Estudio de caso

Caterpillar, Inc.

Neo4j proporciona procesamiento de lenguaje natural (NLP) a gran escala, haciendo que la reparación de equipos sea más eficiente

INDUSTRIA
Fabricante

CASO DE USO
Procesamiento del lenguaje natural / Inteligencia artificial

OBJETIVO
Crear una herramienta de PLN que permita a los usuarios extraer significado de documentos a gran escala.

RETO
La base de datos relacional devolvía cadenas de texto sin analizar que no aportaban ningún significado.

SOLUCIÓN
Uso de Neo4j para crear una herramienta de PLN escalable con la que descubrir tendencias e incrementar la eficiencia del mantenimiento y las reparaciones.

RESULTADOS
- Realización de búsquedas en tiempo real en millones de documentos
- Posibilidad de conectar causas y efectos, y de identificar cuestiones más profundas

Caterpillar, Inc. tiene más de 27 millones de documentos de seguimiento de reparaciones y mantenimiento de vehículos. Mediante el uso de Neo4j para el procesamiento del lenguaje natural (PLN), la empresa puede buscar a gran escala para descubrir tendencias y problemas de reparación, definir soluciones y formular predicciones útiles, lo que aumenta la eficiencia de las reparaciones y del mantenimiento de vehículos en toda la empresa.

La empresa

Caterpillar, Inc. es el principal fabricante mundial de equipos de construcción y minería, motores diesel y de gas natural, turbinas de gas industriales y locomotoras dieseleléctricas. Los clientes recurren a Caterpillar para desarrollar sus infraestructuras, energía y recursos naturales. Su cartera de 20 marcas ofrece servicios y soluciones para satisfacer las necesidades exclusivas de una amplia gama de sectores y clientes en todo el mundo. En 2018, las ventas y los ingresos mundiales de esta empresa de 90 años de actividad superaron los 54 mil millones de dólares.

El reto

Cada vez que llega una pieza de maquinaria Caterpillar para reparación o mantenimiento, un técnico crea un documento de garantía que registra la reclamación, un análisis del problema y la solución.

Tienen un gran repositorio de documentos técnicos, muchos de ellos satisfactorios en cuanto a etiquetado y estándares de lingüística computacional. Pero había multitud de datos dispares que conectar entre sí.

La empresa se dio cuenta de que tenían información valiosa alojada en más de 27 millones de documentos y se propuso crear una herramienta de PLN para descubrir las conexiones y tendencias invisibles.

Durante la última década habían explorado el procesamiento PLN para fines tales como el mantenimiento de vehículos y la gestión de la cadena de suministro. Aunque era posible asociar correctamente un gran porcentaje de los datos en algunos dominios, ello no implicaba que la información fuera interpretable y utilizable debidamente.

"Queríamos crear un sistema que permitiera a un usuario plantear cualquier tipo de pregunta, siempre y cuando fuera parte del dominio", explica Ryan Chandler, científico de datos sénior de Caterpillar. "Para eso había que crear un sistema de diálogo con el que probar el uso de un grafo, demostrar una interfaz de usuario abierta capaz de responder preguntas y desarrollar una capacidad para crear una interfaz de comunicación oral máquina-persona".

Estudio de caso

"Es probable que el tema del diálogo del lenguaje natural entre personas y máquinas sea el análisis, y el mecanismo para lograrlo es el procesamiento del lenguaje natural, que se adapta perfectamente a las bases de datos de grafos."

– Ryan Chandler,
Científico de datos sénior,
Caterpillar, Inc.

La solución

Debido a que un grafo es el nivel más bajo de la estructura y proporciona una enorme flexibilidad, las bases de datos de grafos son ideales para el procesamiento del lenguaje y el aprendizaje automático.

El procesamiento del lenguaje suele dividirse en estructuras de dependencia, que se basan en el verbo para trazar arcos del verbo a la relación de las otras palabras con relación al verbo; o puede descomponerse en un árbol de constituyentes. Ambas estructuras son grafos.

Caterpillar utilizó Neo4j para estructuras de datos de grafos con las que crear una forma lógica de conocimiento. Esta alternativa NoSQL a las bases de datos relacionales les permitió crear ontologías y realizar deducciones.

Para pasar del lenguaje natural a los resultados de las consultas de grafos, el equipo creó una arquitectura de datos que ingiere texto a través de un kit de herramientas de código abierto PLN, que usa Python para combinar oraciones en cadenas, corregir límites y omitir "basura" del texto. Los datos también se importan de sistemas SAP ERP y sistemas que no son SAP ERP.

La herramienta de clasificación de aprendizaje automático aprende de la porción de datos ya etiquetados con términos como causa o reclamación para aplicarlo al resto de los datos.

Utiliza WordNet como un diccionario lexicográfico para proporcionar definiciones de las palabras, el analizador de dependencias Stanford para analizar el texto y Neo4j para encontrar patrones y conexiones, crear jerarquías y agregar ontologías.

Con todo eso dispuesto, los usuarios pueden realizar búsquedas significativas con consultas Cypher simples.

Los resultados

La consecuencia natural es una acción prescrita como, por ejemplo, el siguiente paso requerido si un motor martillea; y descubrir cuáles son el problema y el diagnóstico correspondientes.

"Este tipo de soluciones son lo más alejado de la inteligencia artificial comercial", dice Morgan Vawter, directora de análisis de Caterpillar, Inc. "Representan el cerebro de la organización, su conocimiento del dominio y, por lo tanto, son el producto de la traducción minuciosa del mensaje de la persona a la máquina."

Neo4j es el líder en tecnología de base de datos de grafos. Siendo la base de datos de grafos más utilizada en el mundo, ayudamos a marcas globales - que incluyen Comcast, NASA, UBS y Volvo Cars – a revelar y predecir como las personas, procesos y sistemas se interrelacionan. Usando el enfoque en las relaciones, aplicaciones desarrolladas usando Neo4j afrontan problemas de conexión de datos tales como análisis e inteligencia de datos, detección de fraude, recomendaciones en tiempo real y grafos de conocimiento. Más información en neo4j.com.

¿Tiene preguntas sobre Neo4j?
Contacte con nosotros en todo el mundo:
info@neo4j.com
neo4j.com/contact-us

© 2021 Neo4j, Inc. Todos los derechos reservados.