WHAT IS A KNOWLEDGE GRAPH?

MARCH 31 2020
NOTES - EXTRACTS OF EMAILS

Contributions and Comments from Ontology Summit 2020 Members (From Speakers will be added Next Week)

Gary Berg Cross, Todd Schneider, Ravi Sharma, Ken Baclawski, Janet Singer, John Sowa, others …..

Scope:
View KG as a whole?
View elements of KG?
Theories of what is KG?
Minimum essentials of KG?
Ontologies and KG?
etc.

Currently (circa 2020) there are three types of persistence mechanisms that are at times labeled 'Knowledge Graphs'. These include 'graph database', 'property graph', and 'knowledge graph'.

A Graph Database is one that uses graph structures for semantic queries with nodes, edges, and properties to represent and store data. The graph model explicitly lays out the dependencies between nodes of data, the relational model and other NoSQL database models link the data by implicit connections. A graph database is a database designed to treat the relationships between data as equally important to the data itself.

A Property Graph is a key/value-based, directed, multi-relational graph that is a collection of interlinked descriptions of entities. The descriptions have a structure that allows both people and computers to process them and the entity descriptions contribute to one another, forming a network, where each entity represents part of the description of the entities, related to it.

Knowledge Graphs combine characteristics of several data management paradigms. The Knowledge Graph be a specific type of database, because it can be queried via structured queries. A graph, because it can be analyzed as any other network data structure and knowledge base, because the data in it bears formal semantics, which can be used to interpret the data and infer new facts.

There are many definitions but here is one way to distinguish property graphs from knowledge graphs:

1) A property graph is a pattern. It has a graph for each property. Attributes are properties with literals (e.g., strings) as values.
2) An ontology is also a pattern.
3) A knowledge graph is a pattern morphism from an ontology to a property graph

Knowledge Graph Category
● Let O be a pattern that will be used as an
ontology.
● The knowledge graph category for O has:
● Objects are pattern morphisms: O → A
● A morphism from: O → A to: O → B is a morphism: A → B such that this diagram commutes: O
A B

Knowledge graph is or maybe:
0) It's a model of a portion of some domain,
1) Is allows more fidelity of the representation of a domain via relations (the relations are the key),
2) The additional fidelity of representation (via relations) allows more useful and actionable distinctions.
3) A KG is graph-based, flexible formal data/knowledge representation that can easily map to other data formats using generic tools, and pipelines for updates. It is now a De Facto and popular format for representing knowledge, supporting many applications to semantic search engines, question answering systems, and recommender systems.

Consider the label 'Knowledge Graph' and the two terms used to distinguish it. 'Knowledge' and 'Graph'.

First consider the notion of 'Graph'. This is a well-defined entity in Mathematics. In fact, there's an entire field devoted to it, 'Graph Theory'.
Second consider the qualifier in the label, 'Knowledge'. This is a less well-defined notion. Among the common (English) definitions that can be found are the following:

- facts, information, and skills acquired by a person through experience or education, the theoretical or practical understanding of a subject
- what is known in a particular-field, or in total: facts and information

Let's leave out epistemological considerations for the moment and focus on the bits about "information" and "in a particular domain".
So, we may consider that a 'knowledge graph' is information about 'things' in a particular-domain, using graph constructs to represent them. If we use a simple definition of a graph as a set of nodes and arcs that 'connect' the nodes, then the common interpretation of a knowledge graph is that the nodes correspond to sets of things that are similar and the arcs are relations among the nodes, or relations among dissimilar 'things'.

Having only a set of nodes that represent collections or sets of things in the domain that are similar does represent some portion of knowledge about the domain: what's similar and what is

not similar. However, that's not much 'knowledge'. It may be useful for some applications, but certainly not many.

So, what else is missing to have 'knowledge'? The answer is the relations among those collections of similar entities.
For instance, what makes 'things' similar? Do they have the same mass? The same color? What else might they have in common? Each of these criteria can be expressed as a relation (of some sort, binary, ternary, or in general n-ary). Having the combination of a collection of similar 'things' and the relations among them is knowledge about the domain (from which they are chosen).
This raises comments and questions, listed are either uses or requirements on what needs to be represented.

- We can move further and not only limit ourselves to similarity and differences among entities (nodes) in a graph but also construct complex or many types of correlations among entities as well.
- Also, we can seek dependencies and hierarchies and frequencies of interest in entities for example how many A type Entities are related to B type and how often in this domain of interest, for example, in these Graphs.
- Building on these notions, we can begin to ask such variety of questions relating to graph - set or collection of graphs with common nodes, etc., almost disconnected graphs, sparse, dense, singleton nodes, etc. etc.
- We can then call this set of collections and questions and queries made on graphs and generalize the domain and set of Questions and Answers as a “knowledge” about the domain.
- We recognize that often we are limiting to graphs constructed using triples.
- Are there other examples of non-triple graphs, and would they be relevant to ontologies and to the Ontology Summit 2020 topic?
- Can we construct a prototype or a demo knowledge graph for a (small) sub-domain or just a topic for illustration?
- What would be the role of Data (info sources) and data models, CDM and LDM, and in implementation mode, PDMs? Will these be called Ontologies related to Knowledge Graphs?
- Role of vocabularies, glossaries in construction of KGs?

Can queries around “graphs” enrich them to become “knowledge graphs”? An example could be binary logic of semiconductor gates that would help us build logical and executable circuits and correct the design errors?

If we can understand “graphs” (nodes and links) by examining the things and relationships by some domain related query, logic, criteria and apply rules to stay within a topic or subdomain, can we convert that graph into a “knowledge graph” and in a narrow sense one means “knowledge” to be that collection of graph elements identified using these queries and other reasoning techniques used, to collect useful information or knowledge of that domain. In the end it would be a filtered graph by above criteria called KG.

Staying at elementary level to simply show (perhaps an example later as we develop) that Knowledge Graph is a modification of that general graph with which we started, that addresses Knowledge about a domain or topic.

Other comments:

How the 'things' in a knowledge graph are interpreted.
To what extent are the possible interpretations constrained by the representation (i.e. in OWL using Object and Data property constraints) or by humans applying their own interpretation(s) of the (usually) natural language terms used to create identifiers (for the things in the knowledge graph).
This latter item (interpretation) is what could be used to distinguish a 'knowledge graph' from a 'property graph'.
Another distinction among these two is the ability to reason. In a property graph the additional 'attributes' that can be assigned to the relations (aka properties) are extra-logical entities.

It is too early but perhaps the summit should recommend (Include?) as part of a definition of a knowledge graph:
a) allow extra-ontological entities (e.g. as in property graphs and the ability to have arbitrary 'attributes' associated with properties)
b) the ability to effectively use a fragment of First Order Logic reasoning on the representation.

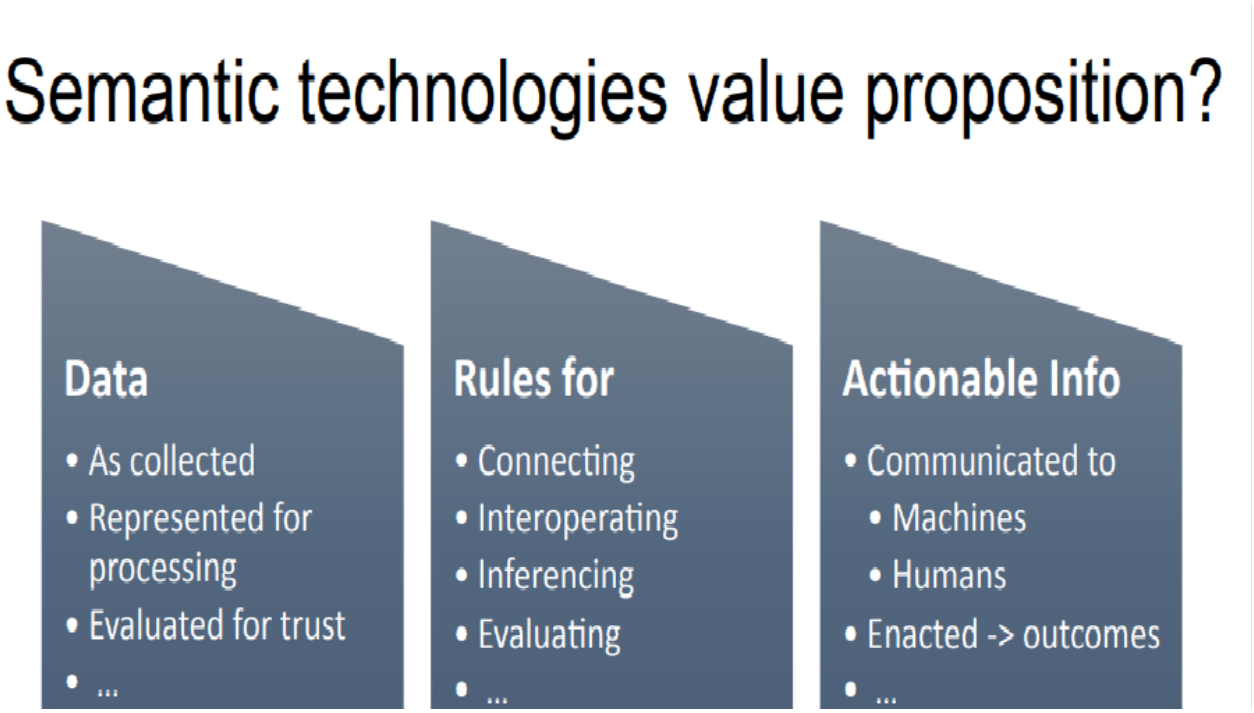
Appendix:

"A knowledge graph (i) mainly describes real world entities and their interrelations, organized in a graph, (ii) defines possible classes and relations of entities in a schema, (iii) allows for potentially interrelating arbitrary entities with each other and (iv) covers various topical domains."	H. Paulheim. Knowledge Graph Refinement: A Survey of Approaches and Evaluation Methods. Semantic Web Journal, (Preprint):1–20, 2016.
"Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities."	M. Kroetsch and G. Weikum. Journal of Web Semantics: Special Issue on Knowledge Graphs. http://www.websemanticsjournal.org/index.php/ps/announcement/view/19 [August, 2016].
"Knowledge graphs could be envisaged as a network of all kind things which are relevant to a specific domain or to an organization. They are not limited to abstract concepts and relations but can also contain instances of things like documents and datasets."	A. Blumauer. From Taxonomies over Ontologies to Knowledge Graphs, July 2014. https://blog.semantic-web.at/2014/07/15/from-taxonomies-over-ontologies- to-knowledge-graphs [August, 2016].
"We define a Knowledge Graph as an RDF graph. An RDF graph consists of a set of RDF triples where each RDF triple (s, p, o) is an ordered set of the following RDF terms: a subjects∈UUB,a predicate p∈U,and an object UUBUL. An RDF term is either a URI u ∈ U, a blank node b ∈ B, or a literal l ∈ L."	M. Färber, B. Ell, C. Menne, A. Rettinger, and F. Bartscherer. Linked Data Quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web Journal, 2016. http://www.semantic-web-journal.net/content/linked-data-quality-dbpedia-freebase- opencyc-wikidata-and-yago [August, 2016] (revised version, under review)
"[...] systems exist, [...], which use a variety of techniques to extract new knowledge, in the form of facts, from the web. These facts are interrelated, and hence, recently this extracted knowledge has been referred to as a knowledge graph."	J. Pujara, H. Miao, L. Getoor, and W. Cohen. Knowledge Graph Identification. In Proceedings of the 12th International Semantic Web Conference - Part I, ISWC '13, pages 542–557, New York, USA, 2013.

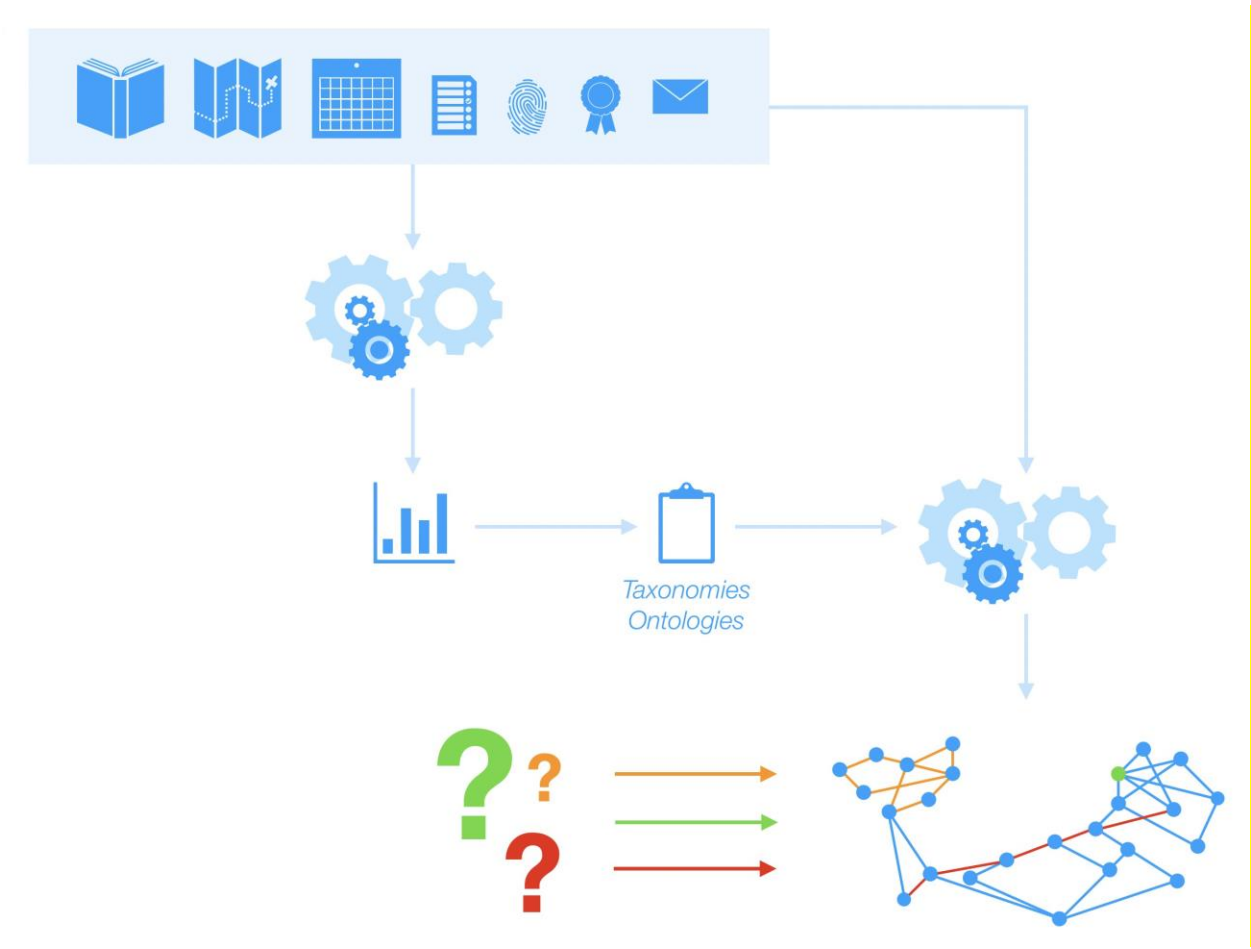
References

P. Mika, T. Tudorache, A. Bernstein, C. Welty, C. A. Knoblock, D. Vrande˘ci´c, P. T. Groth, N. F. Noy, K. Janowicz, and C. A. Goble, editors. The Semantic Web - ISWC 2014 - 13th International Semantic Web Conference, Riva del Garda, Italy, October 19-23, 2014. Proceedings, Part I, volume 8796 of Lecture Notes in Computer Science, Springer, 2014.

Visualizations related to Knowledge Graphs From earlier discussions:



Visualization of Knowledge Graph
From Anirudh Prabhu (Keck) talk and
https://towardsdatascience.com/knowledge-graphs-and-machine-learning-3939b504c7bc



Knowledge Graphs Visualization attempt -Ravi

Pool Party - white paper or advertisement?

https://www.poolparty.biz/what-is-a-knowledge-graph

What are Knowledge Graphs?

A Knowledge Graph is a model of a knowledge domain created by subject-matter experts with the help of intelligent machine learning algorithms. It provides a structure and common interface for all of your data and enables the creation of smart multilateral relations throughout your databases. Structured as an additional virtual data layer, the Knowledge Graph lies on top of your existing databases or data sets to link all your data together at scale – be it structured or unstructured. The fluidity of the structure also allows for your Knowledge Graph to grow organically each time new data is introduced. The more relations created, the more context your data has – allowing you to get a bigger picture of the whole situation and helping you to make informed decisions with connections you may have never found.

Conclusions

With the never-ending snowball effect of accumulated data, organizing your data without a strategy means not being able to stay competitive and relevant in the digital age. Large digital corporations are disrupting the market by making sure that their data works for them, rather than the other way around. Ensuring that your Enterprise Knowledge Graph is created and maintained in-house will ensure risk mitigation and transparency for both you and your clients, avoiding the dreaded 'black box' effect. Implementing a Knowledge Graph in combination with AI and Machine Learning algorithms will help put context and rationale in your data. This will enable you to not only discover deeper and more subtle patterns but also to make decisions smarter and faster.
Knowledge is every enterprises' most valuable asset – so, why isn't every enterprise prioritizing it?

TO DO:
SPEAKERS EXAMPLES

Look at their slides that define KG
Quote them

Look at chat and their or others What is KG related comments

Make a pictorial like Janet Graph what how - on what

List in the communique draft different views of KG by orientation, expertise, math, graphic examples of use such as category theory
Domain sub domain
Source target database etc.
Create viewpoints