

Google's Knowledge Vault: It's Features and Comparisons with other Knowledge Bases

Knowledge graphs are a unique way to showcase the intricate connections between multiple objects and get a more semantic way of representing text data. These objects can represent an array of things, starting from abstract ideas and concepts to real life scenarios and events. The knowledge graph takes all these relationships and stores the information in a graph database. The data can also be displayed through a graph structure. Knowledge graphs are also able to apply natural language processing and machine learning to get deeper semantic meaning from their datasets. A complete knowledge graph is used in applications such as search systems, which need to know how certain words and documents are related to each other in order to retrieve the correct information for a query. While the knowledge graph is very useful, it is very hard to associate entities to the right types and to find accurate relations, which means many mistakes can be made.

A widely used knowledge graph is Google's Knowledge Graph. It is specifically used for Google's search engine and uses many different sources to collect its ever growing dataset. Google has gone one step further and extended their current knowledge graph with the introduction of Google's Knowledge Vault. The Knowledge Vault goes beyond the traditional knowledge graph by being able to gather and sort a large collection of information without the need for human evaluations. Additionally, while the old knowledge graph relied mostly on test-based extractions, the vault combines getting information from different web content, such as human annotations, with prior information that's been collected by older knowledge bases. The size of the

vault is also exponentially larger than any of its predecessors. Google's ultimate goal with their newest creation is to create an almost "all-knowing" system that is often present in fictional stories with superior advanced technological capabilities.

Many others have attempted to build something similar to Google's technologies. One of them is the knowledge base Wikidata. A knowledge base stores unstructured or structured data which can be used by a larger system and it is a more general form of a knowledge graph. Wikidata differs in the fact that it is an open knowledge base which means that it can be edited by both machines and people. It is also specifically for storing data for all the projects associated with Wikimedia. Google's knowledge bases also credit Wikipedia as the source of a lot of its data. Wikidata is available for public use and can even be used together with other datasets. This is very different from Google's knowledge bases that are much more secure and mostly handled by the company itself. Wikidata and Google's knowledge bases are similar in the way that they can both be optimized to help improve search engines. Google's knowledge graph can actually use the data from Wikidata to populate its databases so that it can be more attuned to the user's needs.

Similar to Wikidata, DBpedia is another open knowledge base which makes structured data available for public use. DBpedia also extracts entities and their relationships from Wikipedia data. It has even been extended beyond Wikipedia and made applicable to other datasets. With the extracted data DBpedia also allows us to manipulate and look at the data in multiple different ways. Because of this, DBpedia and Wikidata can almost be seen as complementing each other.

Google's Knowledge Vault has had many previous iterations which has allowed it to learn and to get to the advanced form that it has today. While all these knowledge bases have different implementations and levels of complexity, they all aim to create solutions for a lot of the same problems. Whether it be aiding in improving search engines, finding out semantic relations through natural language processing, providing machine language services, or exploring new artificial intelligence related advancements, through Google's Knowledge Vault and other knowledge bases we've been able to get a complex understanding of all of these topics.

Citations

<https://www.ibm.com/cloud/learn/knowledge-graph>

<https://research.google/pubs/pub45634/>

<https://searchengineland.com/google-builds-next-gen-knowledge-graph-future-201640>

https://www.wikidata.org/wiki/Wikidata:Main_Page

<https://www.wikinative.com/why-is-wikidata-optimisation-essential-to-search-engine-results/>

<https://medium.com/openlink-software-blog/what-is-dbpeda-and-why-is-it-important-d306b5324f90>