

Vidensorganisation¹

Skal bibliotekarer² organisere al information på Internettet?

Af Birger Hjørland

Indledning

En 1. årsstuderende på Danmarks Biblioteksskole foreslog i en opgave at bibliotekarer i fremtiden skal organisere "al information" på Internettet. Dette forslag er, synes jeg, et godt udgangspunkt til at diskutere hvad vidensorganisation er, hvad bibliotekarer gør i dag og må uddanne sig til at gøre i fremtiden, og hvad biblioteks- og informationsvidenskab er og bør videreudvikles til. Én holdning, der i al fald tidligere har været udbredt blandt bibliotekarer, dokumentalister og informationsspecialister, har været ønsket om eet fælles og standardiseret klassifikationssystem (f.eks. UDK³) for alle biblioteker⁴. Konstruktørerne af UDK havde jo oprindeligt den idé at dette system skulle anvendes til at lave en komplet bibliografi over alverdens bøger, artikler og alle andre dokumenter ordnet efter eet system, en egentlig universel bibliografi. Den idé er det jo ikke svært at overføre til Internettet, som på en måde netop bliver en teknologi, der kan muliggøre denne drøm. Dømmen om en universal bibliografi brast i begyndelsen af det 20 århundrede. I stedet for rekommanderedes det at hvert land udarbejder sin egen nationalbibliografi.⁵ Stillet op ved siden af hin-

anden skulle disse nationalbibliografier tilsammen udgøre en slags universalbibliografi. Dette fungerer også nogenlunde for bøgernes vedkommende, men ikke for artikler m.v. Og det er ikke særligt brugervenligt. Hvem har adgang til et sådant bibliotek med alle nationalbibliografierne⁶? Og hvem har tid til at søge så mange forskellige trykte værker igennem for at søge efter nogle bøger, hvis man ikke kender udgivelseslandet? I dag udarbejdes disse nationalbibliografier naturligvis som elektroniske databaser (og der sker en hastig retroinddatering af poster fra før edb-katalogiseringen slog igennem). Det er ikke svært rent teknisk at lægge disse databaser sammen, så de kan søges under eet⁷. Det synes ikke længere at være nogen umulig utopi at forestille sig én database med oplysninger om alle kendte bøger udgivet i alle lande til alle tider. Lad os derfor forestille os, at disse problemer snart bliver løst. Vi har alverdens nationalbibliografier lagt op i et system af elektroniske databaser med fælles registre og/eller søgemaskiner m.v., således at de kan søges under eet⁸. »Onesearch« i *Dialog* har i dag har løst disse praktiske problemer med andre databaser, og virker meget fantastisk (og findes relevant af mange brugere). Lad os se lidt på begrænsningerne i forhold til målsætningen hos

vores studerende fra Danmarks Biblioteksskole, der foreslog at bibliotekarer i fremtiden skal organisere al information på Internettet.

- I. For det første udgør bibliografiske beskrivelser af bøger/monografier kun en meget lille del af al »litteratur« og en endnu mindre del af »al information«. Dette er spørgsmålet om „bibliografisk kontrol.“⁹
- II. For det andet er det et spørgsmål om én endelig repræsentation af dokumenterne kan varetage alle behov? Den mængde af f.eks. bøger, der udkommer i verden er naturligvis en endelig mængde, selvom den er stor. Kan den registreres objektivt og udtømmende, så alle brugerbehov varetages i samme omfang? (Naturligvis med brug af al den teknologiske snilde, der kan opvises?). Dette er spørgsmålet om vidensrepræsentation og genfindingsmuligheder.

Disse to spørgsmål (I&II) er i virkeligheden forbundne (via begrebet „relevans“).

Ad I: Bibliografisk kontrol

Med hensyn til det problemet om at repræsentere „al information“ vil vor førsteårsstuderende måske mene, at man kan tage det i tur og orden. Vi kan udvide fra bøger til også at dække artikler og herfra gå videre indtil „al information“ er registreret. Lad os se på nogle af de centrale problemer, som dette indebærer. Måske er en „komplet“ registrering af alle bøger udgivet i et land en ret overkommelig mulighed. Jeg vil dog påpege nogle historiske og geografiske forskelle bag motiverne til denne bibliografiske kontrol. Formålet hermed er at betone, at den måde en given aktivitet udføres på altid er præget af nogle bagvedliggende forudsætninger og interesser og ikke blot „objektiv“. Den oprindelige interesse for bibliografisk kontrol kan føres tilbage til *enevældens behov for kontrol* og censur med, hvad der udkommer i kongeriget. Såvel den danske pligta everingslov som Det kongelige Bibliotek og andre mægtige instanser i det bibliografiske kontrolapparat bærer stadig præg af denne oprindelige interesse¹⁰. Uden

for monarkiernes kreds kan den dominerende kontrolinteresse være *at sikre udbydernes copyright* på de udgivne dokumenter. Dette medfører helt andre principper, bl.a. frivillighed i a everingen, som det kendes fra USA. Men naturligvis er det ikke blot den oprindelige eller den dominerende interesse, der er bestemmende for, hvordan systemet fungerer i dag. Naturligvis har forlagernes og boghandlernes interesser også i tidens løb påvirket udformningen af f.eks. det danske system¹¹. Den kontrolinteresse, der er det centrale for vores perspektiv, er den interesse, der muliggør at brugere af systemet kan identificere *den bedst egnede tekst til et givent formål* (Wilson, 1968). Sådanne interesser vil jeg til forskel fra de censurmæssige og økonomiske interesser kalde *de erkendelsesmæssige interesser*. Sådanne interesser er af en helt anden karakter, og selvom de tiltag, der sikrer f.eks. censurinteresser og økonomiske interesser også kan være med til at sikre de erkendelsesmæssige interesser, så er dette jo næppe den optimale måde at forvalte dem på. Naturligvis er bibliotekernes dominerende interesser i dag de erkendelsesmæssige og naturligvis arbejdes der yderst seriøst for at sikre en bibliografisk kontrol med dette perspektiv, men man kan altså ikke fuldstændigt abstrahere fra de sociale og kulturelle forudsætninger, der altid vil præge menneskelige aktiviteter, bl.a. i definitionen af de enheder, der skal registreres. Det har hidtil været muligt at definere kriterierne for den bibliografiske kontrol formelt og operationelt ud fra bogtrykningen. Dette bliver imidlertid mere og mere problematisk efterhånden som trykketeknikken udbredes så voldsomt som tilfældet er. Problemet forstærkes yderligere ved publicering på Internettet. Måske bliver det i fremtiden svært at definere hvad en bog er?

Så vidt bøger. Lad os hernæst se på en tilsyneladende simpel udvidelse til også *at dække artikelregistreringen „nationalbibliografisk“*. Det skulle vel ikke være en vanskelig opgave. Vi har vel allerede i Danmark med „Artikelbasen“ (modsvarende den svenske „Artikelsök“) en af de mest omfattende nationalbibliografiske artikelbaser i verden. Hvad er der i vejen for at udbygge denne og etablere til-

svarende i alle andre lande i verden? Den danske Artikelbasen udelukker nogle tidsskrifter, aviser og artikler fra indexering, hvilket ikke er tilfældigt. Det danske og internationale forlag *Munksgaard* udgiver f.eks. mange høj-videnskabelige tidsskrifter, hvis artikler IKKE registreres i den danske Artikelbasen. Hvorfor ikke? Hovedsagelig fordi brugerne af disse tidsskrifter især finder de artikler, de skal bruge gennem internationale databaser som *MEDLINE* og *SciSearch*, mens de typiske brugere af Artikelbasen omvendt ikke er så interesseret i at finde disse tidsskrifters artikler samme med de øvrige tidsskrifters (så ville de nok anvende de internationale databaser og ikke afgrænse søgningen efter specifikt danske bidrag). Artikelbasen henvender sig i højere grad til folkebiblioteksbenyttere, mens benyttere af forskningsbibliotekerne i højere grad betjener sig af internationale fagdokumentationsbaser. Dagbladet Politiken udgiver også en database POLTEKST, der supplerer Artikelbasen ved at have en grundigere registrering af visse aviser. Dette er et andet udtryk for at én monolitisk database *måske* ikke kan varetage alle behov. Vores studerende kunne naturligvis spørge: Hvorfor udvider Artikelbasen ikke blot sin dækning til såvel internationale forskningstidsskrifter udgivet i Danmark som til avisartikler, der er mindre end 5.000 typografiske enheder? - og øvrige „huller“ i dækningen. Svaret er bl.a., at alting koster penge, og enhver tjeneste bliver nødt til at sikre overensstemmelse mellem omkostninger og behov. Jeg mener selv, at der er behov for at de enkelte lande foretager en registrering, der sikre bedre input til internationale databaser. Der er også behov for sådanne registreringer til forskningsadministrative formål, til videnskabssociologiske studier m.v. Men de eksisterende „markeder“ sætter nogle grænser, der nok kan yttes, men altid vil være der. Det er vigtigt at fastslå, at Danmark er ganske atypisk mht. artikelregistrering. Det ville være utopisk at forestille sig en tilsvarende amerikansk artikelbase. De amerikanske artikler registreres i „internationale“ faglige data databaser rettet mod bestemte brugergrupper. *Medline* og *Chemical Abstracts* henvender sig til forskere og specialister indenfor medicin, kemi og grænseområder, mens f.eks. *Magazine Da-*

tabase dækker mere populærvidenskabelige behov og *Academic Index* dækker mere almene behov osv.

De bibliografiske registreringssystemer er - i al fald delvist - organiseret „organisk“ efter målgrupper og brugerbehov, og ikke blot rationalistisk og monolitisk efter logiske principper, der sikrer gensidigt udelukkende og tilsammen udtømmende bibliografier (dvs. som er gensidigt eksklusive og tilsammen er ekshausive).

Det er naturligvis stadigvæk spørgsmålet, om den oprindelige, rationalistiske tanke om et monolitisk universalbibliografisk registreringssystem ikke er den rigtige idé? Denne idé har nogle argumenter, der taler for den. F.eks. at mange lande i verden er meget utilfredse med den dækning deres nationale forskning har i mange af disse såkaldte internationale databaser¹². Der *er* efter min opfattelse behov for en national (eller overnational, f.eks. europæisk) bibliografisk kontrol med inputtet til faglige dokumentationsbaser¹³. Den umiddelbare begrænsning i tanken om et „komplet“ „nationalbibliografisk“ registreringssystem for artikler ligger især i det forhold at de faglige dokumentationsdatabaser opererer med fagspecifikke klassifikationssystemer, thesauri og emnedatasystemer i øvrigt (jfr. nedenfor pkt. II), at de er mere eller mindre organisk tilpasset deres respektive domæner, og at de selekterer de dokumenter, der registreres efter relevanskriterier, der er specifikke for deres domæne og målgrupper. Hvis de fagbibliografiske databaser erstattes af et sådant »monolitisk universalbibliografisk registreringssystem« risikerer vi måske at de bliver for stive og uspecifikke til målgruppernes behov¹⁴.

Delkonklusion 1:

Vi har nu set, at udvidelsen af tanken om nationalbibliografisk registrering fra at omfatte bøger/monografier (og fysiske enheder) til også at omfatte artikler og andre „bibliografiske enheder“ ikke er helt uproblematisk. Noget tyder på at dette spring i dækningen involverer nogle nye behov, som den nationalbibliografiske organisationsmodel måske ikke er optimal til at takle. Dette problem bliver større og

klarere, hvis vi går fra „publikationer“ til andre typer af informationskilder.

Tre begreber: Publikationer, dokumenter og information

Biblioteker tager sig normalt kun af bøger, tidsskrifter, aviser, musikaler, m.v. som er publiceret, dvs. som er *publikationer* (trykkekriteriet har som sagt historisk været betydningsfuldt, men er i dag ved at blive meget problematisk, hvorfor begrebet må revurderes). Her eksisterer der en historisk udviklet arbejdsdeling overfor arkiver¹⁵, der tager sig af ikke-publicerede dokumenter (f.eks. manuskripter og andre arkivalier) og museer, der tager sig af museale genstande. Mange af disse ting vil naturligvis glide sammen i fremtiden, hvor såvel biblioteker som arkiver og museer mere og mere arbejder med Internettet og andre digitale medier. Det er imidlertid en vigtig pointe, at der er væsentlige forskelle i de principper og vilkår, hvorunder dokumenter på den ene side og arkivalier og museumsgenstande på den anden side indsamles, opbevares og registreres. Arkivalier er således *unika*, hvoraf de mest relevante er udvalgt og oceaner af dokumenter kasseres. Hvert eksemplar har normalt ikke samme offentlige interesse som en bog eller et tidsskrift, der jo er trykt fordi udgiveren mener at den vil være af interesse for tusindevis af læsere. Derfor har bøger som sådan en bred kulturel interesse, og der investeres i at udbrede kendskabet til dem, og de klassificeres enkeltvis efter emne (hvorimod arkivalier typisk organiseres efter proviensprincippet).

Pointen er, at publikationer og arkivalier udgør to forskellige former for dokumenter, ud fra hvilke der historisk har udviklet sig ret forskellige organisationer til at varetage opbevaring, registrering og formidling. Der er faktisk tale om to forskellige former for vidensorganiseringssystemer. Selve publikationsprocessen udgør derfor i sig selv en grundlæggende for vidensorganisation, der varetages af forlagene/udgiverne.

Det er klart, at vores førsteårsstuderende kan forestille sig, at bibliotekerne i fremtiden overtager forlagenes rolle. Det er også én mulighed, ligesom det er en mulighed, at forlagene overtager (dele af) bibliotekernes rolle. Hvad jeg vil fremhæve er, at det er væsentligt man forstår de eksisterende funktioner og processer førend man udarbejder forkromede planer om at lave det hele om. Vi må forestille os, at i fremtiden ligger på Internettet såvel arkivalier som publikationer side om side. F.eks. bøger side om side med manuskripterne til disse bøger. I nogen tilfælde, når der er tale om berømte forfattere, vil hver eneste lille manuskriptændring være af interesse for fremtidige forskere, i de fleste tilfælde vil manuskripterne være uden interesse. Ligesom der i det nuværende samfund prioriteres hvor mange ressourcer, der kan anvendes på hvad, således må der også i fremtiden investeres i ressourcer i formidlingen af mere væsentlige dokumenter end af mindre væsentlige. Den minutiøse registrering af manuskripter og udgavevariationer foretages af de mennesker, der udgiver tekstkritiske udgaver og kildelitteratur. Der er en arbejdsdeling imellem bibliotekernes registreringer og sådant tekstkritisk arbejde. Det er ikke helt den samme form for ekspertise, der er involveret. (Når en tekstkritiker så har publiceret sit arbejde, registreres og formidles det typisk af BDI-sektoren).

Delkonklusion 2:

Udvidelsen af den bibliografiske kontrol fra at omfatte publicerede dokumenter til også at omfatte ikke-publicerede dokumenter indebærer andre problematikker og principper, som den traditionelle nationalbibliografiske organisationsmodel nok i endnu mindre udstrækning er gearret til at varetage.

Lad os nu gå fra dokumenter til „information“. Vores førsteårsstuderende mente jo at bibliotekarere skulle indekserer *al information* på Internettet. Når begrebet „information“ anvendes på denne måde adskiller det sig klart fra den betydning, hvor information måles i bits og bytes. Det, der er tale om i bibliografisk henseende er tekster eller dokumenter i en udvidet betydning af dette ord, inkluderende elektroniske dokumenter, billedfiler, homepages etc.(se Hjørland,

2000a). Ideen om en komplet registrering minder om et forskningsprojekt, der er omtalt i den informationsvidenskabelige litteratur som Projektet LIBRU (Mariscal, 1993). Heri opstilles den målsætning, at man skal kvantificere *al information i verden*. De nævner, at det drejer sig om kort, gravsten, lysbilleder, annoncer, lydoptagelser, videooptagelser, disputatser og afhandlinger, periodika, papyrus, knogler, julekort etc. Men hvordan vil Mariscal dog bære sig ad med at afgrænse hvad der er „information“. Er det ikke sådan at alle genstande og processer kan være informative. At f.eks. en træstup i sine årringe indeholder information om træets alder, årstidernes nedbør osv.? Eller at ethvert menneskeligt produkt informerer om sit ophav, om sin kultur o.m.a.? Er det ikke problematisk på denne måde at operere med „information“, som noget objektivt, der kan tælles? Er information ikke tvært imod noget, der informerer, noget der besvarer spørgsmål? Afhænger en given genstands eller et givent dokument „information“ ikke af de spørgsmål, der bliver stillet til det, og som det er med til at besvare? Kan en sten på en mark ikke fortælle f.eks. en arkæolog og en geolog to helt forskellige historier, der må integreres i to helt forskellige videnssystemer? Hvis vi tager Mariscals julekort, indeholder det da ikke uendelig megen information? Kan det ikke besvare spørgsmål om afsenderen såvel modtageren, såvel som om motivet og kunstneren, papirkvaliteten, farverne, frimærket og poststempellet? Dette eksempel kan synes ekstremt, men ligger der ikke ofte i BDI-professionen lignende forestillinger om fuldstændige og udtømmende samlinger og registreringer? Ligesom med tekstkritikerne ovenfor mener jeg helt grundlæggende, at der findes forskellige arbejdsdelinger i samfundet i forhold til „information“. Nu kan man definere information forskelligt, f.eks. som noget menneskeskabt eller noget fysisk. I begge tilfælde er der imidlertid om noget, der er meget bredere end BDI-fagets virkefelt.

Eksempel 1: Såfremt vi betragter lyset fra stjernerne som information om verdensrummets opbygning, så er det astronomerne (ikke bibliotekarerne), der er eksperter til at opspore, registrere, fortolke og organisere denne information¹⁶. Først

når astronomerne har skrevet bøger og artikler m.v. herom bliver informationen til dokumenter og dermed til genstand for BDI-fagets vidensorganiserende aktivitet.

Eksempel 2: Såfremt vi betragter potteskår, redskaber, menneskeknogler, bygninger m.v. som information, så er det bl.a. arkæologerne (ikke bibliotekarerne), der er eksperter til at opspore, registrere, fortolke og organisere denne information. Først når arkæologerne har skrevet bøger og artikler m.v. herom bliver informationen til dokumenter og dermed til genstand for BDI-fagets vidensorganiserende aktivitet

Eksempel 3: Såfremt vi betragter breve, traktater, arkivalier m.v. som information, så er det bl.a. historikerne (ikke bibliotekarerne), der er eksperter til at opspore, registrere, fortolke og organisere denne information. *Denne form for informationskilder er dokumenter, men ikke publikationer*. Først når historikerne har skrevet bøger og artikler m.v. herom bliver informationen til *publikationer* og dermed til genstand for BDI-fagets vidensorganiserende aktivitet.

Eksempel 4: Såfremt vi betragter kunstgenstande, malerier, skulpturer m.v. som information, så er det bl.a. kunsthistorikerne (ikke bibliotekarerne), der er eksperter til at opspore, registrere, fortolke og organisere denne information. Først når kunsthistorikerne har skrevet bøger og artikler m.v. herom bliver informationen til dokumenter og dermed til genstand for BDI-fagets vidensorganiserende aktivitet

Eksempel 5: Såfremt vi betragter personoplysninger og oplysninger om virksomheder og institutioner m.v. som information, så er det især vejviserne (ikke bibliotekarerne), der har varetaget arbejdet med at opspore, registrere, fortolke og organisere denne information. Først når der er udarbejdet vejvisere, biografiske håndbøger, telefonbøger m.v. herom bliver informationen til dokumenter og dermed til genstand for BDI-

fagets vidensorganiserende aktivitet. Ganske vist repræsenterer arbejdet med at organisere sådanne vejvisere næppe kvalifikationer, der går ud over, hvad bibliotekarer kan klare¹⁷, men prioriteringen af arbejdet er udviklet ud fra hvilke produkter, der er økonomisk rentable, dvs. det er nærmere bestemt af markedsvilkår end af en topstyret national målsætning.

Organiseringen af viden og information er derfor grundlæggende aktiviteter, der omfatter en arbejdsdeling mellem mange forskellige faggrupper og *foregår parallelt på flere planer*. BDI-professionens særlige opgaver og ekspertise i relation til vidensorganisation og informationssøgning bliver efter min opfattelse svær at forstå, hvis vi ikke fastholder *dokument- og publikationsbegrebet*. Det skal dog siges, at træstubbe, lys fra stjernerne, knogler, kunstgenstande og potteskår og andre informative genstande jo ikke i sig selv kan findes på Internettet. Der findes højst repræsentationer af dem, som jo godt kan betragtes som dokumenter. Sådanne repræsentationer („klargjorte kilder“) er i endnu højere grad end publikationer og arkivalier afhængige af fagspecifikke principper/kvalifikationer for fortolkning og organisation. Såfremt man selektivt udtager oplysninger fra dokumenter og sammenstiller disse oplysninger til nye dokumenter, er vi efter min opfattelse nede på et niveau, hvor vi kan tale om „information“. I informationsvidenskabens historie har der givet ligget forestillinger om at man ad teknisk eller mekanisk vej kan destillere information ud af dokumenter, og at dokumentbegrebet herved mister sin betydning. Dette opfatter jeg som en fatal fejlslutning (Jfr. Hjørland, 2000a). Også svar på faktuelle spørgsmål som „hvornår er Napoleon født“ bygger på oplysninger fra dokumenter, og en teori herom må indbefatte en teori om teksters autoritet, om primære og autoritative kilder (f.eks. Kirkebøger som primære kilder). I stedet for at operere med „fakta“, „viden“ og „information“ i informationssystemer er det vigtigt at operere med påstande om viden (knowledge claims) fordi dette principielt åbner for muligheden for vurderinger og korrektioner.

Mekanisk registrering af al information på nettet

Man regner med, at der pr. februar 1999 findes godt 15 terabytes information på Internettet (ca. 6 terabytes tekst efter fjernelse af HTML-koder, kommentarer og ekstra mellemrum) fordelt på ca. 800 millioner sider og denne mængde er hastigt voksende (mange biblioteker har f.eks. planer om indskanning af tidsskrifter og bøger i fuld tekst). Heraf har 83% et kommercielt indhold, 6% har et videnskabeligt eller uddannelsesmæssigt indhold og 1,5% har et pornografisk indhold¹⁸. De ti største søgemaskiner dækker hhv:

Dækningsprocenterne for de 10 største søgemaskiner på Internettet

16,0%	Nothern Light;
15,5%	Snap;
15,5%	AltaVista,
11,3%	HotBot,
8,5%	Microsoft;
8,0%	Infoseek;
7,8%	Google;
7,4%	Yahoo;
5,6%	Excite
2,5%	Lycos.

(NEC Research Institute)¹⁹

Disse søgemaskiner anvender elektroniske robotter (webcrawlers eller webspiders) til at trevle nettet for websider, registrere alle ordene i en database og følge sidernes links videre til andre websider. Denne metode bevirker at sider, der er centrale fordi mange er linked op til dem registreres effektivt, mens isolerede websider registreres meget ueffektivt. I de senere år er søgemaskinernes „bibliografiske kontrol“ faldet drastisk. I 1997 dækkede den bedste søgemaskine mere end 33%, i dag kun 16%. De 11 bedste søgemaskiner tilsammen dækker kun 42%, hvilket betyder at næsten 60% af informationerne på nettet ikke i dag kan findes via søgemaskinerne. Dette skyldes at nettet er vokset hurtigere end søgemaskinernes kapacitet. Uden at der ligger nogen bevidst politik bag, er søgemaskinernes

dækning ikke repræsentativ. Amerikanske websider har større synlighed end Europæiske, kommercielle websider har større synlighed end private, populære websider har større synlighed end upopulære sider og nye sider har svært ved at konkurrere med etablerede sider. Søgmaskinerne er også blevet sammenlignet med telefonbøger, hvor mange af bladene er revet ud. Disse problemer med søgemaskinernes kapacitet må man naturligvis løse rent teknologisk ved at gøre maskinerne mere effektive. Et problem er, hvorvidt gammel information på nettet skal gemmes af dokumentationshensyn. Den er vel i princippet opfattet af pligta everingslovene i f.eks. Danmark og Norge. Det er imidlertid næppe realistisk med dagens teknologi at håndhæve sådanne bestemmelser (og det er vel et uklart problem i hvilket omfang dette er ønskeligt eller påkrævet). Spørgsmålet er imidlertid hvorvidt automatiske teknikker kan sikre den bibliografiske kontrol eller om menneskelig registrering er nødvendig og om den er rentabel? Her står vi også overfor problemet med „information overload“ og med problemet om omkostningerne pr enhed når gennemsnitsenheden måske repræsenterer en meget lav kvalitet og værdi. Et hovedproblem ved den rent mekaniske registrering vil netop være, at den ikke kan skelne god kvalitet fra dårlig kvalitet, men som svar på et spørgsmål måske giver tusinder eller sågar millioner af websider (altså en ekstrem lav precision i visse typer af søgninger). Hermed er vi inde i relevansproblematikken og i spørgsmålet om hvordan viden (dvs. videnspåstande) skal repræsenteres for at sikre genfindingen optimalt. Vi kan foreløbigt konkludere, at de automatiske indexeringsformer må betragtes som parallelle til de tidligere omtalte former for vidensorganisering af forskere, tekstkritikere, BDI-personale, osv. Vores arbejdshypotese er, at der er problemer omkring relevansvurdering, hvor maskiner ikke kan erstatte menneskelig vurdering og fortolkning.

Delkonklusion 3:

Det vidensorganiserende arbejde i samfundet og på internettet foregår parallelt på tre planer under udøvelse af mange forskellige slags fagekspertise, hvoraf den bibliotekariske er én blandt mange²⁰.

To hovedtyper af principper kan siges at stå overfor hinanden. På den ene side formalistiske principper, der forsøger at basere organiseringen på objektive, indholdsneutrale og rationalistiske metoder. På den anden side indholdsorienterede principper, der bygger på relevans for givne brugergrupper, på pragmatiske kriterier og på værdimæssige kriterier. Fordi der knytter sig en betydelig usikkerhed til vurdering af dokumenters relevans og værdi, søges ofte løsninger, hvor konsekvenser af fejlslån søges elimineret, bl.a. ved indbyggelsen af omfattende bibliografisk kontrol baseret på formelle kriterier. Sådanne metoder bliver dog mere og mere problematiske jo mere omfattende registreringen skal være. Som udgangspunkt må vi gå ud fra, at de forskellige slags dokumenter på internettet ikke skal behandles ens, men skal behandles i overensstemmelse med de traditioner og principper (og af de faggrupper), der gennem århundreder er udviklet til at formidle dem. Fra dette udgangspunkt kan man naturligvis udvikle nye principper. Fra et BDI-perspektiv er det vigtigt at danne et overordnet billede og en overordnet teori om disse vidensorganiserende processer og at fastholde de væsentligste demokratiske og oplysningsmæssige traditioner og idealer bag bibliotekernes og videnskabernes virke. Den særlige BDI-faglige ekspertise og funktion skal efter min opfattelse knyttes til en videnssociologisk og epistemologisk teori²¹.

Ad II: Vidensrepræsentation og genfindingsmuligheder

Der eksisterer forskellige syn på hvordan viden bør repræsenteres.

- Vi så ovenfor at nogle mennesker forestiller sig, at én konsekvent klassifikation af alle dokumenter efter et standardiseret system (f.eks. UDK) vil være det bedste ideal at stræbe efter.
- Andre mennesker forestiller sig, at hvis alle tekster i fremtiden bliver tilgængelige fuldt tekst på Internettet, så behøver man ikke foretage nogen form for emneanalyse, klassifikation, indexering eller lignende. Man må kunne lave

nogle kunstige intelligente agenter, der kan klare ethvert søgebehov (og måske kende vore behov bedre end os selv?).

- Atter andre mennesker mener at behovene måske kan klares ved hjælp af bibliometriske metoder (citationssøgning m.v.), eller
- at der bør udvikles thesauri, semantiske net osv.

Ideerne er mangfoldige, og der er en tendens til at forskellige personer eller organisationer går varmt ind for en bestemt af disse ideer, men har mindre forståelse for behovet for de andre. Spørgsmålet om hvilken repræsentation, der er den optimale til et givent formål (økonomien taget i betragtning) er naturligvis et spørgsmål, der skal belyses sagligt og videnskabeligt og ikke blot ud fra en forkærlighed eller fordom for eller imod den ene eller anden type af løsninger.

Delkonklusion 4:

Vores opgave som forskere og professionelle er især at belyse stærke og svage sider ved de forskellige muligheder, og at opstille begrebsapparater, teorier, metoder m.v., der kan udforske dette fra et overordnet perspektiv.

Objektive og subjektive søgemuligheder

Udviklingen i registreringen af dokumenter (og i BDI-forskningen) har i vid udstrækning været båret af udviklingen indenfor informationsteknologi. I Hjørland (2000d) opstiller jeg følgende fem udviklingstrin:

- I 1876- Manuel klassifikation og indexering af bøger i biblioteker
- I 1895- Klassifikation og indexering i fagbibliografier (dokumentalistbevægelsen og videnskabelig kommunikation)
- I 1950- Computerbaseret „information storage and retrieval“
- I 1963- Citationsindexering
- I 1990- Fuldtekst, hypertekst & Internet

Generelt mener jeg, at denne udvikling inddrager ere og ere søgeelementer, og præsterer bedre objektive søgemuligheder. En database, hvori et dokument f.eks. kun er registreret med ord fra titlen søgbare som „subject access points“ er objektivt fattigere end en database, hvor også ord fra abstracts og fuldtekst er søgbare. Da nogle felter kan producere „støj“ i søgeprocessen (dvs. non-relevante dokumenter) er det dog en forudsætning for at man kan tale om ubetinget objektive fordele, at man selv kan bestemme hvilke felter, man vil søge i. Givet denne forudsætning øges de objektive søgemuligheder i al fald indtil et givet dokument er registreret i „fuld tekst“ (incl. citationer). Hvorvidt søgemulighederne kan forbedres yderligere, er et teoretisk spørgsmål. Det er min egen opfattelse at dette godt kan lade sig gøre fordi et givent dokument kan være implicit om sit eget emne, hvorfor indexører kan tilføje „value added information“ ved at eksplicitere sådan implicit emneinformation (Jfr. Hjørland, 1997). Hvilken slags „value added“ information, der kan og bør tilføjes er et spørgsmål, der varierer med teksttyper, brugerbehov osv. (Dette spørgsmål er stærkt underbelyst rent forskningsmæssigt, de este forskere synes rent intuitivt at have antaget at selve fuldtekstrepræsentationen er en fuldt tilstrækkelig repræsentation, og de har betragtet alle emneord m.v. som værende blot „semantiske kondensationer“. Søgeudbyttet afhænger ikke kun af de objektive søgemuligheder, men også af de subjektive, dvs. af søgerens dygtighed. Denne er bl.a. er betinget af viden og færdighed med hensyn til databasernes opbygning, fagsproget, konventioner i såvel de repræsenterede dokumenter selv som i indeksering m.v. Det er f.eks. spørgsmål, om brugeren lægger den samme betydning i en given term som forfatteren til det repræsenterede dokument eller den informationsspecialist, der har beriget posten med en deskriptor eller klassifikationskode. Men skelner almindeligvis mellem to hovedformer for ekspertise: BDI-faglig ekspertise, der bl.a. består i generel søgeerfaring, erfaring med specifikke databaser, erfaring med forskellige typer „subject access points“ mv. og domænespecifik ekspertise, der især består i kendskab til faglige begreber og teorier. Der eksisterer mange mellemformer, og disse former for

ekspertise kan ikke blot ses som „udvendige“. En teori om subjektive søgeforudsætninger er relateret til teorier om informationssøgningsadfærd, der må inddrage det indhold og de kontekster, hvori søgningen foregår (jfr. Hjørland, 2000e). I biblioteks- og informationsvidenskabens forsøger man at måle værdien af forskellige former for indeksering og vidensrepræsentation. Man måler hhv. „recall“²², dvs. hvor mange af de relevante dokumenter, der faktisk identificeres i søgeprocessen og „precision“²³, dvs. hvor meget „støj“, der kommer med. Begge disse mål forudsætter, at man kan objektivt kan bestemme, hvad der er relevant (hvilket mig bekendt aldrig har været betragtet som et videnskabsteoretisk problem omkring „paradigmer“ og forhåndsantagelser, jf. Hjørland, 2000). Selvom det efter min opfattelse er problematisk at operere med et sådant objektivistisk, positivistisk relevansbegreb, så finder jeg at denne metodologi kan være hensigtsmæssig til at måle den relative værdi af f.eks. titelord, indekstermer, ord fra abstracts, ord fra selve artikelteksten (i fuldtekstbaser), af klassifikationskoder, af søkning i artiklers litteraturlister og andre „subject access points“ for søgeudbyttet. Forskningen tyder stærkt på, at forskellige metoder kompletterer hinanden og at ingen separat metode (inkl. fuldtekstsøgning er tilstrækkelig i sig selv). Forskellige søgemetoder giver som regel overlappende resultat, og nogle metoder er mest egnede til visse typer af spørgsmål (og til „browsing“, „navigering“ og lignende). Vi har en vis viden om statistiske sammenhæng, men står dog endnu i dag med en ret ufuldstændig viden, når det gælder de kvalitative sammenhænge. Der eksisterer et stort behov for forskning, som kan afsløre mekanismerne bag sammenhæng af denne art. Man kan eksempelvis sammenligne citationssøgning med termsøgning i databaser. Det er to avancerede måder at organisere og søge dokumenter på. Begge metoder har deres store fortalere og fortrin, og bibliotekarer skal naturligvis beherske dem begge. Teoretisk er det imidlertid vigtigt, at vi kan sige noget om hvilken af disse metoder, der er bedst egnede i en given situation, dvs. udvikle teorier om deres respektive styrker, svagheder og anvendelsesområder. En besvarelse af dette spørgsmål kræver en dybere

forståelse af såvel den måde sproget anvendes på og den måde litteratur citeres på i databaserne, og begge forhold kan variere fra kontekst til kontekst. Dette er objektive forhold, som forskningen kan kortlægge, og som brugere derefter kan oplæres i (dvs. subjektivt tilegne sig)²⁴. Vi er hermed dybt begravet i problemer omkring „database-semantik“, og vi er langt ud over simple tekniske løsninger, der er optimale uanset indhold og opgave.

Delkonklusion 5:

For fordi man ikke kan søge i noget, der ikke er der, er søgeudbyttet først og fremmest bestemt af objektive forhold, dvs. hvilke informationer om et givent dokument, der er repræsenteret i informationssystemet og gjort søgbare. Fremkomsten af søgbare databaser med abstracts i 1960'erne markerede en meget væsentlig forbedring i søgemulighederne. Herefter kommer citationsbasernes enestående supplerende muligheder, efterfulgt af fuldtekstbaser, hypertext og internet i 1990'erne. Forbedringen har hidtil i vidt omfang været båret af forbedringer i den tilgrundliggende teknologi og kun i mindre omfang af udviklingen af indexeringsteori m.v. Med fuldtekstbaserne er der nået et punkt, hvor forbedringer mere afhænger af kvalitative end af kvantitative forhold, hvorfor teorier om semantik o.lign. bliver centrale. Søgeudbyttet afhænger også af søgerens subjektive forudsætninger, der igen afhænger af en relevant undervisning, der igen afhænger af relevant forskning.

Kriterier for optimal repræsentation

Repræsentationen af dokumenter, viden og information skal sikre at „informationssystemet“ kan give optimale svar på de spørgsmål, der stilles til det. Dette er dog ikke det eneste krav. Herudover sikrer gode klassifikationssystemer, thesauri m.v. at brugeren kan danne sig et overblik over et emneområde, kan navigere og browse og dermed ikke blot få svar på færdige spørgsmål, men være med til at udvikle og raffinere spørgsmålene.

Delkonklusion 6:

Repræsentationen af dokumenter/viden/information skal sikre

- 1) at der gives optimale svar på spørgsmål
- 2) at der gives mulighed for feedback, navigering, browsing osv. med henblik på at udvikle og raffinere spørgsmål.

Tendensen i forskningen indenfor dette felt har været, at man dels har besvaret første spørgsmål ved at sige at „fakta“ repræsenterer det optimale svar på et spørgsmål og dels ved at betragte f.eks. et bibliografisk informationssystem isoleret fra andre informationssystemer, hvor der findes en rigdom af forskellige tiltag, der supplerer hinanden. Der er jo ingen tvivl om, at man i forbindelse med en bibliografisk søgning kan have behov for f.eks. at slå ord op, få dem oversat, læse om forfatteren, få kortfattet leksikalsk information osv. Sådanne oplysninger udarbejdes almindeligvis i mange forskellige sammenhænge. Nogle udarbejder fuldtekstbaser, andre bibliografiske databaser, nogle ordbøger, nogle leksika, nogle biografiske artikler, andre oversigtsartikler osv. En „ideel“ bibliografisk database kan derfor opfattes som en database, som på imperialistisk vis opsuger alle andre opgaver i sig selv. Derfor finder jeg, vi bør gå den anden vej og i lighed med f.eks. UNISIST-modellen (1971)²⁵ beskæftige os med hvilke specialiserede typer af arbejdsopgaver, der eksisterer og bør udvikles indenfor videnskabelig kommunikation (og vidensformidling i øvrigt). Først udfra et sådant overordnet billede kan man afgrænse de enkelte delsystemers funktion overfor hinanden, og hermed også definere, hvad de enkelte subsystemer ideelt set bør registrere/repræsentere. F.eks. fastlægge arbejdsdeling imellem nationalbibliografier, nationale og internationale fagbibliografier, thesauri, modersmålsordbøger og fagordbøger, vejvisere osv. (selv dette er et meget centralistisk og rationalistisk synspunkt, men jeg opfatter det som i det mindste værende et bedre skridt på vejen mod en mere åben og holistisk opfattelse).

Hvad angår synet på det optimale svar som „fakta“ (og måske oven i købet som en let tilgængelig, „smart“ præsentation af fakta), vil jeg som udgangspunkt mene, at det er det modsatte ideal, vi skal stræbe efter: at ingen viden må opfattes som

ubetvivlelig, og at al det, der er repræsenteret i systemerne derfor skal opfattes som påstande, ikke som fakta. Alle videnskabelige dokumenter bør opfattes som argumenter og påstande, der er udarbejdet af bestemte personer og institutioner på bestemte tider og steder ud fra bestemte forudsætninger, interesser og udgangspunkter. Foruden at repræsenterer dokumenternes egen selvforståelse, terminologi og betydning (der jo i vid udstrækning er imødekommet ved fuldtekst-repræsentation), bør informationssystemerne forsøge at repræsentere viden om dokumenterne, deres ophav, traditioner, implicite udgangspunkter osv. Efter denne opfattelse bliver den mest centrale værdi, som BDI-institutioner kan tilføre dokumentrepræsentationen, en epistemologisk/videnskabsteoretisk analyse. Den viden, som bibliotekarer har som særlig kompetence, hænger sammen med deres overblik over vidensproduktionen og brugergrupper samt bestemmelse af det enkelte dokumentets plads i overordnede vidensstrukturer. Videns(skabs)sociologisk og epistemologisk/videnskabsteoretisk viden er BDI-kernefag og i mange dokumenter, som bibliotekarer skal repræsentere, er denne viden udtalt, implicit eller problematisk. I relation til folkebibliotekspraksis mener jeg bibliotekarernes vigtigste fortolkningsbaggrund er knyttet til almen kultur & idéhistorie og til folkebibliotekernes egen oplysningstradition og erfaringsbase. I relation til forskningsbiblioteker finder jeg at faghistorie, videnskabsteori og videnskabssociologi er de vigtigste almene fag. Det er klart, at viden om alle de fag bibliotekarer skal repræsentere er en væsentlig forudsætning og ballast, men dette udgør speciel/specialiseret viden. Informationsteknologien øger generelt mulighederne for at søge og finde information på dramatisk vis. Når vi har nået op på det niveau, hvor vi nu befinder os, bliver problemerne omkring optimering af søgesystemerne mere subtile. Det er ikke længere først og fremmest et spørgsmål om at få mere emnedata repræsenteret eller få hurtigere computere, men i langt højere grad spørgsmål af kvalitativ art. Problemstillingen forskydes i mod mere humanistiske spørgsmål omkring betydning, relevans og sprog. Traditionelle metoder til kundskabsrepræsentation har overvejende bygget på em-

piristiske og rationalistiske principper. De har f.eks. bygget på statistisk behandling af termfrekvenser og på logiske opdelinger af vidensuniverset i gensidigt udelukkende og udtømmende klasser. Det er min opfattelse, at sådanne metoder må suppleres med andre metoder. Vidensorganisationen må bygge på studier af vidensproduktionen og den intellektuelle arbejdsdeling i samfundet. Den må omfatte studiet af vidensdomæner og af ”discourse communities”, dvs. sociale grupper, som udvikler fælles viden med byggende på fælles fagsprog, fælles kommunikationskanaler, databaser mv. Vi må udvikle metoder til domæneanalyser, som bl.a. inddrager sociologiske, bibliometriske, filosofiske og lingvistiske perspektiver. Problemstillingen forskydes hermed også på en anden måde. I stedet for en rent teknologisk problemstilling om en neutral effektivitet, stilles spørgsmålet om hvilket behov og hvilke interesser vidensorganisationen især skal tjene. Dette lyder måske umiddelbart overraskende og usandsynligt. Viden organiseres vel efter objektive kategorier såsom geografiske områder, tidsperioder, administrative inddelinger og videnskabelige taxonomier (som f.eks. kendt fra Linnés system). Ja, jeg er enig i, at valget af kategorisystem i nogen grad ligger fast og ligger udenfor informationspecialisters egen ind ydelse. Sverige har f.eks. en bestemt administrativ organisation, vilken er et objektivt faktum, som må være udgangspunkt for deler af vidensorganisationen. Et væsentligt synspunkt i moderne videnskabsteori er imidlertid at observationer er *teori-afhængige*, „Det sete afhænger af øjnene, der ser“. (Se Chalmers, 1995, kapitel 3). Videnskab ses ikke som en neutral indsamling af objektive fakta, men som en proces styret af antagelser, teorier, „paradigmer“ og interesser. Alle videnskabelige kategorisystemer kan derfor principielt problematiseres (selvom der er stor forskel på, hvor frugtbart det er at kritisere forskellige kategorisystemer). Dette synspunkt må også have væsentlige implikationer for vidensorganisationen og informationsvidenskabens og en af forskningens opgaver er udrede disse implikationer og formulere nye teorier om vidensorganisation, som er i overensstemmelse hermed. En ny svensk doktordisputats af

Joacim Hansson (1999) viser hvordan de svenske folkebibliotekers vidensorganisation (i SAB-systemet) hænger sammen med folkebibliotekernes samfundsmæssige opgaver, sådan som disse blev defineret af ideologiske faktorer omkring begyndelsen af det 20. Århundrede. Jeg tror at det er væsentligt at have oplysningstraditionens idealer og perspektiver i baggrunden, når man arbejder med vidensorganisation i en folkebibliotekssammenhæng. Jeg tror også at det betyder, at der er en væsentlig forskel mellem rent kommercielle principper for vidensorganisation (som f.eks. udformet hos Amazon.com) og offentligt oplysende principper (som udformet i folke- og forskningsbiblioteker). Selvom de sidstnævnte utvivlsomt kan lære meget, er deres formål ikke at formidle det mest populære, ikke blot at synliggøre det, der i forvejen er mest synligt, men derimod at formidle demokratiske og kulturelle værdier. Dette må bl.a. afspejle sig i hvilke dokumenter, der henvises til, eller rettere: i den relative synlighed som de enkelte dokumenter får i systemerne. Fælles for vidensorganisationen i folke- og forskningsbiblioteker i forhold til kommercielle systemer bliver en opprioritering af de „paradigmer“, dvs. af de forskellige teoretiske og ideologiske synspunkter, der ligger bag givne dokumenter og informationer. Ved at synliggøre disse giver man brugerne mulighed for at blive bevidst om dem, og dermed også at foretage selvstændige og bevidste valg i sin søgeproces.

Konklusion

Forskellige faggrupper vil konkurrere om at gøre sig gældende i fremtidens vidensorganisation, f.eks. på Internettet. Edb-folk vil især satse på algoritmiske løsninger til smartere søgemaskiner. Forlagsfolk vil især satse på at berige informationer, udgive publikationer, og organisere viden, f.eks. i tidsskrifter og serier. Internetboghandelen *Amazon.com* er et eksempel på en væsentlig nyskabelse hvor bøger, anmeldelser, søgemaskiner, best-seller statistik, diskussionsklubber, rekommendationer baseret på tidligere køb og lignende eksisterer side om side, og nok også et eksempel på en service som tiltrækker sig

mange brugere - delvis på bibliotekernes bekostning som redskab til at identificere bøgerne. Forskerne selv vil organisere viden efter deres disciplinære tilhørsforhold efter konferencer, og i deres referencelister og hyperlinks indbygge et vidensorganiserende net af forbindelser til andre forskere, institutioner og publikationer. BDI-professionen er den profession, hvis uddannelse og oplæring giver de bredeste forudsætninger (svarende til denne artikels indhold). Jeg har ikke i denne artikel givet noget definitivt nej til spørgsmålet i overskriften. Jeg mener ikke, der er nogle bestemte grænser for, hvad vi kan stræbe efter at uddanne bibliotekarer til. Jeg har derimod implicit opfordret til respekt for andre faggruppers særlige ekspertise og til en vis ydmyghed overfor det store ansvar, der ligger i at have ansvar for forvaltningen af offentlige vidensressourcer. Fremfor alt må man ikke automatisk opfatte sig som kvalificeret, men man må anstrenge sig for at blive det, dvs. for at kunne yde en indsats, der gør en forskel, og en forskel, man kan være bekendt.

Noter

1. Artiklen er baseret på et foredrag holdt i Borås 7. februar 2000, og er tidligere bragt i Human IT.
2. Begrebet „bibliotekarer“ her anvendt i videste forstand om personer, der har en BDI-faglig uddannelse, incl. forskningsbibliotekarer, dokumentalister og informationspecialister. Hvilke kvalifikationer, der er behov for, er dybest set et teoretisk problem, som også dette indlæg forsøger at belyse.
3. Den Universelle Decimalklassifikation (UDK) blev udviklet af de belgiske dokumentalister Paul Otlet og Henri La Fontaine og først udgivet 1905-1907. Paul Otlet & La Fontaine omformer i denne forbindelse det mest udbredte biblioteks-klassifikationssystem Dewey „DDC“ til et langt mere udbygget og videnskabeligt orienteret system „UDC“, dvs. de grundlægger klassifikationssystemet „Universal Decimal Classification“, der stadig anvendes (også i danske og svenske biblioteker).
4. De havde også andre drømme, bl.a. om en ny slags verdensencyclopopædi, der har forbindelse til tankerne om A WORLD BRAIN. (Jfr. Rayward, 1999).
5. Universal Bibliographic Control (UBC) blev i 1970'erne formuleret som et program under IFLA, International Federation of Library Associations & Institutions, med det mål at styrke udviklingen af et verdensomspændende system for kontrol med og udveksling af bibliografisk information. Formålet hermed var at gøre bibliografisk information universelt tilgængelig, og styrke mulighederne for udveksling og genbrug af bibliografiske data. Midlerne hertil var bl.a. udviklingen af fælles retningslinier for beskrivelse og klassifikation og kompatibilitet mellem de formater, der anvendes. UBC er baseret på nationale tiltag, specielt i form af nationalbibliografier, hvor de enkelte lande påtager sig ansvaret for bibliografisk kontrol med de dokumenter, der publiceres i landet. UBC-programmet hænger nøje sammen med et andet af IFLA's store programmer: Universal Availability of Publications (UAP), hvis formål det er at sikre alle brugere den videst mulige adgang til selve dokumenterne, d.v.s. fremme deres tilgængelighed ved lån eller på anden vis, nationalt såvel som internationalt. UBC-systemet er bl.a. beskrevet i Anderson (1974 & 1982).
[Denne note er en forkortet version af en artikel, min kollega *Mona Madsen* har skrevet til den af mig redigerede bog *Informationsvidenskabelige grundbegreber* (supplement 1996)].
6. Forfatteren har arbejdet på Det kgl. Bibliotek i København, hvor de næsten alle er faktisk tilgængelige. De anvendes især til verifikation af bøger, hvis udgivelsesland kendes.

7. Det er faktisk mest et ophavsretsligt spørgsmål om producenterne vil frigive deres bibliografiske data. Det er vistnok vedtaget i Danmark. Det er vel også rimeligt at borgerne gratis kan søge i de data, de selv har betalt via skatten. Og en sådan åbenhed og tilgængelighed er vigtig for at andre projekter, personer og organisationer skal kunne genbruge disse data til nye formål uden at gentage for meget arbejde.
8. Hvad enten det er gratis eller med betaling til den pågældende udgiver af de enkelte delbaser og til databaseværten.
9. Såfremt dokumenter er registreret i en væsentlig bibliografi, er de omfattet af bibliografisk kontrol, uanset hvordan dokumentet er beskrevet. Den nationalbibliografiske og fagbibliografiske kontrol er to systemer, der ofte i praksis er ret uafhængige, i al fald i forhold til visse idealer. Central litteratur om bibliografisk kontrol er: Anderson (1974, 1982); Davinson (1981), International Congress on Universal Availability of Publications (1982), *Manual on bibliographic control*. (1983), Plassard (1994), Wilson (1968):
10. Den oprindelige kontrolinteresse kan siges at ytre sig ved tvangsaspektet, centralismen, langsommeligheden og bundetheden til tryk-teknikken. Hermed være ikke sagt, at disse forhold ikke kan være yderst nyttige funktioner i de samlede bestræbelser, blot at systemet ikke i udgangspunktet er designet ud fra brugerbehov.
11. Måske er det dybest set dette, der ligger bag det forhold, at det i Danmark ikke er nationalbiblioteket, men det privatejede Dansk Bibliotekscenter, der er hovedkraften bag nationalbibliografien?
12. Der findes undersøgelser af f.eks. dansk forsknings repræsentation (f.eks. Alstrup, 1996; Søgaard Larsen, 1993 & 1994). Flere sådanne undersøgelser ville være relevante forskningsprojekter .
13. Om den så skal være „nationalbibliografisk“ eller „fagbibliografisk“ er et andet spørgsmål. Med hensyn til at sikre national bibliografisk kontrol forekommer Tyskland at være et forbillede. (Jfr. Hjørland, 1983).
14. Igen kan man komme med modargumenter. Hvor godt er disse databaser tilpasset deres målgrupper? Er det rationelt at de har stor overlapning og samtidig store huller i dækningen? Disse spørgsmål trænger til nærmere empirisk udforskning.
15. Nogle biblioteker har også arkiver, f.eks. indeholder Det kongelige Bibliotek i København manuskriptarkiver, billedarkiver, musikalearkiver mv. Nogle biblioteksskoler i verden underviser og forsker også i arkivvæsen, f.eks. University College London, School of Library, Archive and Information Studies (SLAIA)
16. Hvordan astronomer og andre forskere opsporer og registrerer information er et spørgsmål, der hører til under videnskabsteori og epistemologi. Der findes her mærkværdige empiristiske teorier, der forudsætter objektive og neutrale registreringer af sansedata, rene, forudsætningsløse beskrivelser. Det er indlysende, at en stillingtagen til sådanne teorier er afgørende for enhver videnskab og ikke mindst for informationsvidenskaben.
17. Omvendt går det måske heller ikke ud over hvad så mange andre faggrupper kan klare. Begrundelsen for at én faggruppe bør foretrakkes til en bestemt slags arbejde bør naturligvis være teoretisk begrundet
18. Kun 34% af hjemmesiderne anvender simple HTML „keywords“ og „description“ metatags, og kun 0,3% anvender the Dublin Core metadata standard.
- 19 Jfr. Lawrence & Giles (1999).

20. Bibliotekarisk ekspertise er naturligvis ofte opdelt i specialer, hvorfor man kan sige at den ikke udgør én men mange former for ekspertise. Sådanne specialer kan være fagspecifikke (f.eks. musikbibliotekarer) eller opdelt på andre måder (børnebibliotekarer, katalogiseringseksperter etc.
21. BDI-faglige registreringer af, hvad der produceres af litteratur og hvad der udlånes/efterspørges er et basalt udtryk for den implicite videnssociologiske baggrund som faget efter min opfattelse bygger på.
22. „Recall“ (Genfindingsforholdet) er et mål for et informationssystems ydelse. Først anvendt i „Cranfield II-eksperimentet“ (Cleverdon et al, 1966). Recall defineres ved forholdet mellem relevante dokumenter, der faktisk er fremfundet i forhold til alle relevante dokumenter i det pågældende system: $\text{Recall} = a : (a + c) \times 100\%$, hvor a = antallet af fremfundne, relevante dokumenter, c = antallet af ikke-fremfundne, relevante dokumenter (betegnes nogle gange „stilhed“; engelsk: „silence“). „Recall“ er således et udtryk for hvor *udtømmende* en dokument søgning er. „Recall“ udgør sammen med *„Precision“ de mest udbredte mål (og begreber) indenfor måling af genfindingssystemers ydelser. I informationssøgning står en lang række strategier til disposition for at øge genfindingsforholdet, f.eks. søgning i såvel „naturligt sprog“-felter som i kontrollerede emnefelter, søgning på synonyme udtryk (i naturligt sprog), søgning på overbegreber, begrænsning i anvendelsen af logisk „og“ i boolesk logik etc. Sådanne strategier er dog sjældent omkostningsfrie, men har en tilbøjelighed til at mindske „precision“.
23. „Precision“ (*Præcision*) er et mål for et informationssystems ydelse. Først anvendt i den såkaldte „Cranfield II-undersøgelse“ (Cleverdon et al., 1966). Precision defineres ved forholdet mellem relevante dokumenter, der faktisk er fremfundet i forhold til alle fremfundne dokumenter i den på-

gældende søgning: $\text{Precision} = a : (a + b) \times 100\%$, hvor a = antallet af fremfundne, relevante dokumenter, b = antallet af fremfundne, ikke-relevante dokumenter. »Precision« er således et udtryk for hvor meget støj en dokument søgning indebærer. I ældre litteratur betegnes »precision« ofte for „relevance“. Præcision bør ikke forveksles med begrebet *„relevans“. Sidstnævnte er et begreb, der er en forudsætning for, at man kan måle »precision«. I informationssøgning står en lang række strategier til disposition for at øge præcisionen, f.eks. begrænse søgningen til bestemte felter (titel, „deskriptor“ og „identifier“ benyttes ofte, når søgning i abstracts eller fuldtekst indebærer for lav precision), søgning på mere veldefinerede termer etc. Sådanne strategier er dog sjældent omkostningsfrie, men har en tilbøjelighed til at mindske *„Recall“ (genfindingsforholdet).

24. Brugere kan naturligvis også lære på baggrund af deres egne erfaringer.
25. Der er hårdt brug for en opdatering af UNISIST 1971 af mange grunde. Bl.a. fordi den aldrig har været tænkt til at dække områder udenfor naturvidenskab og teknik.

Referencer

Alstrup, Erik (1996). Danske historikere i internationale fagbibliografier. En kvantitativ og kvalitativ undersøgelse af dansk historieforsknings repræsentation i internationale fagbibliografier i tidsrummet 1977-1994. *Biblioteksarbejde nr. 46*, side 19-48.

Anderson, Dorothy (1974): *Universal Bibliographic Control, a long-term policy, a plan for action*. PULACH, München: Saur.

Anderson, Dorothy (1982): *UBC. A survey of Universal Bibliographic Control*. London: IFLA International Office for UBC. (Occasional Papers; 10).

Chalmers, Alan F.: *Hvad er videnskab? En indføring i moderne videnskabsteori*. Oversat fra „What is This

Thing called Science“ (University of Queenland Press, 1976 & 1982) af Gitte Lyngs. København: Gyldendal, 1995. 272 sider. (Serie: Gyldendal Filosofi).

Cleverdon, C.W.; J.Mills & E.M.Keen (1966): *Factors determining the performance of indexing systems*. Vol. 1-2. Cranfield: College of Aeronautics.

Davinson, Donald (1981). *Bibliographic control*. 2nd ed. London: Clive Bingley.

Hansson, J. (1999). *Klassifikation, bibliotek och samhälle. En kritisk hermeneutisk studie av „Klassifikationssystem för svenska bibliotek“*. Borås: Valfrid.

Hjørland; B. (1983) Informations- og dokumentationscentre: Den tyske model. *DF-revy. Danmarks forskningsbiblioteksforenings blad*, 6, nr. 3, side 49-55.

Hjørland, B. (1997) *Information Seeking and Subject Representation. An Activity-theoretical approach to Information Science*. Westport & London: Greenwood Press.

Hjørland, Birger (1998). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25(1/2), 16-31.

Hjørland, B. (2000a). Documents, Memory Institutions, and Information Science. *Journal of Documentation*, vol. 56(1), 27-41.

Hjørland, B. (2000b). Birger Hjørland. Professor i Biblioteks- og Informationsvetenskap. Side 12-15 i: *Installation av Professorer vid Högskolan i Borås Våren 2000*. Borås: Högskolan i Borås.

Hjørland, B. (2000c). Kunskapsorganisation. Installationsforelæsning i Borås 17 marts 2000. IN PRESS; Tidsskrift för Dokumentation, (også på <http://www.db.dk/bh/instalfor.htm>).

Hjørland, B. (2000d). Library and information science:

practice, theory, and philosophical basis. *Information Processing and Management*, 36(3), 501-531.

Hjørland, B. (2000e). Information Seeking Behaviour. What Should a General Theory Look Like? ISIC2000.

International Congress on Universal Availability of Publications (1982). 3-7 May 1982. Main working document including annotated programme and summary. Paris: Unesco. (PGI-82/UAP/2).

Lawrence, Steve & Giles, Lee (1999). Accessibility and Distribution of Information on the Web. *Nature*, 400, 107-109 resume: www.wwwmetrics.com (jfr. Politiken, 10.2.2000 3. Sekt. Side 2+4) NECs rapport.

Mariscal, A. R. (1993). Quantification of information. *International Forum on Information and Documentation*, 18(3-4), 21-23.

Manual on bibliographic control. (1983). Compiled by the IFLA International Office for UBC. General Information Programme. Paris: Unesco. (PGI-83/WS/8)

Plassard, Marie-France (1994): The IFLA Core Programme for Universal Bibliographic Control and International MARC (UBCIM). Recent developments and current state. I: *Alexandria. Vol 6(2)*, 145-153.

Rayward, W.B. (1999). H.G. Wells's idea of a world brain: A critical reassessment. *Journal of the American Society for Information Science*, 50(7), 557-573

Søgaard Larsen, Michael (1994). Dansk pædagogisk forskningsrepræsentation i internationale databaser 1-2. *Biblioteksarbejde* nr. 38, 1993, 14.årg., side 71-80 & *Biblioteksarbejde* nr. 41, 15.årg., side 55-64.

Unisist (1971). Study Report on the feasibility of a World Science Information System. By the United Nations Educational, Scientific and Cultural Organization and the International Council of Scientific Unions. Paris, Unesco

Wilson, Patrick (1968). *Two Kinds of Power. An Essay on Bibliographical Control*. Berkley: University of California Press, 1968. (Publications in Librarianship, 5).