# Generations of Knowledge Graphs: The Crazy Ideas and The Business

##### Keynote Speech

### Abstract

Knowledge Graphs (KGs) have been used to support a wide range of applications, from web search to personal assistant. In this talk, we describe three generations of knowledge graphs: entity-based KGs, which have been supporting general search and question answering (e.g., at Google and Bing); text-rich KGs, which have been supporting search and recommendations for products, bio-informatics, etc. (e.g., at Amazon and Alibaba); and the emerging integration of KGs and LLMs, which we call dual neural KGs. We describe the characteristics of each generation of KGs, the crazy ideas behind the scenes in constructing such KGs, and the techniques developed over time to enable industry impact. In addition, we use KGs as examples to demonstrate a recipe to evolve research ideas from innovations to production practice, and then to the next level of innovations, to advance both science and business.

*"Science is to test crazy ideas; engineering is to bring these ideas into business." – Andreas Holzinger*

### Introduction

Since the birth of modern Knowledge Graphs (KGs) around 2007 (in the same year, Yago [4], DBPedia [1], and Freebase [2] were released)[[1]](#footnote-0) , the area has been broadly researched in a multitude of research communities (to name a few, NLP, IR, Data Mining, Databases, Semantic Web). The industry deployment started about a decade ago, when Google launched Knowledge Panels in web search in 2012; since then, KGs have been used broadly to support web search (e.g., Google and Bing web search), voice assistants (e.g., Amazon Alexa, Apple Siri, and Google Assistant), and so on, and have made profound business impact.

KGs model the real world in a graph representation, where nodes represent real-world entities or atomic (attribute) values, and edges represent relations between the entities or attributes between entities and atomic values. A piece of knowledge can be considered as a triple in the form of (subject, predicate, object), such as (Seattle, located\_at, USA). The data instances in a KG follow the ontology as the schema, which in itself is represented in a graph form and can be taken as a part of the KG. The ontology describes entity classes, often organized in a hierarchical structure and also called taxonomy, and meaningful relationships between classes.

KGs can be considered as semi-structured: on the one hand, it enjoys clean semantics of structured data powered by the rigidity of schemas (i.e., ontology); on the other hand, it embraces the flexibility of unstructured data by allowing easily adding new classes and relationships. An additional advantage of KGs is that it can seamlessly connect a large number of domains through common entities across domains or relationships between domains (e.g., the Movie and Music domains can be connected by people who are both actors/actresses and singers, and by the featured\_song relation). These advantages give KGs a unique position that is both understandable to machines (through ontology) and easy-to-understand by human beings (blessed by the structure), suitable to facilitate understanding in search, question answering (QA), and dialogs, to power recommendation through the graph structure, and to display information for human understanding (in attribute-value pairs), comparison (in tables), and explanation (in paths in the graph).

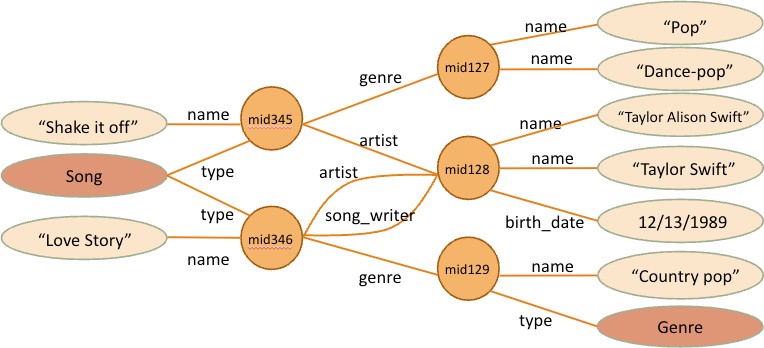
With the widespread applications of KGs, how to model and capture all valuable knowledge in the world has emerged as a prominent research area. This talk delves into this subject through the author’s journey in the past decade, enriched with extensive scientific re- search and production deployment experiences gained at esteemed companies like Google, Amazon, and Meta.

### Generation of knowledge graphs

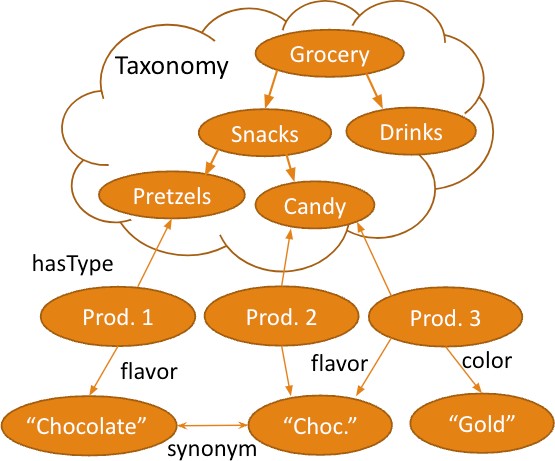
In this talk, we discuss a few generations of KGs. The first generation is entity-based KGs, where both ontology and data are more rigorous, and nodes in the graphs are mostly entities that have one-to-one correspondence with real-world entities (see Figure 1(a) as an example). Most well-known generic KGs, such as Yago [4] from academia and Google KG [3] from industry, are entity-based KGs.

The second generation is text-rich KGs, where ontology and data allow much more ambiguities, and nodes in the graphs are more often just free texts. With the text nodes, the graph is mostly in the form of a bipartite graph, as depicted in Figure 1(b). Text-rich KGs are often used to model domains where structure is sparse while ambiguities are abundant, with vague and fluid semantic boundaries between values and even classes, such as Product, Bio-informatics, and Health.

The upcoming generation is not fully shaped yet and we call it dual neural KGs for now. It encodes knowledge explicitly as triples (as in KGs) and implicitly as embeddings (as in language models). The same piece of knowledge may co-exist in both forms or stay on one side that is more suitable, and there is smooth transition between the two forms to allow harmonic blending. We discuss why we believe co-existing is the key for success, at least in the near future.



(a)An example entity-based KG in the music domain. Nodes are mostly entities, each with an ID.



(b) An example text-rich KG in the product domain. The top depicts the taxonomy, which can be a rich and deep hierarchy. The bottom depicts data instances, where attribute values are mostly texts; as such, it is mostly in the form of a bipartite graph (except edges like "synonym").  
**Figure 1: Example knowledge graphs.**

### The recipe from innovation to practice

KG is an area that has witnessed success both in research and in industry. As we discuss evolution of KGs, we employ it as an example to illustrate the cycle from innovation to production practice, and subsequently to the next round of innovation. This iterative cycle often comprises several stages, each contributing to impacts from initial to profound.

1. **Feasibility:** The cycle first starts with a (or a series of) prototype or an experiment, showing the feasibility of a crazy idea, which sometimes seeds a new field.
2. **Quality:** The second stage focuses on gradually improving the quality of the solution (a model, an algorithm) to pro- duction quality, which enables trustworthy and pleasant user experiences. This is the key stage to land an innovation as a tangible product: unless attaining production quality, a research idea will only remain research.
3. **Repeatability:** Once we achieve success with the initial product, usually within a limited scope (a few domains, or working under a set of constraint conditions), the next stage is to repeat the success for larger scopes like broader domains. This stage often emphasizes building pipelines to facilitate automation, and employing machine learning (ML) models to minimize manual work. It is a stage leading to much higher business impact.
4. **Scalability:** Although repeatability could lead to impact enhancement of 1-2 orders of magnitude, it oftentimes falls short of achieving true scalability, demanding impacts of thousands or millions of times. Scalability often necessitates a new set of solutions that substantially reduce costs and eliminate all manual work from the loop.
5. **Ubiquity:** Finally, ubiquity seeks to maximize the scope of applicability, to encompass long-tail use cases, to remove any underlying assumption in the solutions. Pursuing such solutions often triggers a new round of innovations, initiating the next cycle (sometimes even scalability can lead to the next cycle).

This talk interweaves the discussions of the three generations of KGs and the innovation-to-practice-to-innovation cycle. Through the former, we illustrate how development of techniques leads to larger and larger business impact; through the latter, we shed insights on how the pursuit of large business impact sparks new innovations. Finally, we reflect on critical factors for production success.

### References

[1] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In Proc. of ISWC.  
[2] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, and Jamie Taylor. 2008. Freebase: a collaboratively created graph database for structuring human knowledge. In SIGMOD. 1247–1250.  
[3] Amit Singhal. 2012. Introducing the Knowledge Graph: Things, Not Strings. Google Official Blog.  
[4] Fabian M. Suchanek, Gjergji Kasneci, and Gerhard Weikum. 2007. YAGO - A Core of Semantic Knowledge. In [WWW.](http://www/)

## Speaker

| Xin Luna DongHead Scientist at Meta Reality Lab | Principal Scientist at Meta Reality Lab. Prior to joining Meta, she was a Senior Principal Scientist at Amazon, leading the efforts of constructing Amazon Product Knowledge Graph, and before that one of the major contributors to the Google Knowledge Vault project, and has led the Knowledge-based Trust project, which is called the “Google Truth Machine” by Washington’s Post. She has co-authored books "Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases" and “Big Data Integration”, was awarded ACM Distinguished Member, and VLDB Early Career Research Contribution Award for “Advancing the state of the art of knowledge fusion”. She serves in the VLDB endowment and PVLDB advisory committee, and is a PC co-chair for KDD'2022 ADS track, WSDM 2022, VLDB 2021, and Sigmod 2018 |
| --- | --- |

1. There are two knowledge bases before all KGs discussed in this talk, Cyc (cyc.com) and WordNet (wordnet.princeton.edu); they are limited in scope and scale because of hand-crafting. [↑](#footnote-ref-0)