ontotext

Graph databases like GraphDB are popular for a variety of reasons. They make it easy for you to import data without creating complex schemas. They store relationships extracted from unstructured data. You can use them to combine Linked Open Data with your own data and extend your knowledge about facts like people, places, organizations and events. As a result, the types of queries you can perform and the intelligence returned expands.

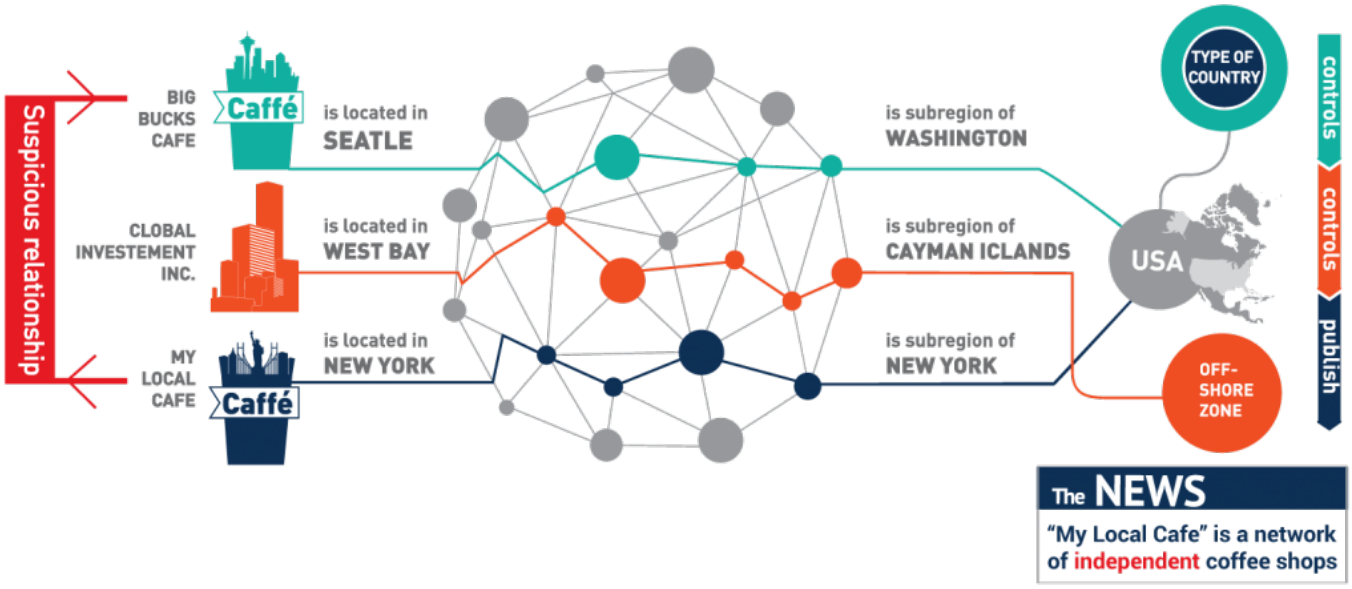
There are dozens of reasons why organizations are adopting this exciting new form of database. One of the most important aspects of graph databases is that data is stored in the form of relationships. These relationships tell you something about the entity. For example, “John works at Banking Corp” or “Sally lives in Nottingham”. As you create more and more semantic links (known as triples – the atomic form of intelligence inside a graph database) you uncover more meaning because of connections across the triples.

This new found intelligence can be used to identify unknown or non-obvious relationships and linkages between facts. Two of the most important attributes of graph databases are inference and semantic data integration. The first allows you to create new facts from existing facts. The latter allows you to integrate many forms of data while maintaining connections back to the original sources. Keeping all of your data in synch and materializing new facts using inference are two important aspects of graph databases and semantic technology.

What is Inference?

Inference is the ability to materialize new facts from existing facts. For example, if we know that Fido is a dog and we know that a dog is a mammal, then we can infer that Fido is a mammal.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com



How can inference help your business? Let's use the graph database example above. A business person analyzing entities such as companies may need to know relationships that exist between different companies. Some of them may not be obvious. In the example above, we know that "Big Bucks Cafe" controls a company called "Global Investment Inc.". We also know know that "Global Investment Inc." controls a chain of coffee shops called "My Local Cafe".

As the diagram shows, data about "My Local Cafe" was also extracted through a text mining pipeline from a news article on the Cafe and stored inside a graph database. Because of the transitive properties of graph databases, we can infer (red dashed lines) that "Big Bucks Cafe" controls "My Local Cafe".

Why Semantic Data Integration?

In a graph database like GraphDB, we can also observe other facts about the world that have been integrated. These facts may come from Linked Open Data. For example, we know that "Big Bucks Cafe" is in Seattle and Seattle is a subregion of Washington State. We know that "Global Investment Inc." is in West Bay and West Bay is a subregion of the Cayman Islands. And we know that the Cayman Islands are classified as an offshore zone for investment purposes.

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

Most importantly, we can infer that there is a suspicious relationship between “Big Bucks Cafe” and “My Local Cafe” using inference rules that take into account the location of the two entities and the relationships they have to each other. Without connected facts and inference, you simply could not determine that all of these relationships actually exist.

Integration of Text and Data

Semantic data integration, when done correctly, has the ability to maintain real time feeds from text mining pipelines into your graph database. One of the biggest challenges organizations face is extracting meaning from unstructured data. Therefore, including text mining in your semantic stack is essential if you want to analyze free flowing text, create triples on the fly and store them inside a graph database.

Closely aligned with text mining is something called disambiguation or identity resolution. As you analyze text, identify entities and classify them, you will inevitably uncover names that refer to the same entity. For example, Robert Smith, RJ Smith, Bob James Smith and Bobby Smith may actually be referring to the same person. Optimizing the storage of facts that refer to the same entity is an important aspect of the graph database enabling fast queries and inference.

Graph databases hold the keys to unlocking hidden meaning in your data. Because GraphDB is a special type of graph database, it provides you with extremely powerful qualities that other graph databases do not have. It can load, query and infer new facts simultaneously and at high rates speed. It has direct connections to text mining pipelines allowing you to extract meaning from your unstructured data and create new facts in real time. It ensures that the semantic triples in GraphDB are kept in synch with changes to your content stores. It allows you to develop hybrid queries that include semantic facts and full-text search within unstructured data.

Graph databases allow you to tell a story. They allow you to connect the dots. When you use this powerful type of database, true meaning is one query away.

Want to learn more about graph databases like Ontotext’s GraphDB?

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com

ontotext

White Paper: The Truth About Triplestores

Download Now

Copyright: Ontotext New York | Sofia | Basel
Reach out to us: info@ontotext.com