Case Study

Caterpillar, Inc.

Neo4j Provides Natural Language Processing at Scale, Making Equipment Repair More Efficient

INDUSTRY
Manufacturing

USE CASE
Natural Language Processing / Artificial Intelligence

GOAL
- Create an NLP tool to allow users to extract meaning from documents at scale.

CHALLENGE
- Relational database returned unparsed strings of text that didn't provide meaning.

SOLUTION
- Used Neo4j to create a scalable NLP tool to uncover trends and make maintenance and repairs more efficient.

RESULTS
- Conduct real-time searches for millions of documents
- Can connect cause and effect, and identify deeper-level questions

Caterpillar, Inc. has more than 27 million documents that track vehicle repairs and maintenance. By using Neo4j to perform natural language processing (NLP), the company can now search at scale to uncover repair trends and issues, prescribe solutions and make valuable predictions, all of which increase the efficiency of vehicle repairs and maintenance across the company.

The Company

Caterpillar, Inc. is the world's leading manufacturer of construction and mining equipment, diesel and natural gas engines, industrial gas turbines and diesel-electric locomotives. Customers turn to Caterpillar to help them develop infrastructure, energy and natural resource assets. Its portfolio of 20 brands offers services and solutions to meet the unique needs of a variety of industries and customers around the world. In 2018, sales and revenues for the 90-year-old company exceeded $54 billion worldwide.

The Challenge

Any time a Caterpillar machine is brought in for repair or maintenance, a technician creates a warranty document that chronicles the complaint, an analysis of the problem and the solution.

There's a large-scale repository of technical documents, much of which was quite good from the outlook of labeling and computational linguistics standards. However, there was a lot of disparate data to connect.

The company recognized there was valuable data housed in more than 27 million documents and set about creating an NLP tool to uncover these unseen connections and trends.

For the last decade, they had already been exploring NLP for purposes such as vehicle maintenance and supply chain management. Although a large percentage of the data could be mapped correctly in some domains, it didn't mean they could represent this knowledge and leverage it in a meaningful way.

"We wanted to create a system that would allow someone to ask any type of question as long as it was in the domain," said Ryan Chandler, Chief Data Scientist at Caterpillar. "This meant creating a dialog system to test the use of a graph, demonstrate an open-ended user interface capable of answering questions and to develop a capability to create spoken human machine interface."

Case Study

"The topic of natural language dialogue between people and machines is probably going to be analytics, and the mechanism to make that happen is natural language processing – a perfect fit with graph databases."

–Ryan Chandler,
Senior Data Scientist, Caterpillar, Inc.

The Solution

Because a graph is the lowest level of structure and provides massive flexibility, graph databases are a natural fit for language processing and machine learning.

Language processing is often broken down into either dependency structures, which looks at the verb and draws arcs from the verb to the relationship of the other words relative to the verb, or it breaks down into a constituency tree. Both of these structures are graphs.

Caterpillar employed Neo4j for graph data structures to create a logical form of knowledge. This NoSQL alternative to relational databases allowed them to build ontologies and perform deduction.

To get from natural language to graph query results, the team created data architecture that ingests text via an open-source NLP toolkit, which uses Python to combine sentences into strings, correct boundaries and omit "garbage" in the text. Data is also imported from SAP ERP systems, as well as non-SAP ERP systems.

The Machine Learning Classification tool learns from the portion of data already tagged with terms such as cause or complaint to apply to the rest of the data.

It uses WordNet as a lexicographic dictionary to provide definitions for the words, the Stanford Dependency Parser to parse the text and Neo4j to find patterns and connections, build hierarchies and add ontologies.

Once this is all put together, users can conduct meaningful searches with simple Cypher queries.

The Results

What naturally follows is a prescribed action, such as the appropriate next step if an engine is "knocking," and to uncover the associated problem and diagnosis.

"These types of solutions are the furthest thing from off the shelf AI," says Morgan Vawter, Chief Analytics Director at Caterpillar, Inc. "They embody the mind of the organization, its domain knowledge and, therefore, they are the product of the painstaking translation from (wo)man to machine."

Questions about Neo4j?

Contact us across the globe:
info@neo4j.com
neo4j.com/contact-us