemented, maintained, and deployed. Chapter 3 introduces relevant application layers that can be built with such Knowledge Graphs. We explain how inference can be used to define views on such a graph, making it a useful resource for open and service-oriented dialog systems. The proof of the pudding is in the eating. Chapter 4 elaborates application of Knowledge Graph

technologies for e-tourism and use cases and pilots for other verticals. Chapter 5 provides a summary and sketches directions for future work. In Appendix A we introduce an abstract syntax and semantics for our domain specifications that are used to adapt schema.org to specific domains and tasks.

Innsbruck, Austria
Innsbruck, Austria
Innsbruck, Austria
Innsbruck, Austria
Innsbruck, Austria
Innsbruck, Austria
Telfs, Austria
Telfs, Austria
Telfs, Austria

Dieter Fensel
Umutcan Şimşek
Kevin Angele
Elwin Huaman
Elias Kärle
Oleksandra Panasiuk
Ioan Toma
Jürgen Umbrich
Alexander Wahler

Acknowledgment

We would like to thank Andreas Harth and Aidan Hogan for the very fruitful and helpful discussions and reviews of the book. Also, we would like to thank all participants of the MindLab project as the book summarizes the essence of their work.¹ Additionally, the very fruitful discussions we had with the participants of the 2019 STI Summit on Knowledge Graphs (especially with Mark Musen, Juan Sequeda, Rudi Studer, and Sung-Kook Han) helped significantly in sharpening our conclusions.² Finally, we would like to thank Andreas Lackner for acting as a pathfinder for us to find through the labyrinth of (Tyrolean) e-tourism.

¹<https://mindlab.ai/>

²<https://www.sti2.org/events/2019-sti2-semantic-summit>

Contents

1	Introduction: What Is a Knowledge Graph?	1
1.1	Introduction	1
1.2	A Conceptual Definition of Knowledge Graphs	2
1.3	An Empirical Definition of Knowledge Graphs	6
1.3.1	Open Knowledge Graphs	6
1.3.2	Proprietary Knowledge Graphs	9
2	How to Build a Knowledge Graph	11
2.1	Introduction	11
2.2	Knowledge Creation	13
2.2.1	Knowledge Creation Methodology	14
2.2.2	Our Modelling Language	15
2.2.3	Knowledge Generation Tools	20
2.2.3.1	Manual Editing	20
2.2.3.2	Semi-automatic Editing	23
2.2.3.3	Mapping	23
2.2.3.4	Automatic Annotation Tools	24
2.2.3.5	Evaluation	28
2.3	Knowledge Hosting	31
2.3.1	Collection, Storage, and Retrieval of Semantic Annotations	32
2.3.2	Collection, Storage, and Retrieval of Knowledge Graphs	34
2.4	Knowledge Curation	35
2.4.1	A Maximal Simple Knowledge Representation Formalism	36
2.4.2	Knowledge Assessment	37
2.4.2.1	Literature	37
2.4.2.2	Task Types	42
2.4.2.3	Methods and Tools	43

- 2.4.3 Knowledge Cleaning 46
 - 2.4.3.1 Literature 47
 - 2.4.3.2 Task Types 47
 - 2.4.3.3 Methods and Tools 49
 - 2.4.3.4 Summary 51
- 2.4.4 Knowledge Enrichment 52
 - 2.4.4.1 Literature 53
 - 2.4.4.2 Task Types 55
 - 2.4.4.3 Methods and Tools 55
 - 2.4.4.4 Summary 61
- 2.4.5 Summary on Knowledge Curation 61
- 2.5 Knowledge Deployment: The Use of the Pudding
Is in the Eating 62
- 3 How to Use a Knowledge Graph 69**
 - 3.1 Introduction 69
 - 3.2 Merging Artificial Intelligence and the Internet 70
 - 3.2.1 60 Years of AI in a Glimpse 70
 - 3.2.2 The Web (for Bots) 71
 - 3.2.3 Summary 78
 - 3.3 Knowledge Access Layer 78
 - 3.3.1 Loosely Connected TBoxes Defining Logic-Based
Views on Knowledge Graphs 79
 - 3.3.2 Dynamic and Active Data: Semantic Web Services 83
 - 3.4 Open and Service-Oriented Dialog Systems 85
 - 3.4.1 Open Dialog Systems 85
 - 3.4.2 Service-Guided Dialogs 91
 - 3.4.3 Summary 92
- 4 Why We Need Knowledge Graphs: Applications 95**
 - 4.1 Introduction 95
 - 4.2 The Market 96
 - 4.3 Motivation and Solution 97
 - 4.4 Touristic Use Cases 101
 - 4.5 Energy Use Cases 107
 - 4.6 Further Verticals 110
 - 4.7 Summary 111
- 5 Conclusions 113**
- Appendix 117**
- References 129**
- Index 145**