CASE STUDY

Knowledge Graph-Powered Diabetes Research

DZD
German Center for Diabetes Research

The DZD created a knowledge graph that incorporates clinical trial data, public datasets, and medical research. Graph algorithms uncover patterns that move the DZD closer to a cure.

BY THE NUMBERS

1.8B nodes

4.9B relationships

30M publications ingested

PLATFORM

Neo4j Enterprise Edition

INDUSTRY

Life Sciences/Medical Research

USE CASE

Data Discovery

OBJECTIVE

Combine DZD data with public data and medical research

CHALLENGE

Empower researchers with internal and external data

SOLUTION

Enrich DZD's knowledge graph with data on related diseases and automatically update graph with the latest medical research using NLP

RESULTS

- Quick retrieval of data from clinical trials
- Find patterns for further research
- Faster identification of diabetes subtypes

The Company

Founded in 2009, the German Center for Diabetes Research e.V. (DZD) is funded by the German Federal Ministry of Education and Research (BMBF) and the participating states. The DZD brings together experts at the national level to develop effective prevention and treatment measures for all types of diabetes across disciplines with the help of modern biomedical technologies.

The Challenge

About seven million Germans suffer from diabetes, one of the most widespread diseases in the nation. To gain new insights and develop effective prevention and treatment measures, the DZD studies the disease from different angles. "Our goal was to enable access to the data across locations, disciplines, species, and data types," said Prof. Dr. Martin Hrabě de Angelis, DZD board member. "At the same time, the more than 450 scientists in the DZD should also be able to access external expertise."

The Strategy

The first challenge was connecting clinical trial data. "With Neo4j, we were able to transfer the metadata from clinical studies into a graph very quickly," explains Dr. Alexander Jarasch, Head of Data and Knowledge Management, DZD. "Using Neo4j Bloom for visualization, questions could be answered quickly and easily. How many blood samples from patients under 69 do we have? Which studies did the samples come from? What parameters were measured?"

The next phase incorporated cross-disciplinary research data, including data on diseases associated with diabetes such as stroke, heart attack, cancer, and Alzheimer's disease. Human data, which is limited, will be supplemented with standardized data from animal models, such as research on mice. This linkage allows the DZD to draw conclusions about humans and to investigate similarities in individual genes and metabolic processes. For example: Which type of diabetes can be traced back to which genes? What is the effect of external factors?

For Jarasch, Neo4j beats relational databases: "In the graph, I see genes, proteins, and metabolic pathways visualized pictorially. I can follow links across multiple nodes, dive into groups of data, and move freely in all directions. This is investigative research in the truest sense of the word."

The graph database is similar to a library with thousands of publications – except that in the graph we can hold the right book with the right information in our hands within seconds and come across unexpected connections."

Dr. Alexander Jarasch, Head of Data and Knowledge Management, DZD

The Solution

Launched in 2017, DZDconnect, the DZD's knowledge graph built on Neo4j, serves affiliated healthcare and medical professionals. Layered on top of the DZD's relational databases, DZDconnect links the systems and data silos of the health centers.

Knowledge graphs offer a rich platform for incorporating and connecting more and more data, at scale. DZDconnect is updated with the latest medical research. Natural language processing (NLP) reads in and automatically annotates more than 30 million publications from the PubMed data. Algorithms perform a semantic analysis of the texts, classify relevant entities, and link them to internal information in the database.

"Reading and absorbing information from the latest publications is simply not feasible without assistance from technologies such as NLP," Jarasch explains. "Currently, it still takes about 1.5 seconds to analyze an abstract on a decent machine. While that sounds fast, it would actually take about a year and a half to summarize all 30 million publications. Our approach of using NLP and graph technology runs in parallel and is automated in the background."

The Neo4j Graph Data Science Library plays an important role. One goal is to identify different subtypes of type 2 diabetes to provide better therapy (precision medicine). With the help of the integrated graph algorithms, scientists can subdivide the dataset. Based on predefined parameters, the community detection algorithm identifies patient clusters, allowing researchers to investigate them more precisely. Algorithms find attributes of the diabetes subtypes and identify shared characteristics (e.g., height, weight, medication, or genetic defect).

The Results

The DZD knowledge graph has 1.8 billion nodes and 4.9 billion edges. Instead of manually searching various databases, as was previously the case, all relevant data can be mapped holistically in the graph. This increases the speed of queries and reduces the susceptibility to errors when extracting and aggregating data. The diversity and depth of detail of the data also allows a new perspective.

"In DZDconnect, we combine external knowledge with internal data," says Jarasch. "The graph database helps us to retrieve the relevant data quickly and in a targeted manner. It's similar to a library with thousands of publications – except that in the graph we can hold the right book with the right information in our hands within seconds and come across unexpected connections."

Questions about Neo4j?

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