

Fra katalog til kunnskapsgraf: Lars Dahles boksamling som forskningsobjekt

Lars G Bagøien Johnsen (Nasjonalbiblioteket), Jennifer Thøgersen (VID vitenskapelige høgskole)

Hvilke historier kan en samling fortelle – om sin samler og verdenen rundt ham? Og kan AI-teknologier hjelpe til å synliggjøre disse historiene?

Midt på 1900-tallet fikk misjonsskolen i Stavanger (nå VID) boksamlingen til den kjente norske misjonæren og akademikeren Lars Dahle. Vi digitaliserte nylig kortkatalogen for denne samlingen og jobber med å importere deler av samlingen til Alma. Dette er viktig for å gjøre samlingen søkbar, men samtidig ønsker vi å løfte fram katalogens verdi som forskningsressurs i seg selv. Potensialet strekker seg langt utover bibliografisk gjenfinning. Hvert kort bærer spor av tid, sted og tanke: forfatternavn, titler, utgivelsessted og -år danner et historisk, geografisk og tematisk kart over en intellektuell horisont i bevegelse – med misjonærens samling som ett eller flere punkt i dette bildet. Katalogen er ikke bare en inngang til bøker – den er en bok i seg selv, skrevet i metadata.

Når vi leser katalogen som data, åpner det seg nye forskningsoppgaver. Titlene kan tematiseres, språk identifiseres, og digitale versjoner spores opp i nasjonale og internasjonale samlinger.

Utgivelsesstedene kan geolokaliseres og danne grunnlag for kartbaserte analyser, mens forfatternavn gir mulighet for kjønnsstudier og nettverksanalyser basert på publiseringsmønstre.

For alt dette, utforsket vi bruk av språkmodeller for å rengjøre, berike og strukturere dataene. Vi skal presentere våre suksesser og utfordringer med denne fremgangsmåten og hvordan vi håndterte modellenes mangler. Når det arbeidet er fullført, kan vi løfte katalogen videre – fra digitalisering til strukturering. Resultatet kan bli en kunnskapsgraf som både gir grunnlag for Alma-oppføringer og åpner for digitale humaniorastudier: tematisk, språklig og historisk.

From catalog to knowledge graph: The Lars Dahle book collection as a research object

What stories can a collection tell – about its collector and the world around him? And can AI help reveal these stories?

In the mid-1900's the mission school in Stavanger (now VID) was gifted the book collection of Lars Dahle, a well-known Norwegian missionary and scholar. We recently digitized the card catalog for this collection and are working toward importing portions of the collection into Alma. This is important to make the collection searchable, but we also want to highlight the value of the catalogue as a research resource itself. The potential extends far beyond bibliographic retrieval. Each card carries traces of time, place and thought. Author name, title, place and year of publication form a historical, geographical and thematic map of an intellectual horizon in motion – with the missionary's collection as one or more points in this picture. The catalogue is not just an entrance to books – it is a book itself, written in metadata.

When we view the catalogue as data, new research possibilities emerge. Titles can be thematized, languages identified, and digital versions tracked down in national and international collections. The places of publication can be geolocated and form the basis for map-based analyses, while author

names provide the opportunity for gender studies and network analyses based on publication patterns.

To achieve this, we explored the use of language models to clean, enrich, and structure the data. In this presentation, we will discuss the successes and challenges we had with this approach and how we addressed shortcomings in the language models' output. When that work is complete, we can take the catalogue further – from digitization to structuring. The result can be a knowledge graph that both provides the basis for Alma entries and opens up for digital humanities studies: thematic, linguistic, and historical.