



Word Vector Based Search using Open Source Machine Learning Libraries

Andreas Bergman | Stein-Erik Bjørnnes



Oppdragsgiver

✘ Fotoware AS

”High Performance Digital Asset Management Software for Videos, Photos, Graphics and all other files”

✘ Kontaktperson – Andreas Gnutzmann



Problemstilling

”Hvordan kan maskinlæring brukes for å gi en større bredde og relevans på tekstbaserte bildesøk”



Problemstilling forts.

- ✘ Nå søkes det i hovedsak kun ordrett fra bilders metadata, og derfor er det nødvendig å utvide søket slik at relaterte ord inkluderes og søkeområdet utvides
- ✘ Løsningen er å lage en søkemotor som tar hensyn til dette, og gjennom maskinlæringsbiblioteker finner lignende ord og søker over et bredere område og deretter sorterer etter relevanse



Hvorfor denne oppgaven?

- ✘ Helt nytt for både oss og veiledere
- ✘ Kunstig Intelligens er for alvor på vei opp i samfunnet
- ✘ Sluttrapporten blir en beskrivelse på hva som eksisterer av rammeverk og biblioteker
- ✘ Kan være til stor hjelp for andre ute i verden



Resultat

- ✘ Bruker bibliotekene word2vec og fastText
- ✘ Vi fant ut at en kombinasjon av begge bibliotekene var den beste løsningen for gode resultater
- ✘ Webapplikasjonen: Gjør et søk – søkeordet går igjennom maskinlæring og returnerer flere relevante ord som vi søker med i bildearkivet
- ✘ Bilder og metadata er lever av Reuters