



## Caterpillar, Inc.

# Effiziente und skalierbare Instandhaltung von Geräten dank Natural Language Processing (NLP) und Neo4j

### BRANCHE

Fertigung

### USE CASE

Natural Language Processing / Künstliche Intelligenz

### ZIEL

Entwicklung eines NLP-Tools für die aussagekräftige und skalierbare Auswertung von Dokumenten

### HERAUSFORDERUNG

Relationale Datenbanken lieferten nur unbearbeitete Textzeichenketten ohne Bedeutung/Kontext

### LÖSUNG

Mit Neo4j entwickeltes skalierbares NLP-Tool zur Aufdeckung von Tendenzen und für eine effiziente Instandhaltung

### ERGEBNISSE

- Echtzeit-Abfragen in Millionen von Dokumenten
- Ursache-Wirkung-Verknüpfung und Identifizierung von tiefgehenden Fragen

*Caterpillar kann auf mehr als 27 Millionen Dokumente zurückgreifen, um die Instandhaltung und Reparatur von Fahrzeugen zu überwachen. Neo4j wird für die natürliche Sprachverarbeitung genutzt. So kann das Unternehmen skalierbare Abfragen durchführen, Trends beobachten und Störungen identifizieren, entsprechende Maßnahmen vorschlagen und Vorhersagen treffen – was die Instandhaltung aller Fahrzeuge effizienter macht.*

## Das Unternehmen

Caterpillar ist der weltweit führende Hersteller von Bau- und Bergbaumaschinen, Diesel- und Erdgasmotoren, Industriegasturbinen sowie dieselelektrischen Lokomotiven. Kunden vertrauen auf Caterpillar bei der Entwicklung von Vermögenswerten in puncto Infrastruktur, Energie und natürliche Ressourcen. Mit insgesamt 20 Marken bietet das Unternehmen Services und Lösungen für unterschiedlichste Anforderungen in einer Vielzahl an Branchen. 2018 erreichten die weltweiten Verkaufszahlen und Umsätze des 90 Jahre alten Traditionsunternehmens mehr als 54 Mrd. US-Dollar.

## Die Herausforderung

Bei jeder Reparatur oder Wartung einer Caterpillar-Maschine werden sowohl Beanstandungen als auch die Fehleranalyse und Lösung genauestens in einem Serviceheft dokumentiert.

Caterpillar verfügte deshalb über umfangreiche technische Unterlagen, die relative gute Metadaten boten. Allerdings gab es eine Menge separate Datensätze miteinander zu verknüpfen.

Dem Unternehmen war klar, dass in den über 27 Mio. Dokumenten wertvolle Informationen steckten. Caterpillar nahm die Entwicklung eines NLP-Tools in Angriff, um verborgene Zusammenhänge und Tendenzen aufzudecken.

Caterpillar erforscht seit über zehn Jahren wie NLP für die Instandhaltung sowie für das Supply Management eingesetzt werden kann. Ein Großteil der Daten wurde zwar entsprechend zugeordnet, jedoch konnte das gesammelte Wissen nicht so abgebildet werden, dass sich daraus nutzbare Erkenntnisse ableiten liessen.

„Unser Ziel war ein System, das jede Art von Abfrage in der Domain erlaubt“, erklärt Ryan Chandler, Chief Data Scientist bei Caterpillar. „Wir entwickelten also ein Dialogsystem, um das Potential von Graphtechnologie zu testen – einschließlich einer erweiterbaren Benutzeroberfläche zur Beantwortung von Fragen, sowie ein sprachbasiertes Mensch-Maschine-Interface.“

## Fallstudie



„Analytik ist für die natürliche Kommunikation zwischen Mensch und Maschine das zentrale Thema. Natural Language Processing sorgt dabei für die Umsetzung – ideal in Kombination mit Graphdatenbanken.“

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

### Die Lösung

Da Graphen die unterste Ebene einer Datenstruktur darstellen und enorme Flexibilität bieten, eignen sich Graphdatenbanken ideal für Sprachverarbeitung und Machine Learning.

In der Sprachverarbeitung wird entweder die Abhängigkeitsstruktur verwendet, wobei ein Verb im Kontext zu den anderen Worten betrachtet wird. Oder es wird ein Verzeichnisbaum verwendet. Beide Strukturen sind Graphen.

Caterpillar setzt Neo4j ein, um das gesammelte Wissen in eine logische (Graph)Struktur zu bringen. Die NoSQL-Alternative zu relationalen Datenbanken ermöglicht es, Ontologien aufzubauen und Schlüsse daraus zu ziehen.

Um natürliche Sprache in Graph-Abfragen zu übertragen, entwarf das Team eine Datenarchitektur, die Text über ein Open-Source-NLP-Tool aufnimmt, mit Python zu Zeichenketten kombiniert, Abgrenzungen vornimmt und „Text-Müll“ aussortiert. Daten werden sowohl aus SAP als auch anderen ERP-Systemen importiert.

Das Machine Learning Classification Tool lernt aus den Daten, die bereits mit Tags wie „Ursache“ oder „Beanstandung“ versehen sind und überträgt diese auf weitere Daten.

Für die Definition von Wörtern wird das lexikographische Wörterbuch WordNet verwendet. Der Stanford Dependency Parser analysiert den Text. Neo4j erkennt Muster und Verbindungen, baut Hierarchien auf und fügt Ontologien hinzu.

Einmal zusammengeführt, können Anwender über einfache Cypher-Abfragen aussagekräftige Suchen durchführen.

### Die Ergebnisse

Das System kann vordefinierte Maßnahmen empfehlen, z. B. wenn ein Motor „klopfende“ Geräusche macht. Ähnliche Störungen und Diagnosen lassen sich schnell abrufen.

„Diese Services sind alles andere als KI-Standardlösungen“, so Morgan Vawter, Chief Analytics Director bei Caterpillar. „Sie verkörpern den Geist des Unternehmens, nutzen sein Fachwissen und sind das Ergebnis der nicht einfachen Kommunikation von Mensch zu Maschine.“

Neo4j ist die führende Graph-Plattform, die Unternehmen wie Airbus, Comcast, eBay, NASA, UBS, Walmart entscheidende Innovationen und Wettbewerbsvorteile bietet. Tausende von Community- Projekten sowie mehr als 300 Kunden erschließen vernetzte Daten mit Hilfe von Neo4j, um Zusammenhänge zwischen Menschen, Prozessen, Standorten und Systemen aufzudecken. Der Fokus auf Datenbeziehungen ermöglicht es Anwendungen, die mit Neo4j entwickelt wurden, die Herausforderungen vernetzter Daten zu meistern – von künstlicher Intelligenz, über Betrugserkennung und Echtzeit-Empfehlungen bis hin zum Stammdatenmanagement. Weitere Informationen unter [Neo4j.com](http://Neo4j.com) und [@Neo4j](https://twitter.com/Neo4j).

Fragen zu Neo4j?

Kontakt:  
[info@neo4j.com](mailto:info@neo4j.com)  
[neo4j.com/contact-us](http://neo4j.com/contact-us)