An Introduction to Knowledge Graphs

Lateef Muizz Kolapo

May 14, 2021

Abstract

Understanding how to work with knowledge graphs can give data scientists the ability to not just extract interrelated facts and assumptions from massive collections of data, but can also help in understanding how to form contextual connections and understanding from data via linking and semantic metadata which helps provide a unified approach to data analytics and intelligence. This report is an introduction to knowledge graphs, important concepts and their applications.

1 Introduction

Knowledge graphs are used to map collection of data from different sources and create a connection between the different entities in a given subject matter such as people, place or things to provide meaning to the data and remove any Semantic ambiguity (Ristoski and Paulheim, 2016). Graph technologies and analytics offers tools for a system driven by connections—from social connections to financial connections, it helps move data analytics and intelligence from just tabulating and retrieving data to forming real world meanings and relationships (Needham and Hodler, 2019). Knowledge graphs can be considered as a type of semantic network which forms the basis of associations between concepts and entities in a network— connections between real world objects, events, situations or abstract concepts ('Semantic network', 2021). Although they are used quite interchangeably, knowledge base and knowledge graphs are not the same with the fact that information in knowledge graphs are organized using graphs which is not true for all knowledge bases. There have been different versions of knowledge graphs proposed by different organizations with each version having varying levels of complexity such as Google's Knowledge Graph, Knowledge Vault, Microsoft's Satori, Facebook's Entities Graph, etc. In 2012 Google introduced its knowledge graph called "Google Knowledge Graph" harnessing approaches that have been around since the beginning of modern AI to enable its users discover new information quickly and relevant to their queries (Singhal, 2012).

2 Resource Description Framework and Ontologies

Knowledge graphs is a combination of different paradigms, a database to provide a way to explore the data through structured queries, a graph to provide a network data structure, a knowledge base to interpret the data and infer new facts and an ontology to serve as a formal semantic data models used to define the types of objects that exists within the subject matter and the properties used to describe them (Ontotext, no date).

2.1 Resource Description Framework

The Resource Description Framework, is a graph data schema used for describing the semantics, or meaning of information in the form of triples which are based on an Entity Attribute Value (EAV) model, a combination of a Resource, a Property, and a Property value (subject, predicate, object)—forming a Statement used to represent complex relationships between different resources (Wu et al., 2014). This means RDF gives us the opportunity to infer from data by discovering new facts about the data based on a set of rules and due to the semantic nature of RDF we do not need to have explicit relationships store which means RDF can generate new facts from existing triples. In Table 1 the first triple, "Muizz Studies Machine Learning Engineering", Muizz is the subject, Studies is the predicate, and Machine Learning is the object.

It is designed to provide understandable bit of information to computer applications with precreated joins compared to rdbms as seen in the example.

Table 1: Example of a triple.

Subject	Predicate	Object
:Muizz	:Studies	Machine Learning engineering
:Tunde	:Studies	Big Data engineering

The rdf schema in Figure 1 is everything connected to the rdf:type data science, we see that the entities pointing to data science is a topic under data science and we see that project management although have similar characteristics to Data Science-take courses in project management they are not part of the Data science rdf:type and as such are not data scientists. This makes it easy to see connections between entities without having to create complex joins since we know any entity that points to data science is a part of the class. RDF provide interoperability through a uniform structure that supports the reuse of existing standards and ontologies. They use standards which are unique across different systems which facilitates data integration and publishing by enabling different agents, services and applications to share knowledge and intelligence (Noy et al., 2005).

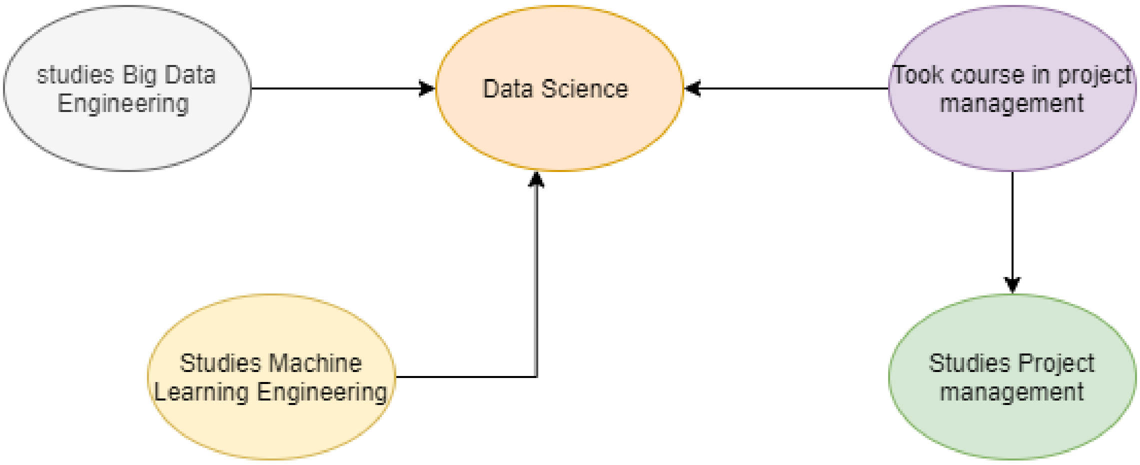

Figure 1: A Sample RDF schema.

2.2 Ontologies

In computer science ontology is how we formally create conventions for naming and definition of the categories, properties and relations between domain of discourse ('Ontology (information science)', 2021). Ontology is different from taxonomy which is a controlled set of vocabulary used to make it easier to find, manage and present related information—they express hierarchical relationships within a category (SCHWEIZER, no date). Ontology is a type of knowledge base which helps build relationship between people, place and things i.e classes, entities and properties,enable users to link multiple concepts to other concepts in a variety of ways. Ontological representations allows semantic modeling of knowledge, this means it adds a basic meaning to the data and the relationships that lie between them and are used to form the knowledge bases in artificial intelligence (AI) applications (SCHWEIZER, no date). Knowledge graphs can be described as a very large ontology with additional features such as data collection, extraction, integration of data from external sources which extends a knowledge-based system and a reasoning engine (Ehrlinger and W oß, 2016). Languages such as Web Ontology Language (OWL) are used to express ontologies, they are designed to represent extensive knowledge about objects and the relations between them by specifying object classes and relationship properties as well as taxonomies (Ontotext, no date).

3 Usability of knowledge graphs

We will be exploring use of knowledge graphs for analytics and business intelligence. Telecommunication is all about connections—interchanging information between people, satellites or computers, this makes knowledge graphs a perfect candidate for modeling, storing and querying telecommunications data of all kinds. Telecommunication system have architectures combining various complex network structures and systems, they provide a diverse range of product offers and bundles and are highly invested in customer satisfaction and churn rate in a highly competitive landscape, implementing a knowledge graph would not only provide an efficient means of monitoring the network systems itself but also enable the business to be more agile in dealing with their customer needs and requirements.

Consider a large telco with millions of customers, agents and partners. They use monitoring systems to monitor customer usage, key performance indicators, application operation and Generation of reports and event notifications (SLAC,2021). Such organization would typically have analytics questions such as:

Table 2: Sample Analytics Questions faced (Stanford CS 520, no date)

Knowledge graph	Analytics Question	Knowledge graph Solution
Customer Intelligence KGS	How do we use data related to the customer to understand customers better in order to create personalized experiences for our customers?	To answer this question we need data about the customers both internal and external. Internal data are usually available in traditional databases with different information about the customers segmented into tables and schemas. External data can be found in hubs for Linked Open Data (LOD) such as FactForge, Google Knowledge Graph, Geonames e.t.c. Combining data from multiple sources with an efficient schema and ontological design helps us create a customer intelligence knowledge graph. This type of knowledge graph design reinforces embedded analytics and customer 360 practices which gives the telco a competitive advantage.

Knowledge graphs reduce cost of the expensive joins used in traditional databases as it requires lesser hops to answer analytics questions[21]. In traditional business intelligence platforms finding connections between different relational databases requires time-intensive data modeling and query operations. With each new business question that arises we would require new dataset and schema which is not a sustainable approach to intelligence and knowledge (Clark, 2020). Implementing a customer intelligence knowledge graph which focuses on the most customer relevant entities, such as most common among its users: bundle types, complaints, data usage and subscription rates would give a competitive advantae to the business. Focusing on those domains that have the greatest chance of delivering the best and delightful user experiences would help improving analytics and intelligence in such organizations (Kempe, 2014).

4 Microsoft Academic Graph

The Microsoft academic graph is a knowledge graph implementation of academic information and data—it has a collection of entities such as people, publications, fields of study, conferences, and locations. It provides connections between researchers and researches related to them which might have been difficult to determine (Noy et al., 2019). Microsoft Academic Knowledge Graph (MAKG) is a large RDF data set with over eight billion triples of information about scientific publications and related entities, such as authors, institutions, journals, and fields of study. MAKG uses data licensed under Open Data Commons Attribution License (ODC-By). This knowledge graph is able to bring together large volumes of schorlarly publications and data to the web by using RDF dump files, data source from the Linked Open Data cloud with URIs that works and connects to multiple data sources (Farber, 2006). Microsoft academic graph is central around the acquisition of knowledge from text through the identification of lexical constructs of semantic objects which represents entities or connections. MAKG uses an hierarchical entity type called concepts to represent the semantic contents of a document, it is an abstract way to define entities such as authors and affiliations and as such no concrete way of defining them (Kuansan et al., 2019).

References

[1] Clark, K. (2020) *What is a Knowledge Graph?*. Available at:https://www.stardog.com/blog/what-is-a-knowledge-graph/ (Accessed: 11 May 2021).

[2] Ehrlinger, L and Wöß, W. "*Towards a Definition of Knowledge Graphs.*" SEMANTiCS (2016).

[3] Farber, M. (2006) *The Microsoft Academic Knowledge Graph: A Linked Data Source with 8 Billion Triples of Scholarly Data.* Institute AIFB, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany.

[4] Feilmayr, C and Wöß, W. "*An analysis of ontologies and their success factors for application to business Data Knowledge Engineering.*" SEMANTiCS (2016).

[5] Kempe, S. (2014) *Why Your Business Needs A Customer Data Knowledge Graph?*. Available at:https://www.dataversity.net/business-needs-customer-data-knowledge-graph/ (Accessed: 11 May 2021).

[6] Kuansan, W., Zhihong, S., Chiyuan, H., Chieh-Han, W., Darrin, E., Yuxiao, D., Junjie, Q., Anshul, K., Alvin, C., and Richard, R.(2019) *'A Review of Microsoft Academic Services for Science of Science Studies'*, Journal of Frontiers in Big Data, (2), pp.45. Available at: https://www.frontiersin.org/article/10.3389/fdata.2019.00045 . (Accessed: 11 May 2021).

[7] Noy, N., Gao, Y,, Jain, A,, Narayanan, A,, Patterson, A, and Taylor, J, (2019) *Industry-Scale Knowledge Graphs: Lessons and Challenges.* Available at: https://cacm.acm.org/magazines/2019/8/238342-industry-scale-knowledge-graphs/ (Accessed: 10 May 2021).

[8] Noy, N., McGuinness, D., and Hayes, P.J. (2005) '*Semantic Integration Interoperability Using RDF and OWL* ', W3C Journals. Available at: https://www.w3.org/2001/sw/BestPractices/OEP/SemInt/ (Accessed: 11 May 2021).

[9] *'Ontology' (2021).* Wikipedia. Available at https://en.wikipedia.org/wiki/Ontology. (Accessed: 11 May 2021).

[10] *'Ontology (information science)' (2021).* Wikipedia. Available at https://en.wikipedia.org/wiki/Ontology_(information_science). (Accessed: 11 May 2021).

[11] Ontotext (no date) Available at: https://www.ontotext.com/knowledgehub/fundamentals/what-is-a-knowledge-graph/ (Accessed: 11 May 2021).

[12] Ontotext (no date) Available at: https://www.ontotext.com/knowledgehub/fundamentals/what-are-ontologies/ (Accessed: 11 May 2021). (Ontotext, no date)

[13] Needham, M. and Hodler, E. A. (2019) *Graph Algorithms*: O'Reilly Media, Inc.

[14] Ristoski, P. and Paulheim, H.(2016) '*Semantic Web in data mining and knowledge discovery*: A comprehensive survey, Journal of Web Semantics', Journal of Web Semantics, 4(2), pp.1-22. Available at: https://doi.org/10.1016/j.websem.2016.01.001. (Accessed: 11 May 2021).

[15] SCHWEIZER, C. (no date) *What is the Difference between Taxonomy and Ontology? It is a Matter of Complexity.* Available at: https://www.earley.com/blog/what-difference-between-taxonomy-and-ontology-it-matter-complexity (Accessed: 11 May 2021).

[16] 'Semantic network' (2021). Wikipedia. Available at https://en.wikipedia.org/wiki/Semantic_network. (Accessed: 11 May 2021).

[17] SLAC National Accelerator Laboratory (2021) Available at: https://www.slac.stanford.edu/xorg/nmtf/nmtf-tools.html (Accessed: 11 May 2021).

[18] Singhal, A. (2012) *Introducing the Knowledge Graph: things, not strings.* Available at: https://blog.google/products/search/introducing-knowledge-graph-things-not/ (Accessed: 11 May 2021).

[19] Stanford CS 520 (no date) Available at: https://web.stanford.edu/class/cs520/2020/notes/What_Are_Some_High_Value_Use_Cases_Of_Knowledge_Graphs.html (Accessed: 11 May 2021).

[20] Wu, B., Zhou, Y., Yuan, P., Jin, H., and O'Hara, J. (2014) 'SemStore: *A Semantic-Preserving Distributed RDF Triple Store.*', In Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14). Association for Computing Machinery, New York, NY, USA, Available at:https://doi.org/10.1145/2661829.2661876 (Accessed: 11 May 2021).