

Søgemuligheder i fremtiden

Emnedata i fremtiden : kontrollerede/ukontrollerede emneord, nye medier, udvalgets arbejde

Ved Anita Sørensen, medlem af Katalogdatarådet, Udvalget for emnedata.

Oplægget er struktureret i 5 klumper, hvis indhold er:

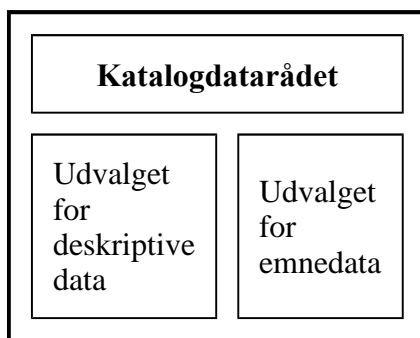
1. [hvem er KED og hvad laver KED](#)
2. [hvorfør bruge emnedata](#)
3. [ord som emnedata](#)
4. [kontrollerede/ukontrollerede emnedata](#)
5. [emnebeskrivelse og søgning i og med nye medier](#)

dvs. i forhold til overskriften starter jeg med det sidste, nemlig udvalgets arbejde, derefter siger jeg noget lidt mere principielt om emnedata, inden jeg går videre med de to andre punkter i overskriften.

ad. 1. hvem er KED og hvad laver KED

KED er en forkortelse af Katalogdatarådet udvalget for EmneData.

Katalogdatarådet er et permanent udvalgt under SBT(Statens Bibliotekstjeneste). Katalogdatarådet har 2 underudvalg - KDD og KED, der detailbehandler spørgsmål til forelæggelse for Katalogdatarådet. Periode 1996-1998

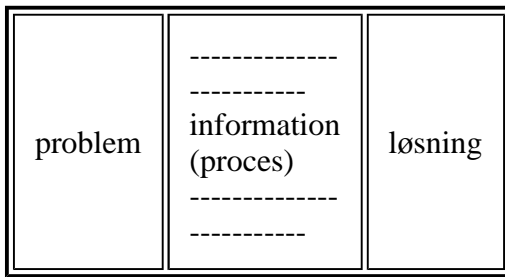


Katalogdatarådet og udvalgene arbejder med alle aspekter af data i de danske bibliotekers elektroniske kataloger, søgemuligheder og grænseflader samt katalogernes udnyttelse i almindelighed.

På emnedataområdet skal KED forbedre adgangen til emnesøgning, specielt bedre mulighed for verbal emnesøgning og fremme koordinering af emneregistrering.

KED's arbejdsområde er alt ang. emnedata, men primært fremme af kontrollerede emnedata. Emnedata er for os både verbale data og klassifikationskoder, samt øvrige søgemæssigt vigtige oplysninger som f.eks. genre (roman...), form (trykt, diskette...), niveau (børnebog...). Emnedata er hovedsaglig knyttet til faglitteraturen.

KED's arbejde er i første omgang at lave indstillinger til Katalogdatarådet om ting, som skal behandles her vedrørende emnedata, men derudover er det vores målsætning at tage nogle mere



Denne model indeholder det element, at der er forskel på information og viden. For en virksomhed er viden værdi. Bearbejdningen af informationerne giver viden.

Emnebeskrivelsen skal præcisere meningen, så den muliggør hurtig udvikling af løsninger. Til det er relationer vigtige, fordi de fremmer forståelse af både problem og løsning. Relationerne er emnebeskrivelsens svage punkt.

Denne model indeholder det element, at der er forskel på information og viden. Bibliotekerne stiller information til rådighed, men er ikke/sjældent involverede i brugen af denne information dvs. om den bliver viden. For f.eks. virksomheder er det nødvendig, at informationerne bliver til viden, da det ikke er informationerne, men viden - og "hurtigt" viden - der for dem er værdi.

Tilbage til selve emnebeskrivelsen. Ideelt set skal emnebeskrivelsen "bryde grænser", hvilket ikke er ligetil, da emnebeskrivelsen faktisk først og fremmest understøtter en søgning, hvor den der søger "kender løsningen". Den, der emnebeskriver, har en faglig viden og emnebeskriver fagligt korrekt. Men den, der søger, formulerer sit spørgsmål på en måde, der afspejler hans/hendes nuværende viden om sit problem. Ofte skal problemet løses på tværs af discipliner.

Konklusionen er, at emnedata er nødvendige for informationsformidling (uanset definition af information). Sat på spidsen kan man sige, at information, der ikke emnebeskrives, det kan heller ikke findes i en søgning. Spørgsmålet er, hvilke emnedata og hvordan de skal "produceres".

ad. 3. ord som emnedata

Som emnebeskrivelseselement er der ingen forskel på, om der anvendes kodede emnedata eller verbale emnedata. Når de verbale emnedata har vundet så stærkt frem, så skyldes det, at de er umiddelbart lettere tilgængelige ved søgning i elektroniske kataloger.

Når jeg her siger "umiddelbart lettere", så skyldes det, at "ord" ikke bare er "ord". De "ord", der skal bruges til emnebeskrivelse skal være "begreber". Det med at skelne mellem "ord" og "begreber" er et forsøg på at sige, at et ord kan være et meningsfuldt element i en tale/ et skrift, fordi det indgår i en sammenhæng, men ved emnebeskrivelse, så skal der findes nogle begreber for den information, der skal formidles. Når man søger, så er man nød til at finde frem til de begreber, der er valgt.

Jeg arbejder på et bibliotek, hvor vi fra 1988 har emnebeskrevet alle materialer med både klassifikationstal og danske emneord. I min dagligdag bruger jeg selv emneordene meget ved søgning, men hvis jeg skal lave en lidt mere omfattende eller en lidt mere kompliceret søgning, så bruger jeg klassifikationerne. Disse har nogle fordele, som det kræver mange søgninger på emneord at opnå.

Selv når vi taler om ord, der er begreber, så er det sådan, at for hver faglig disciplin har et begreb sin egen betydning dvs. at et ords betydning er bestemt af den kontekst, som det indgår i. Dertil kommer, at selv indenfor samme disciplin kan forskellige personer bruge forskellige begreber om samme emne.

I emnedataudvalget er vi sikre på, at ord er kommet for at blive som emnedata. Problemet er, hvordan vi sikre en "god" anvendelse af dem, både til emnebeskrivelse og til søgning.

ad. 4. kontrollerede/ukontrollerede emneord

Når ordene kontrollerede/ukontrollerede sættes overfor hinanden på denne måde, så menes, om der emnebeskrives med ord fra en godkendt liste eller der emnebeskrives med frie ord. I begge tilfælde forudsættes det, at der vil blive valgt relevante ord. Listen kan være opbygget efter forskellige principper. Indtil nu, har der været to hovedprincipper, en emneliste knyttet til et hierarkisk klassifikationssystem (subject heading list) og en tesaurus, hvor den første er et prækoordineret system, medens tesaurusen er et postkoordineret system.

PRINCIPIELT, MEN SÅDAN ER VIRKELIGHEDEN IKKE NØDVENDIGVIS

I et hierarkisk klassifikationssystem kombineret med emneord er emneordets betydning dels bestemt af ordet i sig selv og dels bestemt af dets placering i hierarkiet. I en tesaurus er ordet i sig selv ofte utilstrækkeligt, fordi det ikke er placeret i en entydig sammenhæng.

Kontrollerede emneord har en række karakteristika, som jeg med vilje ikke hverken kalder fordele og ulemper eller grupperer som sådan

- et emne beskrives -princielt- altid med samme ord
- emnebeskrivelsessystemet og dermed emneordene kan identificeres
- nye begreber indarbejdes langsomt/gamle begreber bevares -for- længe
- store krav til faglige viden ved søgning
- niveau for emnebeskrivelse fastlagt/fastlåst

Ukontrollerede emneord har en række karakteristika, som jeg med vilje ikke hverken kalder fordele og ulemper eller grupperer som sådan

- materialet ord anvendes
- tidens ord anvendes
- brugergruppens ord kan anvendes
- "informationens" niveau styrende
- mange alternative søgninger nødvendige for at få dækkende søgning

I diskussionen om kontrollerede/ukontrollerede emneord så synes vi, at det er vigtigt at huske på, at biblioteksverdenen er ikke en homogen enhed. Der er mange bibliotekstyper (skole-, folke-, forsknings-, administrative- og private-) og derfor også mange mange forskellige brugergrupper med meget forskellige behov.

Valget mellem kontrollerede eller ukontrollerede emneord - eller måske en kombination - og valget af det konkrete emnebeskrivelsessystem er ikke nødvendigvis ens for alle, hvorfor et emnebeskrivelsessystem er urealistisk. Eks. forskel medicinsk litteratur på et folkebibliotek og den medicinske litteratur, der er på et medicinsk forskningsbibliotek, og der er forskel på de krav som brugerne har de to steder. Vi mener imidlertid, at principperne for emneordstildelingen og udformningen af emneord skal være ens. (materialets niveau, ubestemt flertal)

Jo flere emnebeskrivelsessystemer brugeren møder på biblioteket (egne base, CD-ROM, Internet) jo mere brug for hjælp har den enkelte til at vælge det bedste søgested og den bedste søgestrategi.

Hvis vi fra emnedataudvalget skal formulere et - beskedent? -ønske/krav, så er det, at indenfor en katalog/base skal der være et emnebeskrivelsessystem, der er konsekvent i sin anvendelse og passer til den katalog/base, som man laver. Det anvendte emnesystem skal være sådan, at brugeren let kan

orientere sig i det og anvende det.

Ved søgning vil en vis faglig viden være nødvendig eller rettere, der er forskel på den søgning, som man kan lave med faglig viden og uden.

Det er værd at overveje, om det ville være en fordel, hvis alle danske biblioteker anvender DK5 sideløbende med et andet specifik emnebeskrivelsessystem. Det skal vurderes nærmere, om det ekstra arbejde giver flere og bedre søgeresultater, om der er søgetekniske problemer f.eks. hvis de samme ord bruges i begge systemer (kan være både med samme eller forskellig betydning). Vi forestiller os, at det kan være en fordel for brugere, der ikke sidder fysisk på biblioteket, at de nemmere kan "finde ind i" i basen vha. et fælles system også selvom det for nogle vil være utilstrækkeligt. Og så er DK5 specielt rettet mod dansk litteratur og danske forhold og kan være med til at understøtte vores kulturelle og samfundsmæssige basis.

ad. 5 emnebeskrivelse og søgning i og med nye medier

Nye medier er ikke præciseret nærmere i overskriften. Jeg vælger her at tolke "de nye medier" som IT (informations-teknologi) eller måske rettere IKT (informations-kommunikations-teknologi) mm. i en tekstmæssig sammenhæng.

Den problematik, som vi har med at gøre, er stadig den samme. Dvs. at der er information på den ene side og en person med et spørgsmål på den anden. Informationen skal stadig formidles til personen.

MEN det hovedspørgsmål, som vi kommer til at diskutere og måske allerede er i fuld gang med at diskutere er, hvordan vi håndterer information, der hele tiden ændre sig.

Dette er ikke kun et emnebeskrivelsesproblem.

Problematikken kender vi det i dag først og fremmest på Internettet, men efter min mening vil den hurtige ændring i informationerne blive dagligdagen på informationsområdet også i det rum, som bibliotekerne skal agere indenfor.

Kriterierne for vidensorganiseringen er ikke nødvendigvis de samme i dynamiske som i statiske omgivelser. Det er meget tænkeligt, at mere formalisering/mere omhyggelig registrering bliver nødvendigt for at tilpasse sig til dynamikken. (Hvor få/mange ændringer gør et dokument til et nyt dokument. Emnebeskrive en gang for alle eller "hen ad vejen" etc.)

Min erfaring med den nyeste IT-teknologi er lille, men blandt andet ud fra nogle diskussioner om og arbejde med udformningen af web-brugergrænseflader til bibliotekssystemer ved jeg, at det er vigtigt i en analyse at skelne mellem, hvad informationsteknologien betyder af ændrede betingelser på områder hver for sig

- 1. for søgeteknik/søgesystemer,**
- 2. for emnebeskrivelse**
- 3. for hvad en informationsbærende enhed er**

Samme skelnen vil nok også være vigtig ved formidling til og oplæring af personalet til brug af web-brugergrænseflader.

ad. 1.

På søgeteknikkens område vil ændringerne være kendetegnet af en ændring fra et statisk til et dynamisk system, fra et lineær til et netværk system.

F.eks. har overgangen fra kort til edb betydet, at der kan laves fritekstsøgning, kombinatoriske søgninger, postkoordineret søgning, søgninger i felter, trunkeres, browses, forskellige automatiserede bearbejdnings af søgeresultaterne, vægtede søgninger, dialog system-bruger.....,

men først og fremmest er den verbale emnesøgning og den postkoordinerede emnesøgning blev understøttet

Brugergrænsefladen ændrer sig fra kommandosystem og/eller menusystem til web-flade. Vores nuværende søgesystemers svageste punkter er emnesøgningen. Web-flader har nogle muligheder for forbedringer.

Boolsk søgning har været den dominerende søgemåde i de hidtidige søgesystemer. Elsket af nogen, hadet af andre og i hvert fald på mit bibliotek sort land for de almindelige brugere. Når der arbejdes i store databaser, så er der problemer med for mange fund, derfor kommer andre metoder ind f.eks. vægtning.

Boolske operatorer søger enkelt ord/begreber og kan ikke skelne mellem forskellige sammenhænge f.eks.søgningen "kvinder og biblen" der kan give fund med mindst disse to slags informationer a) kvinder i biblen og b)kvinder diskuterer biblen

Det har været søgt tegn til tegn søgninger (måske med skjult trunkering). Dette giver nogle problemer, hvorfor nogle arbejder med forsøg på stemming(stamming) i stedet for trunkering.

Browsing - som så vidt jeg kan vurdere - bruges meget af lånerne i vores CD-ROM systemer, bliver lettere i web-brugergrænseflader.

Fuldttekstsøgning/analyse af fuldtekst kan supplere søgninger og specielt bruges ved vurdering af fund, men selvom der er fuldttekstsøgning, så er emnebeskrivelse nødvendig bl.a. pga. mange perifere udtryk og mange fund.

Windows- baserede webflader har både fordele og bagdele søgemæssigt. De hidtiedege forsøg er ikke lige vellykkede. De virker tunge, men teknisk set er der nogle spændende muligheder. Specielt ved emnesøgning er muligheden for at clicke videre dvs. søge videre på emnedata en fordel. Her har hierarkiske systemer med emneord en fordel frem for en thesaurus, så vidt jeg umiddelbart kan se.

ad.2

Emnebeskrivelsen er ikke uafhængig af søgesystem. Søgesystemets muligheder påvirker emnebeskrivelsen, tydeligst ses det med skift fra prækoordineret til postkoordineret emnebeskrivelse.

Selv med det samme emnebeskrivelsessystem, så vil der i praksis ske en glidning afhængig af ændrede søgetekniske muligheder.

Eksempel på IT-ændringer, der har/kan påvirke emnebeskrivelsen (ikke hinanden udelukkende)		
prækoordineret	postkoordineret	vægtet
ordningsord	alle ord bestemte felter	alle ord bestemte felter hvor i teksten strukturering af ord

kort/trykt	1.+2. generation	3.+4. generation
------------	------------------	------------------

ad.3

Den største ændring sker med den informationsbærende enhed. Dette kommer til at betyde noget for, hvad information (som ting) er for noget, men også for beskrivelsen af den informationsbærende enhed. I en papirudgave præsenterer informationen sig struktureret og fysisk, medens den i elektronisk form formes af brugeren. Der sker en ændring fra at betragte dokumentet i sin helhed til at betragteet dokument som noget, hvor "man plukker, hvad man har brug for".

Denne udvikling har allerede været i gang længe, men forstærkes af Internettet. På tekstsiden går den over fotokopieringens vækst til -delvis- downloading, på musikområdet fra at man spiller dele af musik til sampling, på billedområdet fra at man laver collager til billedmanipulation.

Både denne "delvise" brug og den hurtige adgang til information og den hurtige ændring af informationen påvirker vores opfattelse af, hvad information er og stiller krav til (emne)beskrivelsen. Hidtil har vi (emne)beskrevet et dokument/materiale en gang for alle og som et hele. Nu skal vi finde en løsning på,

- at definere den enhed, der skal (emne)beskrives, hvornår det er den oprindelige enhed og hvornår en ny enhed,
- at (emne)beskrive dele af dokumentet/materialet, fordi det er informationerne i dokumentet/materialet ikke dokumentet/materialet, der har interesse.

Vi får på emnebeskrivelsesområdet nye metoder med de nye medier, nemlig at emnebeskrivelsen kan automatiseres (afhænger af krav til emnebeskrivelsen) f.eks. ud fra matching af emnestruktur med tekst, titel etc. Ukontrollerede "emneord" kan genereres ud fra titel, tekst, forfatterens egen beskrivelse. Der kan anvendes sætninger eller fraser til beskrivelsen

Konklusionen er, at nye medier betyder mere, hurtige og stadig ændrende sig information. MEN formidlingsproblemet er der stadig.

Afslutning.

Når man arbejder med emnebeskrivelse i en stadig mere fancy verden, en verden hvor alt går hurtigt, teknikken "bare kan alt", hvor vi svinger os ud på internettet osv., så kan man godt nogen gange spørge sig selv, om man sidder i fortiden eller er et skridt foran andre inde i fremtiden.

Kort sagt, er emnebeskrivelse et overstået fænomen eller bliver det stadig mere presserende?

Har vi søgemuligheder om ikke nu så lige om hjørnet, der kan løse emnesøgningsproblemerne? Også alle de svære hjørner som vi bakser med med den menneskelige hjernes indsats?

For mig er det uden for enhver diskussion, at "menneskegenererede" emnebeskrivelser af information og/eller ekspertformidlet information er to vigtige arbejdsområder i fremtiden. Formidlingen er stadig nødvendig, måske endda af voksende betydning, selvom der også kommer nye muligheder med ændrede søgesystemer.

MEN bliver det også bibliotekernes arbejdsområder?

Det bliver det kun, hvis vi går ind i det med det, som vi kan, nemlig sortere information. ("I gamle dage" bogvalg, emnebeskrivelse, udvælgelse ved søgning).

Og til sidst et citat:

"Subject analysis is one of the most complex and least understood aspects of bibliographical control" (157)

"Unfortunately, while it is possible to determine time, effort and money expended in developing and maintaining a system, it is much more difficult to measure accurately the time, effort and effectiveness of the retrieval process" (158)

Citat Nancy J. Williamson "Standards and rules for Subject Access" i "Cataloging & Classification Quarterly" vol 21 no.3/4 1996 s.155-176

[FABITA](#) | [Konferencen Søgemuligheder i fremtiden.](#)

Sidst opdateret 27.11.96.Informationer på disse sider må benyttes såfremt kilden angives.