

Novel high-performance hybrid system for semantic factoring of graph databases

September 20 2011



Massive amounts of data can be analyzed using a novel hybrid system for semantic factoring in graph databases. Each line represents a frequent relationship type between entities in the dataset.

Imagine trying to analyze all of the English entries in Wikipedia. Now imagine you've got 20 times as much information. That's the challenge scientists face when working with gigabyte data sets. Scientists at Pacific Northwest National Laboratory, Sandia National Laboratories and Cray, Inc. developed an application to take on such massive data analysis challenges. Their novel high-performance computing application uses semantic factoring to organize data, bringing out hidden connections and threads.

The team then used their applications to analyze the massive datasets for

the [Billion Triple Challenge](#), an international competition focused on demonstrating capability and innovation for dealing with very large semantic graph databases, known as SGDs.

Why it matters? Science. Security. In both areas, people must turn massive data sets into knowledge that can be used to save lives.

As SGD technology grows to address components from extremely large data stores, it is becoming increasingly important to be able to use high-performance [computational resources](#) for analysis, interpretation, and visualization, especially as it pertains to the innate structure. However, the ability to understand the semantic structure of a vast SGD still needs both a coherent methodology and the [high-performance computing](#) platform to exercise the necessary methods.

The team took advantage of the Cray XMT architecture, which allowed all 624 gigabytes of input data to be held in RAM. They were then able to scalably perform a variety of novel tasks for descriptive analysis of the inherent semantics in the dataset provided by the Billion Triple Challenge, including identifying the ontological structure, the sensitivity of connectivity within the relationships, and the interaction among different contributions to the dataset.

The semantic database system research team is developing a prototype that can be adapted to a variety of application domains and datasets, including working with the bio2rdf.org and future billion-triple-challenge datasets in prototype testing and evaluation.

More information: Joslyn C, R Adolf, S al-Saffar, J Feo, E Goodman, D Haglin, G Mackey, and D Mizell. 2010. "High Performance Semantic Factoring of Giga-Scale Semantic Graph Databases." Semantic Web Challenge Billion Triple Challenge 2010. cass-mt.pnl.gov/btc2010/pnnl_btc.pdf

Provided by Pacific Northwest National Laboratory

Citation: Novel high-performance hybrid system for semantic factoring of graph databases (2011, September 20) retrieved 2 May 2024 from <https://phys.org/news/2011-09-high-performance-hybrid-semantic-factoring-graph.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.