

# Informationens skygge og foran

## En anmeldelse af Marianne Lykke Nielsen

**Claus Poulsen: Informationens skygge og foran : Informationskvalitet, informationsekspllosion og online kataloger. - Roskilde : Roskilde Universitetscenter, 1994. 198 s.**

Ved overgangen fra trykte til online bibliotekskataloger har kun få biblioteker haft ressourcer til at "modernisere" emneindekseringen. Resultatet er, at man ved søgning i mange traditionelle bibliotekskataloger ikke til fulde udnytter de nye genfindingsmuligheder, og mange online kataloger reduceres til at være værktøjer til verifikation.

Claus Poulsen her gennem PARADOKS-projektet forsøgt at skabe et alternativ til det traditionelt emneindekserede bibliotekskatalog for gennem PARADOKS-indeksering at øge mulighederne for at udføre kvalificerede emnesøgninger. Nærværende bog er en beskrivelse af PARADOKS-indekseringens forudsætninger og erfaringer, idet bogen rundes af med evaluering af et PARADOKS-indekseret katalog.

PARADOKS-indeksering (PARAdigmespecifik DOKumentation) består i at lade emnesøgning foregå i en delmængde af bibliotekets bestand, bestående af udvalgt oversigtslitteratur og marginallitteratur. Marginallitteratur karakteriseres som nyskabende litteratur, hvis paradigme adskiller sig fra de dominerende forskningsretninger. Gennem oversigtslitteraturen forventes brugerne at få et overblik over den bedste litteratur indenfor de dominerende forskningsretninger, mens indeksering af marginallitteraturen giver mulighed for en mere nuanceret informationssøgning. Ved PARADOKS-indeksering emnebeskrives de udvalgte dokumenter dokumentorienteret ved at gøre dokumenternes indholdsfortegnelser, abstracts og stikordsregistre fritekstsøgbare.

Bogens første halvdel beskæftiger sig med emnesøgning og indekseringssprog. Med udgangspunkt i indekseringsteori og diverse brugerundersøgelser argumenterer Claus Poulsen for nødvendigheden af nye indekseringsmetoder. Gennem beskrivelse af amerikanske undersøgelser fra først i

80'erne diskuteres det, hvorledes brugerne har vanskeligt ved at foretage emnesøgning og derfor i højere grad kun anvender online bibliotekskataloger til verifikation. Brugernes søger ved anvendelse af forskellige søgestrategier, hvorfor traditionel, kombinatorisk emneords- og forfattersøgning i bibliotekskataloger ifølge Claus Poulsen bør udvides til også at omfatte andre former for søgestrategier, f.eks. browsing i referencer, indholdsfortegnelser og oversigtlitteratur.

Der findes to kategorier af forbedringsmuligheder: de, der satser på at forbedre selve søgeprocessen gennem udvikling af brugergrænseflader, der kan yde support gennem f.eks. automatiseret stavetrol og krydshenvisning, og de, der søger at forbedre katalogernes indhold og indeksering, ved f.eks. at gøre dokumenternes indholdsfortegnelser søgbare.

Begge typer af forbedringstiltag afvises i bogen som utilstrækkelige og uden markant indflydelse på katalogernes anvendelse. Med de store publikationsmængder, der findes idag, er brugerne ifølge Claus Poulsen ikke tilfredse med at få fremvist litteratur, der emnemæssigt matcher deres søgeprofil. Brugernes er interesseret i "den "bedste" litteratur om emnerne - hverken mere eller mindre". Med den "bedste litteratur" må der menes den litteratur, som både modsvarer søgeforespørgslen emnemæssigt og som passer til den aktuelle brugers niveau, motivation, formål etc.

For at kunne udføre sådanne specifikke emnesøgninger og efterfølgende foretage en kvalificeret relevansbedømmelse kræves en dybtgående og nuanceret emnebeskrivelse og et veludviklet, domænespecifikt indekseringssprog. Claus Poulsen foretrækker anvendelse af et ukontrolleret indekseringssprog frem for et kontrolleret indekseringssprog, idet man ved automatisk indeksering oftest får flere søgeindgange pr. dokument, ligesom man undgår den fortolkning og vurdering, som en indeksør nødvendigvis må foretage ved indeksering.

Dette fører frem til følgende konklusioner, der ligger til grund for PARADOKS-indekseringens principper:

*I forhold til emnebeskrivelse ved anvendelse af et kontrolleret indekseringssprog vil emnebeskrivelse i naturlig tekst beskrive emner mere tidsvarende og redundant, hvilket medfører at beskrivelsen er lettere for brugerne at afkode og mere anvendelig som instrument til at afgøre dokumenternes relevans (s. 46).*

*Alle undersøgelser peger på, at potentialet for forbedring ligger i at anvende indholdsfortegnelser, abstracts og lignende fri tekst som emnebeskrivelse (s. 28).*

*Dette kan lade sig gøre, hvis emnesøgning kun foregår i en mindre del af bibliotekets katalog, f.eks. encyklopædiske artikler og anden oversigtlitteratur (s. 28).*

Som det fremgår af citaterne vil udvidede fritekstsøgemuligheder i indholdsfortegnelser, abstracts og stikordsregistre i store bibliotekskataloger resultere i fremfindning af et meget stort antal poster, hvorfor denne søgestrategi ikke uden tilpasning findes velegnet til store forskningsbiblioteker. Den samlede konklusion er således, at forbedringer af emnesøgningsmulighederen i bibliotekskatalogerne går gennem en dobbelt strategi. Denne dobbelte strategi består i at lade emnesøgning foregå i en delmængde af den anskaffede litteratur, der bredt kunne kaldes oversigtlitteraturen, og foretage emneregistreringen ved at genbruge dele af dokumentets tekst så som indholdsfortegnelse, abstracts eller stikordsregister.

Claus Poulsen redegør i sin bog meget sagligt for sine synspunkter. Alligevel har jeg nogle forbehold overfor hans negative konklusioner omkring de forbedringsmuligheder, der ligger i udvikling af mere brugertilpassede brugergrænseflader og en bedre emneindeksering rettet mod specifikke vidensdomæner og brugerbehov samt tilpasset de online søgemuligheder.

Jeg finder Claus Poulsens afvisning af forbedrede brugergrænseflader forhastet, idet de undersøgelser, som hans argumentation bygger på, stammer fra midten af 80'erne, hvor udvikling af mere såkaldt brugervenlige brugergrænseflader først lige var startet. Emneindekseringen var heller ikke dengang ajourført i forhold til de nye emnesøgningmuligheder. Bl.a. OKAPI projektet omtales, hvor man gennem flere år og fortsat arbejder med udvikling af brugervenlige bibliotekskataloger. Som det bekrives i bogen bød OKAPI-projektet i 1987 ikke på "revolutionerende" resultater ved indlemmelse af partiel match søgeteknikker, automatisk stavekontrol og krydshenvisning. Man konkluderede allerede dengang, at det fremtidige udviklingsarbejde skulle koncentreres om udvikling af interaktive søgetesauruser til support ved fremfindning og valg af søgeord.

Projektets nyeste resultater er netop blevet publiceret (1), og heri anbefales tesaurusen stadig som et værdifuldt browsingredskab. I tråd med andre forsknings- og udviklingsarbejder anbefaler gruppen en bruger-styret browsing proces, idet testforsøg med systemdrevne, automatisk anvendelse af tesaurus viser dårlige resultater. Endvidere fremhæves det, at tesaurusen skal være specielt udviklet til det aktuelle bibliotekskatalog og dets brugere for at fungere tilfredsstillende.

I tilknytning hertil er der projekter igang, der koncentrerer sig om udvikling af dynamiske, fleksible og domænespecifikke tesauruser, der giver brugerne mulighed for at lære og forstå et vidensdomænes kompleksitet og forskellige paradigmer, eksempelvis ved at lade brugerne browse i forskellige definitioner af et begreb. Hypertekst forsøges anvendt til implementering af sådanne interaktive tesauruser.

Der arbejdes altså stadig med udvikling af indekserings- og søgesprog, men det er vigtigt at fremhæve, at kvaliteten af både tesaurus og emneindeksering er helt afgørende for gennem denne forbedringsstrategi at opnå bedre emnesøgninger. Det kræver mange ressourcer at udvikle brugertilpasse-

de og domænespecifikke tesauruser, ligesom den nødvendige analytiske indeksering er ressourcekrævende. Jeg er enig med Claus Poulsen, når han i sin bog argumenterer for, at indeksering med henblik på en bedre udnyttelse af bibliotekernes materialebestande bør ske på analyseniveau, ligesom dokumenternes implicite emner og paradigmatisk betragtninger også bør indekseres.

Har bibliotekerne ikke disse ressourcer, kan PARADOKS indeksering ses som et alternativ. Jeg har dog svært ved at forlige mig med at kun udvalgt marginallitteratur og oversigtslitteraturen skal indekseres. Al udvælgelse vil ligesom emneanalyse og -beskrivelse bero på subjektive beslutninger, og vil ligesom indeksering ikke kunne udføres ensartet og konsistent. Iøvrigt er der en indbygget forældelse i oversigtslitteraturen, og brugerne vil blive henvist til materialer via oversigtslitteraturen, som biblioteket ikke nødvendigvis råder over. Herved forsvinder gevinsten ved samtidigt med søgning at kunne verificere de fremfundne materialer. Derfor mener jeg, at hele materialebestanden bør indekseres.

Som beskrevet i bogens teoretiske afsnit om indekseringssprog er kontrollerede og ukontrollerede indekseringssprog komplementære, og vil ved søgning ofte resultere i fremfindning af forskellige dokumenter, hvorfor en kombination af de to indekseringsformer anbefales. Automatisk indeksering udelukker ikke kontrolleret indeksering, og jeg mener, at man også ved PARADOKS-indeksering med fordel kan anvende begge typer af indekseringssprog. Dokumenternes implicite emner og paradigmatilhørsforhold vil ikke altid fremgå tydeligt af hverken titel, indholdsfortegnelse eller abstract. En dygtig indeksør med fagligt kendskab til vidensdomænet og et grundigt kendskab til brugergruppen vil derimod ofte kunne identificere sådanne vigtige aspekter i dokumenterne og indeksere dem. Dette vil som nævnt stille større krav til det kontrollerede indekseringssprog m.h.t. til specificitet og aktualitet samt øge ressourceforbruget i forbindelse med indeksering, men det vil til gengæld være en hjælp for de brugere,

f.eks. studerende, der endnu ikke er inde i fagområdet og fagterminologien.

I bogen rejses endnu en væsentlig problematik, nemlig om brugerne ved fritekstsøgning selv er istand til at imødekomme de problemer, der altid vil forekomme p.g.a. manglende morfologisk og semantisk kontrol. Behovet for både søgeteknisk og begrebsmæssig viden illustreres tydeligt gennem søgeeksemplerne i appendiks 5, f.eks.:

*“(alternativ\$ or vedvarende or renewable or dauern\$) adj ener\$”*

*“forurening\$ or ekolog\$ or økolo\$ or verschmutz\$ or medienpädagog\$”*

Denne væsentlige problematik færdigbehandles ikke i bogen, men at den er væsentlig viser undersøgelsesresultatet fra en brugerundersøgelse på Statsbiblioteket (2) med al tydelighed. Ud af 29 slutbrugere anvender 4 (14%) trunkering i deres søgning, 22 (76%) kombinerer med AND, mens kun 4 (14%) bruger OR operatoren. Kun 3% begrænser deres søgning til et bestemt felt. Undersøgelsesresultatet på Statsbiblioteket er ikke unikt; mange andre undersøgelser viser samme tendens. Slutbrugere mangler ofte søgeteknisk viden for at kunne udføre kvalificerede søgninger. I artiklen anbefales øget brugerundervisning og især vejledning som en væsentlig løsning på problemet, men også hel eller delvis automatisk krydshenvisning kunne overvejes. Ihvertfald eksisterer et væsentlig problem, der må løses, hvis PARADOKS-indeksering skal medføre gode emnesøgninger.

Bogens sidste del beskriver en sammenligning af emnesøgningsfaciliteterne i et PARADOKS-indeksret katalog, bestående af 560 oversigtsværker, og et traditionelt bibliotekskatalog på Danmarks Pædagogiske Bibliotek, repræsenteret ved et udtræk på 1344 poster. Ved evalueringen blev det testet, hvorvidt emneordene fra det traditionelle katalog også fandtes som søgeord i det PARADOKS-indeksrede katalog. Undersøgelsen viste, at de traditionelle emneindekseringer i Danmarks Pædagogiske Biblioteks katalog i omkring 75% af

tilfældene kunne genfindes i det PARADOKS-indeksrede katalog. Det blev endvidere undersøgt, om der var forskel på de muligheder for præcisering, som de to typer af kataloger tilbyder. Sammenligning af søgeresultater viste, at for ca. 80% af søgningerne gav de PARADOKS-indeksrede bedre eller tilsvarende præciseringsmuligheder.

Det er derfor nærliggende, som Claus Poulsen, at konkludere, at PARADOKS-indeksering giver bedre eller i hvert fald lige så gode søgemuligheder som et traditionelt indekseret katalog. Jeg finder imidlertid ikke denne kvantitative testmetode anvendelig. En væsentlig forskel på traditionel indeksering og PARADOKS-indeksering er, at man ved PARADOKS-indeksering kun søger i en delmængde af bibliotekskatalogen. Ved evaluering af PARADOKS-indeksering er det derfor væsentligt at undersøge, hvorvidt det er hensigtsmæssigt at begrænse søgning til kun at omfatte oversigts- og marginallitteratur. Er det i overensstemmelse med brugernes ønsker og behov at begrænse fremfindingsmulighederne? Endvidere skal PARADOKS-indeksering ikke kun vurderes i forhold til de traditionelle indekseringsmetoder, baseret på klassifikation og tildeling af få kontrollerede emneord. PARADOKS-indeksering skal evalueres i forhold til alternative forbedringsmuligheder, f.eks. anvendelse af et domænespecifikt, bruger- og systemtilpasset indekseringssprog.

I evalueringen mangler også en vurdering af, om det er muligt at udtrykke dokumenternes implicite emner, f.eks. paradigmatilhørsforhold, ved automatisk indeksering. Fritekstsøgning i titler, indholdsfortegnelser og abstracts vil give mange specifikke søgeindgange til dokumenterne, men fremfindning af paradigme-specifikke dokumenter ved frittekstsøgning vil både kræve stor søgeteknisk viden, og især skal brugeren have stor faglig indsigt for at kunne opstille en tilfredsstillende søgeprofil. Indplacering af dokumenter i deres faglige kontekst vil ofte kræve en fortolkning. Oversigtslitteraturen er jo netop nyttig, fordi de

forskellige dokumenter er blevet fortolket og placeret i en faglig kontekst.

Det er tiltrængt med en ajourføring af forskningsbibliotekernes indekseringspraksis, og Claus Poulsens bog er et gennemarbejdet og inspirerende indlæg i debatten om, hvorledes denne fornyelse kan foregå. PARADOKS-indeksering er en realistisk metode, som både praktisk og med hensyn til ressourceforbrug kan implementeres. Metoden vil imidlertid stille store krav til brugerne og både kræve et stort emnekendskab og stor søgeviden af brugerne. Sammenholdt med, at PARADOKS-indeksering kun indekserer en delmængde af bibliotekernes materialebestand, idet kun oversigts- og marginallitteraturen indekseres, bør indekseringsmetoden kombineres med en manuel, intellektuel indekseringsmetode, der giver mulighed for en emneanalyse- og beskrivelse funderet på et fagligt kendskab til vidensdomænerne og til bibliotekets brugergrupper og deres specifikke informationsbehov.

## Noter

(1) **Susan Jones, Mike Gatford, Steve Robertson, Micheline Hancock-Beaulieu, Judith Secker, and Steve Walker.** - (1995). Interactive Thesaurus Navigation: Intelligence Rules ok?. - I: Journal of the American Society for Information Science. 46 (1), pp. 52-29.

(2) **Birgit Truelsen.** - (1995). Selvbetjening på CD-ROM. - I: Biblioteksarbejde 16 (45), pp.