From Google Knowledge Graphs to Google Knowledge Vault:

Applications of Ontology

Introduction

The future semantic web pictures a web of data represented using ontology, rather than in databases, and text content of web pages. Google Knowledge Graph is a step towards building a graph database, and Google Knowledge Vault is an enhanced builder with automatic information gathering and machine learning to assign probability of facts.

Build Semantic Web with Ontology

Semantic web is a vision brought up in the article "The Semantic Web" (Berners-Lee et al.) in May 2001. The current World Wide Web (WWW) is a web of documents that are linked in various ways including via hyperlinks. But from the data point of view, the data are not independent of the of documents' contents, and needs to be extracted from the documents by human before being used by automatic information processing methods on the web.

The Semantic Web is an extension of the World Wide Web that will develop a set of protocols, formalisms and standards to transform the Web of pages into a Web of Data. So the links will be between data rather than pages, and the data will be able to accessed independently of the pages and applications contain them. The web of data enables the sharing and integration of data in different formats from different sources.

Apply Ontology to Knowledge Graph

In May 2012, Google launched Knowledge Graph. Google Knowledge Graph is not an individual product to use, but rather appears as a card to the right of the links to the top ranked search results, and contains popular facts about the queried entity (people, places, things, and etc) extracted from high-ranking sites to provide the direct answers. So the users don't have to browse through the returned links to find the answers. Knowledge Graph appears in the search results only when Google deems it relevant according to the type of entity being queried.

On the other hand, Knowledge Graph is also a database of facts used by Google to enhance their search engine behind the scenes,

By May 2020, the information covered by Google Knowledge Graph had grown to 500 billion facts on 5 billion entities.

Add Automatic Crawling and Probability Assignment to Google Knowledge Vault

In 2014, Google raised the concept of Knowledge Vault, an upgraded version of Knowledge Graph that scales and can identify whether something is true or not by checking a "knowledge base of content" in the research paper "Knowledge Vault: A Web-Scale Approach to Probabilistic Knowledge Fusion". A knowledge base is a system that stores information so that machines as well as people can read it. However, a knowledge base deals with facts rather than textual data.

Knowledge Vault is aimed being the "the largest store of knowledge in human history," by using an algorithm to autonomously gather and merge information all over the web(text, tabular data, page structure, human annotations, and etc), and turn the raw data into usable pieces of facts and relationships based on all existing knowledge using machine learning without human editorial involvement.

Knowledge Vault resort to existing knowledge bases such as Freebase to validate facts and assign confidence scores to the truthfulness of facts during its "link prediction in a graph" process. The process uses either "path ranking algorithm (PRA)" method or "neural network model (MLP)" method.

It also uses a "quiz crowdsourcing approach" to find knowledgeable people who might answer questions that add to the Knowledge Vault, titled Quizz: Targeted Crowdsourcing with a Billion (Potential) Users.

The main components of Knowledge Vault include extractors, graph-based priors, and knowledge fusion. The extractors extracts triples (subject, predicate, object) from Google's Knowledge Graph or other large data sets and assign confidence scores. The graph-based priors can learn from existing knowledge bases. The knowledge fusion reaches a final score on factual probability.

By cross-referencing new facts with what it already knows, Google Knowledge Vault has extracted 271 million "confident facts", facts that are considered more than 90 per cent chance of being true, from 1.6 billion pulled facts, 4469 relation types and 1100 kinds of entities in 2016.

Knowledge Graph is already being used by systems to help robots and smartphones to understand what people ask them, and hence to provide answers. Looking forward, Knowledge Vault, if assembled in the future, could make possible all sorts of new applications and computing scenarios, such as future artificial intelligence applications, machine-to-machine communication, augmented reality, predictive models and virtual assistant use cases. In addition, Google may implement Knowledge Vault as a separate searching service like Microsoft's "Bing Entity Search API", or it may choose to integrate Knowledge Vault as part of the automated knowledge gathering component of its current search engine, and continue present it as the enhanced Knowledge Graph.

REFERENCES :

Hall W., O'Hara K. (2009) Semantic Web. In: Meyers R. (eds) Encyclopedia of Complexity and Systems Science. Springer, New York, NY. https://doi.org/10.1007/978-0-387-30440-3_478