

Nyhetssøker for Digin

I dagens teknologilandskap publiseres det enorme mengder nyheter hver eneste dag. For en næringsklynge som Digin, med over hundre medlemsbedrifter innen IKT, digitalisering og innovasjon, blir informasjonsmengden en reell utfordring. Det er rett og slett ikke mulig å manuelt holde oversikt over hva som skjer i bransjen, i regionen og blant medlemmene.

Dette var utgangspunktet for prosjektet mitt: å utvikle et system som automatisk henter inn, filtrerer og oppsummerer relevante nyheter ved hjelp av kunstig intelligens.

[Digin](#) fungerer som en samlingsarena for teknologibedrifter og offentlige aktører i Agder. Bedriftene spenner fra store konsultentselskaper som Bouvet, Sopra Steria og Atea, til mindre spesialister og regionale institusjoner som Universitetet i Agder. Felles for alle er behovet for å vite hva som skjer i fagmiljøet rundt dem. Enten det er nye prosjekter, innovasjoner, forskningsresultater, arrangementer eller muligheter for samarbeid. For å gjøre denne jobben både mer nøyaktig og langt mindre tidskrevende utviklet jeg et automatisert nyhetssøkesystem som kombinerer webscraping med moderne språkmodeller.

Systemet henter data fra to hovedkanaler. Den første er RSS-feeds fra over tjue norske nyhetskilder, som NRK, E24, Digi.no, Teknisk Ukeblad, Stortinget og Regjeringen. RSS gjør det mulig å hente godt strukturert nyhetsdata, men er begrenset til kilder som faktisk tilbyr slike feeds. Den andre kanalen er webscraping av rundt 70 medlemsbedrifter og andre aktuelle nettsider. Her tilpasses scraping-prosessen til hver enkelt side som håndterer variasjoner i layout og struktur. Systemet ignorerer sider der robots.txt ikke tillater scraping for å sikre etisk og stabil drift.

Datainnhentingene fordeles ved hjelp av en strategi som sikrer jevn dekning av alle kilder over tid. Etter at informasjonen er hentet ut, sendes hver artikkel videre til en språkmodell via OpenRouter API, som vurderer relevansen etter et sett med strenge kriterier. Artiklene må være ferske, teknologirelaterte og ha kobling til Agder, Digin eller medlemsbedrifter. Typisk blir 60–80 prosent av alle artikler forkastet, mens de som blir vurdert som relevante får en kort og presis oppsummering av språkmodellen.

Hele prosessen fra innhenting, vurdering, til lagring og eksport tar noen minutter for rundt 50 artikler. Dette står i sterk kontrast til den manuelle prosessen som kunne krevd flere timer daglig og likevel kun dekket en brøkdel av kildene. I tillegg lagres alle artikler i en Supabase-database med full historikk, slik at informasjon aldri forsvinner. Relevante artikler sendes automatisk til Google Sheets, der de er lettere tilgjengelige for Digin. Det gjør det enklere å dele informasjon, oppdage trender og hente ut spesifikke artikler eller temaer ved behov.

På teknisk nivå består systemet av en Python-backend med BeautifulSoup for scraping, Feedparser for RSS, OpenRouter for AI-vurdering og en database i Supabase. Et Streamlit-interface gir visuell tilgang til dataene. Ytelsen er optimalisert ved hjelp av både parallell prosessering og god håndtering av databaseoperasjoner. Artikler lagres kun etter vellykket AI-evaluering og sorteres etter når den er publisert, hvilken kilde den er hentet fra og om den er relevant, med eventuell oppsummering. Systemet er robust, med feilhåndtering som omfatter fallback-modeller, retry-mekanismer og logging for enkel debugging.

Utviklingsprosessen har gitt erfaring innen flere områder: effektiv bruk av LLM-APIer, etisk webscraping, database-design, parallellisering i Python og viktigheten av god logging og feilhåndtering. Blant de største utfordringene var store variasjoner i nettsidestrukturer, gode prompts til LLM og optimalisering av ytelse.

Dette prosjektet viser hvordan kunstig intelligens kan brukes for å løse praktiske utfordringer i næringslivet og fungerer som et verktøy for Digin for å unngå repetitive oppgaver samtidig som det gir et bedre grunnlag for beslutninger. Systemet gjør det mulig å oppdage muligheter tidligere, reagere raskere og styrke verdien Digin tilbyr sine medlemmer.