

GeoGPT - Søk i geografiske data med språkmodeller

GeoGPT - Søk i geografiske data med språkmodeller

Prosjektsammendrag

Dette prosjektet har laget en "proof of concept" ved å benytte språkmodeller og vektordatabase til gjenfinning av geografisk data, for å forbedre søkefunksjonen på Kartverkets portal Geonorge.no.

Som prosjektmetodikk har gruppen hovedsakelig benyttet Scrum med Jira som styringsverktøy.

Prosjektmålet ble definert slik:

"Oppnå en "proof of concept", ved bruk av språkmodell og vektordatabase, for å gjøre søk på Geonorge mer presise og brukervennlige for brukere uavhengig av fagkompetanse"

Prosjektbeskrivelse

Prosjektet har vært utforskende av natur med et høyt nivå av usikkerhet og risiko knyttet til valg av teknologiske løsninger. Gruppen har gjennomført prosjekt fra A-Å, med målsetning om å levere en "proof of concept"-løsning som sluttprodukt. Fiktivt har gruppen vært konsulenter ansatt i Tietoenvry og jobbet på prosjektet for kunden Kartverket.

Prosjektprosess

Oppbyggingen har bestått av innsikts- og analysearbeid innledningsvis, for å få forståelse for kontekst og geografisk informasjonssystemer (GIS). Denne forståelsen var ukjent for gruppen ved oppstart.

Økt forståelse for bruken av Geonorge.no ble skaffet ved hjelp av brukerinnsikt mot Norge digitalt medlemmer. Her ble semi-strukturerte intervjuer benyttet sammen med observasjoner.

Hoveddelen i prosjektet har bestått av å utforske relevante teknologier. Hensikten har vært å synliggjøre fornuftige teknologier som kan tas i bruk for å bedre søkefunksjonen på Geonorge.no, en slags

benchmarking av de ulike alternativene og visning av mulig sammensetning eller arkitektur av teknologiene med en "proof of concept".

Teknologi

Den tekniske løsningen baserer seg på bruk av vektordatabasesøk med språkmodell embeddings og Retrieval Augmented Generation (RAG) for instruks av språkmodell til å besvare spørsmål med relevant kontekst hentet fra Vektor DB. I utviklingen er det brukt Javascript med HTML og CSS. For vektordatabasesøk er det benyttet PostgreSQL med Pgvector. Python, Pandas og Regex er benyttet blant annet for å renske metadata for DOK datasettene fra Geonorge. Docker, shell scripts og batch script er brukt for oppsett og kjøring av Vektor DB.

Ved spørsmål om prosjektet?

Kontakt Lars Gunnar Stokke

Mob: 48113414

Mail: Lg_stokke@hotmail.com

Vedlegg



Rapport bacheloroppgave - GeoGPT-prosjekt



Fullført oppgave

Publisert: 2024-05-16
Grad: Bachelor
Studium: Bachelor IT og informasjonssystemer (BACIT)
Leverings- 2024 - Vår
tidspunkt:
Samarbeid: Kartverket og Tietoenvry

Fagområder

- Ingeniør- og teknologiske fag
- Datateknologi

Fakultet

- Fakultet for samfunnsvitenskap

Emnekode

- IS-304 - Informasjonssys., bachelor

Deltakere

- Aleksander Grimstad Mo
- Carsten Østergaard
- Lars Gunnar Stokke
- Theodor Litlere
- William Peter Johansson Åredal

Del på sosiale medier: [!\[\]\(c507f772dba2b921f86777f01218e570_img.jpg\)](#) [!\[\]\(a75296508989caaa77a08d26cfccd4e5_img.jpg\)](#) [!\[\]\(55463e2fc8fd9dd5cdf6584182081aba_img.jpg\)](#)