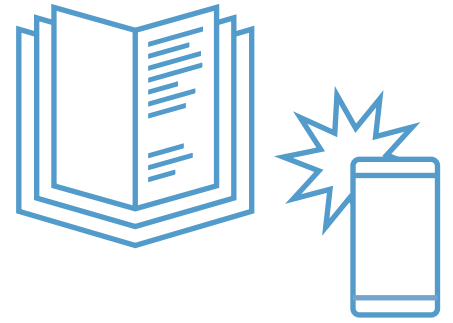


# Med AI katalogiserer vi hurtigere



**På Syddansk Universitetsbibliotek er det lykkedes at skabe et digitalt værktøj, som skal hjælpe med at katalogisere vores særskrifter på en brøkdel af den tid, det ville tage at gøre det manuelt**

□ **Michael Monefeldt**, Akademisk medarbejder, Syddansk Universitetsbibliotek

Der er foregået kvantespring inden for kunstig intelligens i de seneste år, og det er oplagt at udnytte fremgangen til at registrere metadata. Forestil dig: Du hiver mobilen frem og tager et billede af et titelblad, og på få sekunder bliver en katalogpost udformet for dig, helt automatisk. Det er virkeligheden i Team for Særskrifter på SDU's Bibliotek i Odense.

Webapplikationen *Katalogica* læser tilmed de mest krøllede bogstaver i vores ældre biblioteksbøger; og værktøjet kan skelne mellem forfatter, titel, udgivelsessted, etc., så alt metadata havner i de rigtige kasser i systemet.

Det hele er sat op på en hjemmeside, som er designet til at gøre let og intuitivt at uploade billeder og at rette til, hvor den kunstige intelligens laver fejl.

Men kan vi virkelig stole på, at værktøjet er præcist nok til at erstatte den manuelle indtastning? Der er ikke tale om blind tillid – vi sætter ikke *Katalogica* til at gøre alt arbejdet selv. Hver eneste AI-generede post tjekkes af menneskeøjne for at sikre, at kvaliteten er på niveau med traditionelle metoder. *Katalogica* udgør således ikke en komplet automatisering, som erstatter den menneskelige intellekt. Der er snarere tale om et hjælpemiddel, der kan spare bibliotekarerne langt det meste af arbejdsbyrden. Programmet foretager det indledende tastearbejde, og vi skal bare rette til.

Værktøjet er blevet testet flere gange og er ved at komme i brug blandt bibliotekarere og studentermedhjælpere. Indtil videre vidner alt om, at vi sparer ufattelige mængder tid ved at få computeren til at gøre det meste af arbejdet: Det er blevet anslået, at med hjælp fra *Katalogica* kan vi arbejde 10 gange hurtigere. Tilgængeligheden af vores særskrifter bliver øget

# “ Et stort forskningspotentiale kan gå tabt, når studerende og forskere ikke kan finde, hvad de har brug for. Derfor er det vigtigt at eksperimentere med digitale løsninger, der kan sætte farten på processen op

med rekordfart, og vi får overskud til at fokusere mere energi på bevaring og formidling af vores interessante og værdifulde materialer.

## Håb for de usynlige samlinger

Hvis et værk ikke står på hylderne og ikke kan søges frem i bibliotekets database, eksisterer det praktisk talt ikke for brugerne. Gemt væk i en kælder eller skjult i stakke af papkasser i fjernmagasiner er det ikke blot umuligt for forskere og studerende, men også for de ansatte at orientere sig i bibliotekets samlinger. Et stort forskningspotentiale kan gå tabt, når studerende og forskere ikke kan finde, hvad de har brug for. Derfor er det vigtigt at eksperimentere med digitale løsninger, der kan sætte farten på processen op.

Team for Særsamlinger på SDU's Bibliotek har til opgave at tilgængeliggøre universitetets særsamlinger og har længe katalogiseret manuelt i det daglige. Men vi har endnu mange års tastearbejde foran os. Vores dygtige fagreferenter har arbejdet på at hive de mest værdifulde materialer ud af stakkene, så brugerne kan få fat i disse. Men det kan være svært at forudsige, hvad den enkelte ender med at have brug for; og i arbejdet med at lokalisere de vigtigste bøger i samlingerne, ender vi måske med at fravælge det, som faktisk har værdi. Indføringen af *computer vision* i arbejdet vil formodentlig ændre markant ved, hvordan vi griber katalogisering an, da det vækker håb om en fuldkommen registrering.

## Et specialiseret design

Med løbende feedback fra resten af Team for Særsamlinger og andre kolleger på SDU's Bibliotek har jeg egenhændigt designet og udviklet webapplikationen Katalogica. Jeg har finkæmmet katalogiseringsprocessen med det formål at gøre det så hurtigt som muligt

at registrere de "usynlige" bøger i bibliotekets beholdning. Når man har med hundredtusinder af værker at gøre, får det stor betydning, hver gang man kan hente endnu et sekund fra hver registrering.

Projektet har kaldt på mere end blot en *computer vision*-løsning; det har været en øvelse i at strømline en daglig proces ved at skære til og automatisere, hvor end det var muligt (og forsvarligt). Katalogica er endt med at have et højst specialiseret design, som hjælper sin bruger hvert skridt på vejen mod at rette poster til. Programmets intelligens skinner igennem helt ned i detaljen; for eksempel når det automatisk oversætter fra romertal til arabertal med matematisk nøjagtighed.

Katalogica formaterer til det korrekte katalogiseringsformat med fuldkommen præcision. Vi kan komme til at lave menneskelige fejl, når vi taster derudad. Derimod arbejder Katalogica systematisk uden nogensinde at trættes eller at ramme en forkert knap på tastaturet. Hvis programmet opfatter, at noget er uden for normen, gør det registranten opmærksom på mulige fejl. Og der er ikke flere funktioner i Katalogica end de absolut nødvendige for opgaven; med det fokuserede design begrænser vi mængden af personlige præg, som registranten kan komme til at sætte i sine metadata på trods af skarpt optrukne formateringsregler.

Det kræver ressourcer at designe og kode en brugerflade samt at opsætte sikre dataforbindelser mellem hjemmesiden og serveren i skyen. Algoritmen, der skal læse titelblade og sortere metadata, var allerede klar, før jeg startede produktionen af hjemmesiden, og jeg kunne for så vidt være begyndt at katalogisere for længst. Af alt det ekstra arbejde er dog kommet et brugervenligt værktøj, som vil gøre teknologien let tilgængelig for flere – med den klare fordel, at den kan bruges af hele vores team. Selv med hjælp fra kunstig intelligens vil det vare nogle år endnu, før særsamlingerne på SDU's Bibliotek er registreret, og derfor er det vigtigt, at alle kan udnytte Katalogica.

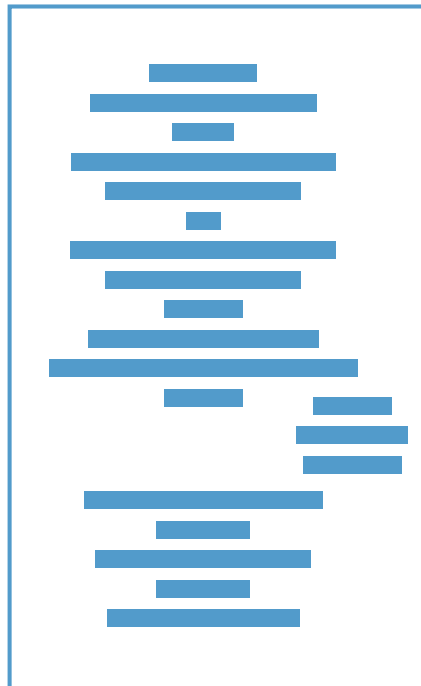
## Katalogica

Katalogica kan aflæse tekst fra titelblade – og af teksten udformer programmet helt automatisk ordnede katalogposter.

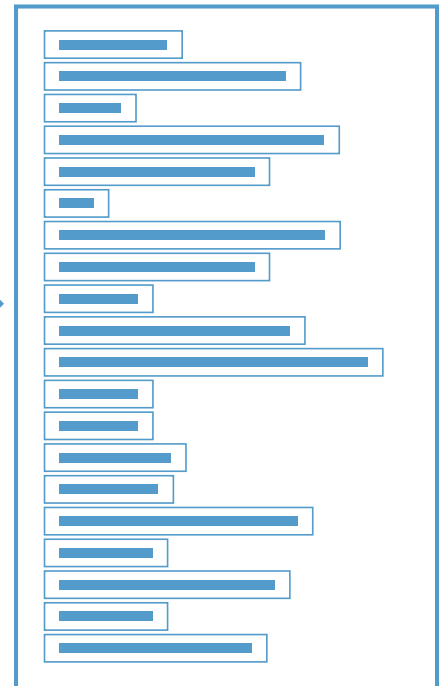
### Titelblad



### Digital tekst



### Ordnet metadata



## Potentialer for fremtiden

Projektets succes gør det oplagt at udforske mulighederne for at registrere andre materialer end de trykte bøger – det værende sig for eksempel magasiner, artikler, nodehæfter mm. Desuden vil det blive relevant at høste metadata fra flere inputs samtidigt. De fleste bøger i særsamlingerne på SDU er fra før det 20. århundrede, og dengang var standarden at nedskrive næsten al information om en bog i titelbladet. Det er ingen større udfordring at videreudvikle webapplikationen, så den kan læse data fra flere kilder, f.eks. en kolofon.

Med lidt videreudvikling ser jeg også et *citizen science*-potentiale i værktøjet. SDU går meget op i at skabe dialog med det omgivende samfund og trække på den hjælp, vi kan få fra borgerne – f.eks. når vi skal måle pH-værdier i vandløb forskellige steder i Danmark. Borgerne kan nemlig være en god kilde til dataindsamling.

Ved at gøre det muligt for borgere at registrere data over internettet, kan vi trække på civilsamfundets engagement og viden, som

en ressource til at berige vores katalog. Deres poster må nødvendigvis også rettes til af bibliotekarer med erfaring inden for katalogisering. Men efter at være set af både en kunstig og en menneskelig intelligens, vil fejlene være minimale, og det vil formentlig blive langt hurtigere at gøre vores samlinger synlige.

Lige nu arbejder jeg på at implementere en mere avanceret søgemekanisme, der skal lade Katalogica fange, hvis et værk allerede findes i bibliotekets katalog. Vi vil gerne undgå dubletposter – og derfor skal Katalogica advare brugeren, når programmet bemærker ligheder mellem nye og gamle poster.