



Knowledge Graphs

A Tutorial on the History of Knowledge Graph's Main Ideas

Claudio Gutierrez
DCC, Universidad de Chile and IMFD

Juan F. Sequeda
data.world

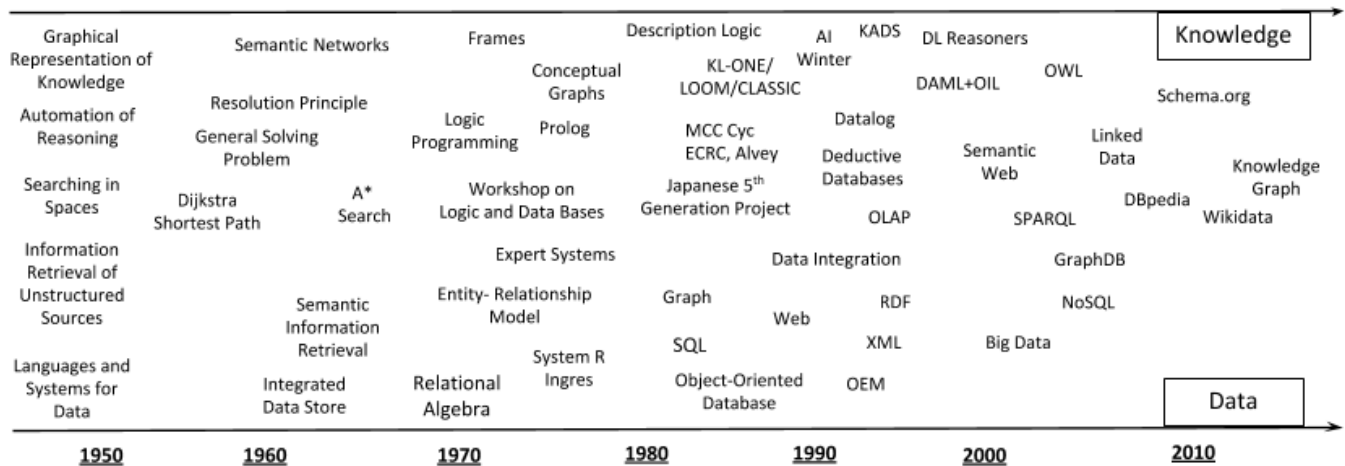


Figure 1: History of Knowledge Graphs

ABSTRACT

Knowledge Graphs can be considered as fulfilling an early vision in Computer Science of creating intelligent systems that integrate knowledge and data at large scale. Stemming from scientific advancements in research areas of Semantic Web, Databases, Knowledge representation, NLP, Machine Learning, among others, Knowledge Graphs have rapidly gained popularity in academia and industry in the past years. The integration of such disparate disciplines and techniques give the richness to Knowledge Graphs, but also present the challenge to practitioners and theoreticians to know how current advances develop from early techniques in order, on one hand, take full advantage of them, and on the other, avoid reinventing the wheel. This tutorial will provide a historical context on the roots of Knowledge Graphs grounded in the advancements of Logic, Data and the combination thereof.

CCS CONCEPTS

• **Information systems** → **Data management systems**; • **Computing methodologies** → **Knowledge representation and reasoning**.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CIKM '20, October 19–23, 2020, Virtual Event, Ireland

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6859-9/20/10.

<https://doi.org/10.1145/3340531.3412176>

KEYWORDS

Knowledge Graph, Knowledge, Logic, Databases, Semantic Web

ACM Reference Format:

Claudio Gutierrez and Juan F. Sequeda. 2020. Knowledge Graphs A Tutorial on the History of Knowledge Graph's Main Ideas. In *Proceedings of the 29th ACM International Conference on Information and Knowledge Management (CIKM '20), October 19–23, 2020, Virtual Event, Ireland*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3340531.3412176>

1 INTRODUCTION

“Those who cannot remember the past are condemned to repeat it” - George Santayana

Knowledge Graphs can be considered to be fulfilling an early vision in Computing of creating intelligent systems that integrate knowledge and data at large scale. The term *Knowledge Graph* was introduced by researchers at the turn of this century and has rapidly gained popularity in academia and industry since it was popularized in the 2010's decade. It is paramount to note that, regardless of the discussions on, and definitions of, the term “Knowledge Graph”, it stems from scientific advancements in diverse research areas such as Semantic Web, Databases, Knowledge Representation and Reasoning, NLP, Machine Learning, among others. The integration of ideas and techniques from such disparate disciplines presents a challenge to practitioners and researchers to know how current advances develop from, and are rooted in, early techniques.

2 MOTIVATION AND TARGET AUDIENCE

Understanding the historical context and background of one's research area is of utmost importance. An early step in the scientific

method is to conduct background research in order to stand on the shoulders of giants. When it comes to the Semantic Web research area, and in particular for one of its most promising developments, Knowledge Graphs, we have observed that students and junior researchers are not very well aware of the history of this research area.

In the talk “Integrating Semantic Web in the Real World: A journey between two cities” by Juan Sequeda¹, he starts out providing a brief historical overview of Logic and Data and asks the audience to raise their hand if they are aware of the Japanese 5th Generation Project. Throughout the fifteen times that this talk has been given, each time, only a few hands are raised, which are usually from senior researchers attending. This recurring event is what has motivated this tutorial.

Claudio, with similar motivations, presented in several places a short scheme of the history of Knowledge Graphs (“A concise account of the notion of Knowledge Graph”) that sparked unpredictable interest from young researchers.

These recurring events are what motivated this paper and tutorial.

The primary audience would be students and junior researchers. The secondary expected audience would be senior researchers who would want to contribute and complement the content of the tutorial.

3 OUTLINE OF THE TUTORIAL

The goal of this tutorial is to provide a high level overview of the key ideas, theories and events that have occurred over the past 50 years which has led to the development of the Semantic Web and Knowledge Graphs.

It is important to acknowledge that this is not a survey. We will not dive into details but present a map and guidelines to navigate through the most important ideas, theories and events that have signaled and triggered current development, in order to understand what worked, what did not work and reflecting how it inspired the next ideas.

The tutorial will be divided by decades. The material will be presented with respect to three contexts: 1) Knowledge 2) Data and 3) combination of Knowledge and Data. A proposed schedule is the following:

- Introduction
- I. Advent of the digital age (1950s and 1960s)
- II. Data and Knowledge Foundations (1970s)
- III. Managing Data and Knowledge (1980s)
- IV. Data, Knowledge and the Web (1990s)
- V. Data and Knowledge at Large Scale (2000s)
- Where are we now? Let’s discuss

The tutorial will be given in a lecture style.

The attendees of this tutorial will leave with a general historical context of Logic and Data, how it has been combined leading to the advent of the Semantic Web and Knowledge Graphs. Furthermore, the attendees will leave with homework, namely a list of seminal papers that they should consider reading.

There is no expected prior knowledge, except for a desire to learn and appreciate the history and background of their discipline.

This tutorial will be accompanied by an upcoming review article in the Communications of the ACM. A pre-print can be found: <http://knowledgegraph.today/paper.html>

4 PRESENTERS

Claudio Gutierrez is full professor at the Computer Science Department, Universidad de Chile and Senior Research at the Millennium Institute for Foundation of Data. His research experiences lies in the intersection of Databases and the Semantic Web, focusing in data models and query languages for RDF layer, particularly RDF and SPARQL. Currently is interested in Open Data, Linked Data and Foundations of Data. He has published extensively in the area, edited three books and received several best paper awards at Semantic Web Conferences. Was awarded the SWSA Ten-Year Award from the International Semantic Web Conference, 2016; the Test of Time Award from the Principles of Database Systems, ACM-PODS 2014, and was International Scholar 2015-2016 for the Society for the History of Technology (SHOT). He is involved in both, Database and SW communities, where has been in PC committees of ICDT, PODS, WWW, ESWC, ISWC, RR, and several workshops; currently is in the editorial board of the JWS.

Juan Sequeda is the Principal Scientist at data.world. He joined through the acquisition of Capsenta, a company he founded as a spin-off from his research. He holds a PhD in Computer Science from The University of Texas at Austin. His research goal is to reliably create knowledge from inscrutable data. His research interests are at the intersection of Logic and Data for (ontology-based) data integration and semantic/graph data management and Knowledge Graphs. Juan is the recipient of the NSF Graduate Research Fellowship, received 2nd Place in the 2013 Semantic Web Challenge for his work on ConstituteProject.org, Best Student Research Paper at the ISWC 2014 and the 2015 Best Transfer and Innovation Project awarded by the Institute for Applied Informatics. Juan is on the Editorial Board of the Journal of Web Semantics, member of multiple program committees (ISWC, ESWC, WWW, AAAI, IJCAI). He has served as a bridge between academia and industry as the current chair of the Property Graph Schema Working Group, member of the Graph Query Languages task force of the Linked Data Benchmark Council (LDBC) and past invited expert member and standards editor at the World Wide Web Consortium (W3C).

¹<http://www.juansequeda.com/blog/2019/01/07/my-most-memorable-event-of-2018/>