

ONDERZOEKSVOORSTEL

Toepassen van tijdelijke grafiek transformatie op object traceringsdata voor het trainen van een LLM.

Bachelorproef, 2024-2025

Maarten Van der Schueren

E-mail: maarten.vanderschueren@student.hogent.be

Co-promotor: B. Peirens (Tracked, bart.peirens@tracked.be)

Samenvatting

Hier schrijf je de samenvatting van je voorstel, als een doorlopende tekst van één paragraaf. Let op: dit is geen inleiding, maar een samenvattende tekst van heel je voorstel met inleiding (voorstelling, kaderen thema), probleemstelling en centrale onderzoeksvraag, onderzoeksdoelstelling (wat zie je als het concrete resultaat van je bachelorproef?), voorgestelde methodologie, verwachte resultaten en meerwaarde van dit onderzoek (wat heeft de doelgroep aan het resultaat?).

Keuzerichting: AI & Data Engineering

Sleutelwoorden: LLM, OTG, EPCIS, CosmosDB, Assistant Agents

Inhoudsopgave

| | | |
|---|--------------------------------------------|---|
| 1 | Inleiding | 1 |
| 2 | Literatuurstudie | 1 |
| | 2.1 Cosmos DB | 1 |
| | 2.1.1 Waarom Cosmos DB gebruiken | 1 |
| | 2.1.2 Grafiek modellering | 1 |
| 3 | Methodologie | 2 |
| 4 | Verwacht resultaat, conclusie | 2 |
| | Referenties | 2 |

1. Inleiding

Binnen de Arcelor Mittal Gent bestaan er veel verschillende processen die verspreid zijn over meerdere afdelingen. Hierbij bevat elke afdeling een deel van het proces om van grondstof tot een afgewerkt product te komen. Tussen elke afdeling bevindt zich een deel van de toeleveringsketen, waarbij het resultaat van de ene afdeling het beginpunt is voor de volgende afdeling. Dit zorgt ervoor dat iedere stap zijn invloed heeft op de kwaliteit van het eindresultaat.

Elke afdeling is opgebouwd rond een specifiek proces, zoals de hoogoven die ruwe grondstoffen omzet in ruw ijzer en de staalfabriek die ruw ijzer in staal verandert. Omdat deze afdelingen zeer specifieke processen hebben, hebben ze ook zeer specifieke behoeften en datamodellen. Dit maakt analyse of onderzoek over afdelingen heen zeer complex en tijdsrovend.

De scope van deze bachelorproef is tweedelig, waarbij deel één eruit bestaat om samen met Tracked vanuit een gestandaardiseerd datamodel die zichtbaarheid events bijhoudt om te zetten naar een grafiekmodel. Deel twee bouwt verder voort op dit grafiekmodel om dan een LLM (large

language model) te trainen die eenvoudig informatie eruit kan halen en teruggeven.

De centrale vraag die hier gesteld wordt is: „Hoe kunnen we efficiënt en snel grafiekmodellering toepassen om een LLM te ontwikkelen die in staat is om het waar, wanneer, wat en hoe van gebeurtenissen binnen een proces vast te stellen, ter ondersteuning van klachtenafhandeling bij productfouten?”

2. Literatuurstudie

2.1. Cosmos DB

2.1.1. Waarom Cosmos DB gebruiken

Cosmos DB is een NoSQL-database van microsoft, deze heeft een lage latentie, multi-query-api die makkelijk grote hoeveelheden data kan verwerken en heeft een grote beschikbaarheid zegt van der Put (2020) wat zeer belangrijk is in ons project. Naast deze feiten is CosmosDB horizontaal schaalbaar wat betekent dat we op hoogtepunten een miljoen lees- en schrijfaanvragen kunnen verwerken door het nodige aantal servers toe te voegen. De hoge beschikbaarheid wordt gegarandeerd door replicatie waardoor we snel kunnen overschakelen als er iets fout gaat in de database.

2.1.2. Grafiek modellering

Cosmos DB biedt ook grafiek modellering aan met behulp van Gremlin

Je mag deze sectie nog verder onderverdelen in subsecties als dit de structuur van de tekst kan verduidelijken.

3. Methodologie

4. Verwacht resultaat, conclusie

Referenties

van der Put, M. (2020, augustus 28). *Azure Cosmos DB*. <https://humandigital.nl/nieuws-en-artikelen/azure-cosmos-db/>