

Indekseringsprocessen

Konsistensmål til sammenligning af tilgange til emnebestemmelse og emnebeskrivelse

Af Karen Birgitte Philipson

Abstract

Dette er første, overvejende kvantitative del, af et empirisk studie til belysning af indekseringsprocessen. De deltagende indeksører har beskrevet, hvordan de fandt frem til emner og tildelte emneord. Anden del er planlagt som en kvalitativ analyse af samme materiale. Et centralt spørgsmål er, hvorvidt forskellige overordnede indekseringstilgange resulterer i konstaterbare forskelle.

To grupper indeksører indekserede samme tekst og beskrev processen. Første gruppe brugte fire tilgange, anden gruppe én tilgang hver. Indekseringerne blev sammenlignet gennem optælling og konsistensmål..

Analysen viste generelt højere konsistens på emne-niveau end på emneordsniveau. Anvendelse af en tesaurus gav lidt bedre konsistens end frit tildelte termer, men var vanskelig at bruge for de forholdsvis uerfarne indeksører.

Undersøgelsen viste, at indeksørerne kunne følge de enkelte tilgange og at de resulterede i såvel forskellige emnebestemmelser som forskellige emnebeskrivelser.

Indledning

Emnebeskrivelse er en af de ældste discipliner indenfor biblioteks- og informationsvidenskab, men trods forskning i området (fx Sauperl, 2002; Chu & O'Brien, 1993) er det stadig noget af et mysterium, hvordan indeksører egentlig bærer sig ad med at nå fra dokument til et sæt emneord. Indekseringsvejledninger, som fx DS/INF100, angiver en metode, som indeksørerne kan følge, men det giver ingen garanti for at flere vil nå til samme resultat.

I praksis vil der normalt findes en bagvedliggende hensigt med indekseringen, en form for "ramme" eller nogle overordnede principper for emnebestemmelse. Langridge (1989, s. 1) kalder emnebestemmelsen "*the most important and difficult part*", for det er dér grundlaget lægges for den endelige emnebeskrivelse og dermed genfinding, og samtidig er det den mindst belyste funktion. Udgangspunktet for denne undersøgelse er spørgsmålet om, hvorvidt og hvordan en sådan ramme påvirker indekseringen, og om forskellige rammer fører til forskellige emner og emnebeskrivelser. I det følgende vil disse overordnede principper blive omtalt som "tilgange".

Undersøgelsens grundlag er to empiriske studier, hvor 4 tilgange blev anvendt til indeksering af samme tekst. Der var behov for at kunne sammenligne effekten af de anvendte tilgange, og derfor blev konsistensmål inddraget. Konsistensmål kan give et overblik over om forskelle eksisterer måleligt, og vise størrelsen og arten af de eventuelle forskelle.

Karen Birgitte Philipson er lektor ved Danmarks Biblioteksskole, kbp@db.dk

Som det kvantitative mål konsistensmålet er, er det til gengæld begrænset, hvad det kan åbenbare af baggrunden for eventuelle forskelle. Konsistensberegninger har her primært haft til opgave at udgøre et struktureret grundlag til understøttelse af senere kvalitative analyser og fortolkninger af samme materiale.

Metoden er inspireret af Ricoeurs hermeneutiske tekstanalysemetode, som er en kombination af en struktureret tekstanalyse og en tekstudlægning (Ricoeur, 2002). Overført hertil, en kombination af en kvantitativ analyse af det indsamlede materiales ("teksternes") konstaterbare træk og den fortolkende tekstforståelse. Gennem optælling og strukturering af materialet i konsistensanalysen kan det konstateres, om der er målbare forskelle; det er denne første del af den samlede analyse, som diskuteres her.

Indekseringsprocessen

Indekseringsprocessen kan defineres som de handlinger, som finder sted, fra indeksøren begynder at orientere sig i dokumentet til den endelige emnebeskrivelse foreligger. De overvejelser, de gør sig, når emner bliver valgt ud blandt mangfoldigheden af ord eller tegn, og hvordan de bliver til emnebeskrivelse. *"There is little about how indexers decide what the subject of a document is, how they determine its aboutness. More interest is currently being shown in automatic indexing than in improving human indexing. Essentially, the subject analysis phase is lamented as an intellectual operation resistant to investigation."* (Todd, 1992, side 101).

Det har nu ikke afholdt alle fra at forsøge at beskrive indekseringsprocessens enkelte delprocesser, som regel 2-4 (fx Langridge, 1989; Chu and O'Brian, 1993; Lancaster, 2003). Lancaster (2003) opdeler processen i to trin.

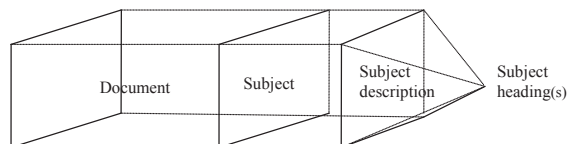
1. conceptual analysis
2. "translation"

Standarden DS/INF100 for indeksering, beskriver processen i tre faser:

- undersøge dokumentet og bestemme dets emner
- bestemme de hovedbegreber, som emnerne består af
- udtrykke emnets hovedbegreber i det valgte indekseringssprog (kilde)

Beskrivelsen opdeler også processen i en konceptuel fase og et konkret udtryk, men adskiller sig ved at opdele den konceptuelle analyse i to delprocesser. "Emner" må her forstås som alt det, som dokumentet omhandler, og "begreber" som overordnede benævnelser for "emnerne", altså i virkeligheden klasser af emner. Den skelnen diskuterede allerede Bliss (1929).

En grafisk, rumlig model af indekseringsprocessen er foreslået af Miksa (1983). Her opfattes indekseringsprocessen i tre trin, som forbinder fire elementer, nemlig dokumentet, dets emner, en emnebeskrivelse og til sidst emneord. Modellen opdeler, som ISO-modellen, den konceptuelle fase i en orienterende del, fulgt af en mere bevidst sproglig formulering af emneindholdet og endelig et udtryk i indekseringssproget.



Figur 1. Indekseringsprocessen

Mai (2000) modificerede modellen til en mere pileformet figur som vist ovenfor, for at præcisere at slutpunktet – emneord – er et meget snævert udtryksmiddel i forhold til det brede dokumentindhold. Begge forfattere opfatter indekseringsprocessen som en række fortolkninger. Den første fortolkning finder sted, når indeksøren orienterer sig i dokumentet. På denne første fortolkning bygger de øvrige, den ene oven på den anden: emnerne bestemmes ud fra dokumentanalysen, en mere præcis emnebeskrivelse foretages ud fra emnebestemmelsen, og endelig vælges emneord ud fra emnebeskrivelsen.

Indekseringsprocessen beskrives som en lineær proces, men spørgsmålet er, om emneindholdet faktisk opfattes i en kontinuerligt snævrere kanal, som indeksøren bruger som "referent", i Mais semiotiske terminologi; eller om indeksøren vil have den brede opfattelse af dokumentet fra dokumentanalysen i baghovedet under hele processen og "referere", altså vælge emneord ud fra denne fortolkning, i modsætning til mekanisk oversættelse af emnebeskrivelsen til de termer, som kan findes i vokabularet.

Mai understreger også netop, at for erfarne indekserers vedkommende må man formode, at indekseringsprocessen er mere helhedsorienteret. Det er endda blevet fremhævet som en faldgrube for erfarne indeksører straks at tænke i indekseringssproget og de begrænsninger, som ligger i det. (Lancaster, 2003). Helhedsorienteret eller lineær proces - det forbliver en antagelse indtil videre. Modellen har været en inspirationskilde til denne undersøgelse, hvor følge idéen i den bliver fulgt.

Metoder til emnebestemmelse

En metode til emnebestemmelse kan være at følge detaljerede retningslinjer og checklister. Eller det kan være indeksørens brug af sin egen sunde fornuft og evne til at skønne hvad der er vigtigt, mindre vigtigt og helt uvæsentligt for den senere genfindning og anvendelse af dokumentet. Eventuelt en blanding af det, som Farrow (1994) beskriver som en intuitiv proces, som bygger enten på indeksørens fornemmelse af dokumentet eller hans viden om verden udenfor dokumentet. Det kan være et indgående kendskab til domænet, brugerne og deres informationsbehov, samlingen osv., som Mai (2004) fx beskriver det i sin diskussion af den domænecentrerede tilgang overfor den dokumentcentrerede tilgang. Eller det kan dække over, hvad der lige falder indeksøren ind; et særlig udpræget eksempel på dette er social tagging, som det kan ses i fx LibraryThing.

Men ”sund fornuft”, hvad enten den er forankret i domænekendskab eller ej, må være baseret på en opfattelse af, *hvad* det er for en form for viden, man gerne vil fremhæve i emnebeskrivelsen og *hvordan* denne viden kan findes. Forskellige opfattelser af det svarer til forskellige tilgange eller principper for indeksering.

Om en regelstyret indeksering er mere eller mindre subjektiv og fortolkende end en principstyret kan diskuteres, ligesom det kan diskuteres om det indledende trin i indekseringsprocessen overhovedet kan standardiseres. *”Indexing guidelines cannot prescribe a standard practice in the determination of the subject matter; they can, however, standardize the expression of the subject matter in the indexing language.”* (Mai, 2004, s. 601). Antagelsen i denne artikel er heller ikke, at emnebestemmelse kan standardiseres, men at det er muligt at styre den i en ønsket retning, forudsat at indeksøren er bevidst om, hvilken form for emner, der skal tildeles.

Hjørland (1992) sammenligner søgen efter emner med søgen efter viden. ”Viden” og måden, hvorpå den kan opnås, er et erkendelsesteoretisk spørgsmål, og Hjørland anbefaler derfor at vi skaber bevidsthed om, hvilke former for viden, det egentlig er, vi giver brugerne adgang til via katalog eller database, når vi har en bestemt tilgang til indeksering og bruger bestemte metoder. Flere forfattere indenfor informationsvidenskaben (fx Mai, 2000; Mai 2004; Hjørland, 1992; Albrechtsen, 1993) har beskrevet forskellige indekseringstilgange og koblet dem til erkendelsesteori.

En måde at inddele indekseringstilgange, som er forbundet med det erkendelsesteoretiske aspekt, er efter, hvorvidt emnerne forventes at kunne findes direkte i dokumentet eller om de skal findes et sted ”udenfor”. Med Farrows begreber: *perceptuel indeksering* henholdsvis *konceptuel indeksering* (Farrow, 1994). Grænsen er flydende. ”Perceptuel” vil sige, at emnerne kan opfattes med sanserne, ”indeksørens fornemmelse af dokumentet”. Den enkleste form er at tælle forekomster af ord, og udnævne de hyppigt forekommende til emner. Tilgangen forudsætter tillid til, at emner – eller viden – er indbygget i dokumentet selv og bare venter på at blive fundet, og den uproblematisk sprogopfattelse, at ord kan betegne et givet fænomen korrekt, uanset kontekst.

I en mere kompliceret perceptuel tilgang søger indeksøren efter en tilkendegivelse af emnet i dokumentet. Forfatteren kan ligefrem have beskrevet det, måske i et forord. Trods større grad af intellektuel medvirken bygger tilgangen også på en forventning om, at emnet vil fremgå med mere eller mindre rene ord. Også her findes en opfattelse af viden- og emnebegrebet, som er nærmest uden kontekst, måske endda objektiv.

Disse tilgange har tilfælles at være fokuseret på dokumentet i sig selv. Andre tilgange har også dokumentet i centrum, men stiller desuden krav til indeksørens evne til at ræsonnere sig til emnerne, eller eventuelt anvende en checkliste. Checklistens mange punkter i DS/INF100 drejer sig udelukkende om dokumentet: ”handler dokumentet om en genstand, en aktivitet, en genstand for aktiviteten” osv. De tre tilgange kan henregnes til de ”dokumentorienterede” tilgange, hvor dokumentet formodes at bære information nok i sig selv.

”Konceptuel” indeksering medfører en intellektuel fortolkning af dokumentet, i forhold til en kontekst. Konteksten består muligvis bare af indekserens person, hvorved indekseringen vil være en subjektiv proces, men kan udvides til at omfatte forfatterskabet eller samlingen, brugerne, arbejdsopgaverne, organisationen, samfundet etc., og tyder så på en holdning om, at emnet hverken er fuldkommen objektivt eller subjektivt, men må forstås i sammenhæng med de omgivelser, som definerer det. Som en yderligere komplikation er et dokument emne efter konceptuel indeksering ikke statisk, men ændrer sig i lyset af videnudviklingen inden for det aktuelle fagområde og skiftende opfattelser af, hvad der er væsentligt og mindre væsentligt.

Mai (2000) definerer fem tilgange, fra perceptuel i den enkleste form til den usædvanligt brugerorienterede tilgang, som han kalder forespørgselsorienteret. En opregning fra det enkle til det meget komplicerede, hvor den ”indholdsorienterede” tilgang, som fx den tredje af de ovenfor beskrevne, danner skillelinjen mellem perceptuel og konceptuel indeksering. Der sker altså en udvikling fra det ene yderpunkts tilsyneladende objektivisme, mod stadig mere subjektive og fortolkende retninger. Senere har Mai tilføjet en ekstra tilgang, nemlig den domænecentrerede tilgang, hvor dokumentanalysen er baseret på analyse af domænet, for at sikre at en vurdering af emneindholdet i overensstemmelse med brugernes diskurs (Mai, 2004). Den domænecentrerede tilgang indebærer ifølge Mai en indekseringsproces i mere end de få trin, som den ellers ofte reduceres til.

Wilsons indekseringstilgange

De tilgange, som er anvendt i undersøgelsen, er de fire, som Patrick Wilson (1968) har beskrevet. De strækker sig tilsammen over hele spektret fra objektiv til subjektiv. Wilsons tilgange er langt fra at være detaljerede retningslinjer, men han forklarer ret omhyggeligt hensigten med hver af dem.

1. I ”The purposive way” er målet for indekseren at finde frem til forfatterens hensigt med værket. Hvis forfatteren selv giver udtryk sit mål, fx i en prægnant titel eller et forord, steder som også DS/INF100 anbefaler at kikke, kan indekseren være fristet til at holde sig til det. Wilson advarer mod for hurtige konklusioner, for en nærmere analyse af dokumentet kan afsløre mål, som ikke er nævnt direkte, eller at forfatteren ikke har levet op til

sine egne hensigter. Tilgangen kan – ideelt – siges at føre til en langtidsholdbar indeksering, for så vidt at forfatterens hensigt ikke pludselig kan ændres indenfor værket. Om indekseringen vil være langtidsnyttig er til gengæld usikkert; det vil forudsætte, at brugerne fortsat vil være interesserede i de samme aspekter af værket.

2. ”The figure-ground way” beskrives som en tilgang, hvor ”det iøjnefaldende” svarer til emnet. Wilson sammenligner det at forholde sig til et skriftligt værk med det at se en film: nogle figurer er i centrum, andre optræder kortvarigt. Nærbilleder af den centrale figur forstærker det dominerende præg, men ellers stammer det især fra det allestedsværende. Lighedstegn mellem det dominerende og emne er en tilgang med faldgruber. Det er fx muligt for forskellige indekserere, at hæfte sig ved forskellige dele af teksten, afhængig af, hvad den individuelle indekserer er klar til at bemærke, så det bliver indekserens subjektive fortolkning af, som afgør hvad emnet er.
3. ”The objective way” kaldes sådan, fordi den indebærer en kvantificering af teksten. Også her drejer det sig om at finde emnet i det mest dominerende, men nu baseret på frekvens, i simpleste form en ordoptælling. Selv med en god stopordsliste, er ordfrekvens ikke nødvendigvis mest sigende for emneindholdet, så Wilson udvider optællingen til også at omfatte henvisninger i teksten. Fx kan en person optræde ved navn, med også med henvisninger som ”han”, ”ministeren” etc. Ting, der refereres til, kan grupperes på forskellige måder, og en stor gruppe vil være kandidat til positionen som emne. Efter Wilsons mening er bestemmelse af emner ved at tælle ord, er en håbløs metode, medmindre de indirekte henvisninger medregnes og der grupperes; men eftersom der er mange måder at gruppere på, og grupperne i øvrigt dannes efter de begreber, som indekseren besidder, bliver den ”objektive tilgang” således alt andet end objektiv.
4. Den 4. tilgang kaldes ”Appeal to unity” eller ”Rules of selection and rejection”. I modsætning til 1. tilgang, hvor forfatterens hensigt er det centrale, er det her hvad forfatteren faktisk har gjort, som indekseren skal forsøge at få mening i. Udgangspunktet er den ideelle tanke, at en forfatter har medtaget lige nøjagtigt det stof, som er nødven-

digte for at få belyst tekstens emne og få det til at fremstå som en helhed. Kan vi finde en "regel", som binder teksten sammen til en helhed, og som kan forklare de forskellige dele af den, har vi identificeret emnet. I modsætning til 2. tilgang, som udvælger emner ud fra mest dominerende position og lader resten ligge, er denne tilgangs formål at finde helheden, som også tager de mindre iøjnefaldende dele i betragtning. Men "*discovering how the writing hangs together may be discovery of one out of several possible ways in which we can make it seem reasonably unified...*" (Wilson, 1968, s. 88).

Wilson konstaterer, at der er flere måder at nå frem til en emnebeskrivelse. De synes alle egnede, skønt der ikke er grund til at tro, at de vil føre til samme emnebeskrivelse, selvom de nok ofte vil gøre det. En enkelt indeksør vil måske stå med 3-4 forskellige resultater; flere indeksører kan opnå forskellige resultater med samme metode, og det er ikke til at sige med bestemt, hvilket er det rigtige. "*The notion of the subject is indeterminate*", skriver Wilson (Wilson, 1968, s. 89), og er således på linje med Bloomfield, som heller ikke har stor tillid til "korrekt" indeksering: "*Machine indexing is rotten and human indexing is capricious.*" (Bloomfield, 2001, side 64).

Wilsons konklusion vedrørende tilgangenes ens eller forskellige resultater, har virket som en udfordring til at afprøve i praksis, om det vil være muligt at give indeksører nogle overordnede principper at arbejde ud fra, og derigennem styre indekseringen i en ønsket retning og opnå høj konsistens.

De fire tilgange er blevet kritiseret (fx af Mai, 2000) for at være udelukkende dokumentorienterede. I sit essay sætter Wilson ganske rigtigt ikke indekseringsprocessen i forbindelse med nogen konkret kontekst, som fx en bestemt brugergruppe eller en faglig disciplin. På den anden side indeholder to af tilgangene implicit en brugerorientering. Når indeksøren i 4. tilgang finder helheden i dokumentet, bliver der lagt en (faglig) synsvinkel på emnet. I 2. tilgang fungerer indeksøren selv som bruger ved at indeksere de aspekter af dokumentet, som er subjektivt fremtrædende for ham. Hvorvidt begrebet subjektivitet dybest set er tilknyttet et enkelt individ, kan diskuteres. "Individuelle" synsvinkler vil, ud fra en hermeneutisk synsvinkel, være påvirket af individets baggrund, kultur, samfund, uddannelse osv.

Konsistens i indekseringen

Denne "modificerede subjektivitet" understøttes af Sauperls (2004) studier af 12 erfarne indeksørers arbejde. Sauperl opdagede, at indeksørerne udmærket var klar over, at der kan findes flere forskellige "meninger" i et dokument, og at de ikke kun brugte deres personlige forståelse, men fx gennem støtte fra lignende poster i katalogen, bevidst udviklede en "katalogiseringsmening". Det vigtigste for dem var at opbygge et fælles grundlag for katalogen. Det vil sige, at ensartethed, konsistens, i katalogen havde høj prioritet. Ifølge Lancaster er det eneste seriøse forsøg på at påvise sammenhængen mellem konsistens og kvalitet i betydningen: genfindingseffektivitet, det studie, som Leonard gennemførte i 1975 (Leonard, 1975, s. 32). Forsøget viste en "moderat til stærk" forbindelse. Sammenhængen mellem konsistens og recall viste "*clearly defined positive relationship.*" (Lancaster, 2003, s. 92-93). Når indeksørerne i Sauperls studier lagde vægt på konsistens kan det tyde på en generel opfattelse af, at kvalitet i katalogen blandt andet hænger sammen med konsistens.

Andre (fx Cooper, 1969; Todd, 1992) stiller sig dog mere skeptisk overfor brugen af konsistens som mål for kvalitet. Cooper med henvisning til, at selvom 3 ud af 4 indeksører er konsistente, er det på ingen måde sikkert, at de også tildeler de mest nyttige emneord, set fra en brugersynsvinkel. Dette kunne være tilfældet blandt Sauperls indeksører, for som hun skriver "*Although all 12 catalogers tried to provide user-friendly subject description, most of them did not consider the different uses someone could make of any particular book*" (Sauperl, 2004, side 61).

Miksa, Mai og mange andre opfatter indekseringsprocessen som en fortolkende proces. Det har konsistensstudier i tidens løb da også har bekræftet, høj konsistens opnås sjældent. "*It is quite clear that indexing is a subjective rather than an objective process.*" (Lancaster, 2003, side 68). Oplæring og socialisering til en bestemt konteksts formål og kultur m.v. har muligvis en normaliserende effekt på indekseringerne, og bliver regnet for én blandt mange faktorer, som påvirker indekseringskonsistens. (Lancaster, 2003, s. 76).

Konsistens i indekseringen drejer sig om, i hvilken udstrækning der er enighed blandt indeksører om hvilke emner, et dokument behandler, om hvilke

aspekter af dokumentet, som skal indekseres og hvilke emneord, som skal tildeles. Formålet med at undersøge konsistens er som regel at konstatere om to eller flere indekserer indenfor en bestemt institution opnår samme resultat. For store forskelle kan tyde på fejl. (Lancaster, 2003). Normalt arbejder indekserer i konsistensundersøgelser under ensartede vilkår, men i undersøgelsen her har forholdene for de to grupper af indekserer været noget forskellige.

Undersøgelsens design

Det empiriske materiale til undersøgelsen blev indsamlet i to omgange. I begge tilfælde var deltagerne studerende ved Danmarks Biblioteksskole. Da indekseringen fandt sted, havde de godt et års studier bag sig, dog kun 6 uger med diskussioner af emnebegrebet, indekseringsprocessen og typer af indekseringssprog. Indeksererne kan ikke karakteriseres som erfarne indekserer, sådan som fx Sauperls forsøgspersoner var det. De har dog tilstrækkelig erfaring med faget og dets terminologi til at kunne læse og forstå faglige tekster.

Mais beskrivelse af indekseringsprocessen (figur 1) blev brugt som model. Udgangspunktet er det ”brede” dokument. Indeksererne blev bedt om beskrive, for hver enkelt tilgang, hvordan de fandt frem til emnerne og skrive dem ned – i første omgang ikke emneord. Desuden skulle de beskrive, hvordan de valgte hvilke af emnerne, som skulle indgå i emnerepræsentationen. Endelig tildelte de ukontrollerede emneord.

I forhold til figur 1 kommer først dokumentet (1. element) som analyseres (1. proces), hvilket resulterer i ”emne” (2. element), fulgt af en emnebeskrivelse (2. proces og 3. element) – den mere formelle beskrivelse af emnet. Emnebeskrivelsen analyseres (3. proces) og til sidst tildeles emneord (4. element). Modellen er således i overensstemmelse med Wilsons adskillelse af emne og emnebeskrivelse.

I den første undersøgelse deltog 65 indekserer. De blev alle stillet overfor samme opgave, nemlig at indeksere en kort tekst (ca. 5 sider)¹ ud fra hver af Wilsons fire tilgange. På den måde kunne vi se, om én indekserer ville nå frem til samme eller forskellige indekseringer, sådan som Wilson diskuterede.

I den anden undersøgelse arbejdede indeksererne med kun én af de fire tilgange hver, men med samme

tekst. Herved kunne den anden af Wilsons muligheder afprøves, nemlig om flere indekserer vil nå til samme indekseringer med samme tilgang. 64 nye personer deltog, men altså kun 16 pr. tilgang. Der var en anden grund til at gentage undersøgelsen: når hver indekserer arbejder med flere tilgange, kan det muligvis være et problem at holde tilgangene adskilt. Det andet sæt materiale fungerer således også som en form for kontrol af det første sæt.

Analyse af det første materiale havde vist stor iderigdom hos indeksererne med hensyn til ukontrollerede termer. Flere forfattere (fx Sievert & Andrews, 1991; Leonard, 1975; Mai, 2004; Lancaster, 2004) mener, at brugen af et kontrolleret vokabular kan have en ”normaliserende” effekt på emnebeskrivelserne. Det kan i et vist omfang styre indekserens analyse af dokumentet og byder samtidig på et begrænset udvalg af indekseringstermer. (Sievert og Andrews, 1991).

Svenonius skriver om kontrollerede vokabularer, at fordi de indeholder semantiske kategorier, er de også ”categories of existence”. *”As such, they determine and limit what is indexable, by predetermining, in Kantian fashion, the perception of what a document can be about. By imposing a constraint on aboutness, facets contribute to defining the vocabulary domain of a subject language: if a candidate term does not belong to one of the facets of the language, it is inadmissible. This same restriction also serves to curtail language subjectivity and inconsistency in indexing.”* (Svenonius, 2001, s.) Som Svenonius påpeger, kan der være både fordele og ulemper ved brug af kontrolleret vokabular. På den ene side kan vokabularet begrænse indekserens frihed til at emnebestemme dokumentet, hvis de fra starten har i tankerne, hvordan det kan udtrykkes; i den aktuelle undersøgelse var indeksererne dog ikke på forhånd fortrolige med vokabularet, og ville derfor næppe lade sig begrænse af det i den indledende emnebestemmelse. På den anden side kan et kontrolleret vokabular begrænse den sproglige rigdom til færre, forhåbentlig velvalgte, udtryk for det samme. I den anden undersøgelse skulle indekserer derfor også tildele deskriptorer fra LISA tesaurus, primært for at se om begrænsningen i udtryksmidlet ville øge konsistens eller mindske antallet af emneord.

Konsistensmål blev inddraget for at få et grundlag for sammenligning. I forhold til formålet med undersøgelsen, som er at se tendenser i indekseringen, ikke

at finde konkrete fejl, valgte vi en kombination af ren optælling og simple procentberegninger.

Emner og emnebeskrivelse for hver tilgang blev skrevet ind i et skema. Skemaet gav et indledende overblik over materialet, og en mulighed for at finde tilbage til den enkelte indekssørs beskrivelse af indekseringsprocessen; det vil blive brugt i den kommende kvalitative analyse.

Resultater og diskussion

Den første optælling viste nogle problemer, flest sproglige, af såvel semantisk som morfologisk karakter. *“... concepts are mental constructs, and, as such, they have unclear boundaries. Do the terms Information and Retrieval represent one, two, or three concepts? It would depend on who was asked..... yet a decision must be made...”* (Svenonius, 2001, s. 133). Valg blev også truffet her: en række ”kriterier for match” blev opstillet. Kriterierne drejede sig hypigst om entals- eller flertalsform og om ”indexing history” og ”history of indexing” og tilsvarende er synonyme.

Et andet problem viste sig i forbindelse med ”objective way”, den 3. tilgang. Mange indekssører fulgte

Wilsons anvisninger på medtælling af henvisninger, og gruppering, men andre talte ord i teksten ved hjælp af et computerprogram. De fleste foretog derefter en intellektuel bearbejdning af de fremfundne termer, men fundene i 3. tilgang må tages med forbehold.

Indeksering med 4 tilgange pr. indekssør

Emneorderne blev optalt med anvendelse af matchkriterierne, og nogle forskelle mellem tilgangene viste sig umiddelbart. Emneord i forhold til hvor mange indekssører, som havde tildelt dem, blev omregnet til procent for at lette sammenligning. Et udsnit ses i tabel 1.

Konsistens på emneordsniveau er ikke høj. Den højeste konsistens indenfor én tilgang står ”indexing” for, med 86,2 % i 1. tilgang, lavest i 4. tilgang med 47,7 %.

Mange andre begreber viser stor spredning mellem tilgangene. Et par eksempler: i 1. tilgang er forfatterens hensigt med værket udgangspunktet, og mange har valgt emneord ud fra den sigende titel og undertitel. Titellordene er repræsenteret langt oftere end alle andre ord. For ”abstracting” var der 69,2 % konsistens i 1. tilgang, men væsentligt lavere i de an-

1. tilgang	%	2. tilgang	%	3. tilgang	%	4. tilgang	%
Abstracting	69,2	Abstracting	36,9	Abstracting	21,5	Abstracting	21,5
Abstracting history	12,3	Abstracting history	6,2	Abstracting history	1,5	Abstracting history	1,5
Abstracts	10,8	Abstracts	15,4	Abstracts	4,6	Abstracts	9,2
		Alexandria	1,5	Alexandria	1,5	Accessibility	1,5
Alphabet	4,6	Alphabet	13,8	Alphabet	23,1	Alexandria	
		Alphabetic	3,1	Alphabetic	23,1	Alphabet	10,8
		Alphabetic indexing	4,6	Alphabetic indexing		Alphabetic	
Alphabetic order	3,1	Alphabetic order	7,7	Alphabetic order	13,8	Alphabetic indexing	
		Alphabetic subject index	1,5	Alphabetic subject index		Alphabetic order	1,5
		Alphabetic subject indexing	1,5	Alphabetic subject indexing		Alphabetic subject index	
		Alphabetization	4,6	Alphabetization	7,7	Alphabetic subject indexing	
		Ancient history	1,5	Ancient history		Alphabetization	1,5
Antiquity	69,2	Antiquity	32,3	Antiquity	13,8	Ancient history	
...				Arrangement	1,5	Antiquity	30,8
		Back of the book indexes	1,5	Back of the book indexes		Arrangement	
						Back of the book indexes	

Tabel 1. Andel af indekssører, som har brugt emneordet, alfabetisk orden og i procent. Udsnit.

dre. Forfatteren bruger ordet i titlen på lige fod med "indexing". Når indekseren så foretager en mere personlig fortolkning af teksten i 2. tilgang, tæller ord som i 3. tilgang, eller søger efter sammenhæng i teksten i 4. tilgang, bliver det klart, at abstracting ikke er særlig fremtrædende i teksten. Omvendt forholder det sig med "alfabet" og variationer af det. Forfatteren har ikke selv fremhævet begrebet, så det optræder sjældent i 1. tilgang, men i 2. og 3. tilgang opdager indekserne begrebets betydning i teksten og tildeler emneord i overensstemmelse hermed. Også i 4. tilgang er begrebet tillagt en vis betydning.

Mange termer opnår kun 1,5 % konsistens, svarende til at én indeksør har brugt termen i den pågældende tilgang. *"Previous studies have noted the relationship between indexing depth and consistency (Svenonius, 1986; Tinker, 1966). That is, the more terms per document, the lower the consistency."* (Sievert & Andrews, 1991, s. 5). Denne antagelse bliver ikke umiddelbart bekræftet af den indledende konsistensmåling, i gennemsnit har indekserne brugt mellem 3 og 5 termer. Den lave konsistens kan skyldes, at enkelte indekser tildeler usædvanlig mange emneord.

Antallet af emneord pr. tilgang varierer. Der er brugt henholdsvis 40, 49, 43 og 46 unikke termer, efter

"normalisering", men langt flere i alt (se tabel 2). Tilsyneladende har indekserne haft brug for flere ord i nogle tilgange end andre, uden at det har betydet store variationer i ordvalg.

Gruppering

Optælling af termer og analyse af forskelle viste nogle tendenser i emneordsvalget. I sin diskussion af indekseringskonsistens siger Lancaster (2003, s. 69), at det kan være mere interessant at se, om indekserne finder frem til dokumentets hovedemner end de mindre aspekter. Med hovedformålet med undersøgelsen in mente, samt for at kompensere for den sproglige variation, blev termene grupperet i aspekter eller facetter.

Svenonius skriver, at problemet med facetter er, at de er vanskelige at definere og derfor vanskelige at klassificere termer ind i, og fortsætter: "The difficulties are caused by the fact that language is unruly. There are terms that cannot be categorized exclusively into only one facet and terms that resist categorization because they are ambiguous or abstract." (Svenonius, 2001, s. 141). Den erfaring blev bekræftet her, men der viste sig alligevel et billede af fire aspekter, som har trukket et stort antal termer. I første omgang blev resten samlet i en gruppe, fantasiløst kaldt "Andre".

Aspekter	1. tilgang antal emneord	2. tilgang antal emneord	3. tilgang antal emneord	4. tilgang antal emneord	emneord i alt
Historie	162	100	72	96	430
Indeksring	78	80	87	58	303
Abstract	64	44	26	25	159
Alfabet	5	24	42	7	78
Andre	11	11	57	24	103
I alt	320	259	284	210	1073
I gennemsnit	4,9	4,0	4,4	3,2	

Tabel 2. Antal emneord, opdelt i aspekter.

Aspektinddeling er udtryk for en fortolkning, og kan ikke gøres "korrekt". Her er grundlaget de tildelte emneord, dvs. at teksten og brugssammenhængen ikke er taget i betragtning. Også her måtte træffes beslutninger, fx om "indexing history" hører til i facetten "Historie" eller "Indeksring". Beslutningen blev at lade begrebet indgå begge facetter, fordi pointen er, hvor mange, som har fundet det vigtigt. I andre tilfælde er det uklart, hvilken betydning, indekseren

har tillagt termen: om fx "knowledge organization" skal indeholde "indexing". Delingen af "alfabetic" og "order" fra især 3. tilgang var også problematisk, men ordene blev stående som de var, som udtryk for indekserens valg.

Aspektinddelingen viser, at "Abstract" stadig vejer tungest i 1. tilgang, hvor "Alfabet" knap nok findes; det er til gengæld rigt forekommende i 3. tilgang,

pænt repræsenteret i den 2., og lavt igen i 4. Aspektet ”Indeksring” er skønnet vigtigt i alle tilgange, dog med forskelle.

Det historiske aspekt er interessant ved, at det har trukket flest emneord til sig (tabel 2), men når det kommer til stykket, har langt fra alle det med i deres valg af aspekter (tabel 3). Analyse af enkelte indekseringer viste, at mange indeksører, som har udvalgt det historiske aspekt, bruger 3 emneord på det, især i den 1. tilgang: ”history”, ”antiquity” og ”middle ages” optræder ofte sammen.

Under aspektet ”Andre” viser grupperingen i tabel 2, at der er brugt 5 gange så mange emneord i 3. tilgang

som i 1. og 2., og over dobbelt så mange som i tilgang 4. Tabel 3 viser et tilsvarende markant mønster.

Ved den mere eller mindre automatiske 3. tilgang, trækkes adskilligt flere specifikke termer ud end ellers. Termer som ”papyrus”, ”works” og egennavne. Den 4. tilgang indeholder også et stort antal unikke termer, men grunden må her være en anden. Hensigten med 4. tilgang er at finde emnet i helheden, og mange indeksører forsøgt at finde et enkelt eller nogle få brede begreber, som kan rumme alle emnerne. Begreber som ”knowledge representation” og ”library and information science” optræder ofte i 4. tilgang. Noget tyder altså på, at ikke alene antal termer og konsistens varierer fra tilgang til tilgang, *kvaliteten* af emnebeskrivelserne gør det også.

Aspekter	1. tilgang %	2. tilgang %	3. tilgang %	4. tilgang %
Historie	95,4	76,9	76,9	80
Indeksring	100	96,9	93,8	63,1
Abstract	93,8	61,5	29,2	33,8
Alfabet	7,7	32,3	72,3	10,8
Andre	2,3	10,8	38,5	24,6

Tabel 3. Andel af indeksører, som har aspektet med som emneord.

Emne og emneord

En beregning foregriber den kvalitative analyse: en optælling af emne og emneord i form af en sammenligning mellem de to trin i indekseringsprocessen, ”conceptual analysis” og ”translation”. Via denne optælling opnås en første indsigt i forholdet mellem emne og emneord.

Indenfor det historiske aspekt udviser optællingen god overensstemmelse mellem indeksørernes fund af emne og deres valg af emneord.

Indekseringsaspektet er jævnt repræsenteret i de to trin i de tre første tilgange. Overraskende er det dog, at flere indeksører har tildelt aspektet emneord, uden at have fundet det som emne. 4. tilgang adskiller sig

Aspekter	1. tilgang Emne %	1. tilgang emneord %	2. tilgang Emne %	2. tilgang emneord %	3. tilgang Emne %	3. tilgang emneord %	4. tilgang Emne %	4. tilgang emneord %
Historie	98,5	95,4	84,6	76,9	81,5	76,9	886,2	80,0
Indeksring	96,9	100	93,8	96,9	95,4	93,8	87,7	63,1
Abstract	80,0	93,8	67,7	61,5	47,7	29,2	53,8	33,8
Alfabet	6,2	7,7	44,6	32,3	86,2	72,3	9,2	10,8
Andre	7,7	12,3	26,2	10,8	663,1	38,5	26,2	24,6

Tabel 4. Aspektet som emne sammenlignet med aspektet som emneord, fordelt på andel af indeksører.

ved, at emnet ”indeksering” er fundet af 87,7 %, men kun 63,1 % har det med i emnebeskrivelsen. Som nævnt, er der tendens til at benytte brede begreber i denne tilgang; aspektet er altså ikke nødvendigvis blevet glemt eller til sidst skønnet uvigtigt, men kan simpelthen være inkluderet i fx ”knowledge organisation”.

Abstractaspektet viser kun i 2. tilgang nogenlunde konsistens mellem emne og emneord. Den manglende overensstemmelse mellem de to trin falder forskelligt ud i tilgangene. 1. tilgang er atter et eksempel på, at flere har aspektet som emneord end som emne, men i 3. og 4. er det modsat, og forskellen er større. Ovenfor har vi diskuteret, hvordan indeksererne – overladt til den subjektive vurdering i 2. tilgang - fandt abstractaspektet mindre væsentligt end forfatteren (tilsyneladende) har ment, det var. I 3. tilgang er forskellen 47,7 % som emne og 29,2 % som emneord, svarende til, at kun 19 af de 31 indekserer har udtrykt det i emneord. Den kvalitative analyse af indekserernes beskrivelser af indekseringsprocessen vil nok afsløre hvorfor, men foreløbig kan vi forsøge med et par antagelser. Arbejdet med de andre tilgange kan have påvirket indekserernes opfattelse, så de har haft svært ved at slippe aspektet, indtil de har nærstuderet optællingen efter 3. tilgang og konstateret, at begrebet optræder forholdsvis sjældent. I 4. tilgang er abstractaspektet muligvis inkluderet i et mere generelt begreb, ligesom indekseringsaspektet.

Alfabetaspektet viser høj konsistens i 1. og 4. tilgang, men kun få har det med, så udsvingene kan næsten ikke være store. I 2. tilgang er aspektet 8

gange blevet fravalgt som emneord – som om indeksererne alligevel ikke har stølet på deres egne observationer, eller har vurderet det mindre betydende. I 3. tilgang er aspektet fravalgt 9 gange, til trods for at såvel automatisk som manuel optælling viser, at det er hyppigt forekommende.

Aspektet ”Andre” har kun trukket få termer i 1. tilgang. Indeksererne kan have ment, at emnet var rigeligt dækket med forfatterens egne prægnante titelord. Forfatteren fremhæver ikke eksplicit andre begreber (hvis vi ser bort fra frekvens). Igennem de andre tilgange får indeksererne øje på flere emner, men har så fravalgt en del af mulighederne i emnebeskrivelsen. De har stort set fastholdt dem i den 4. tilgang, eventuelt fordi de allerede i emnebestemmelsen har søgt efter store linjer, som uproblematisk kunne videreføres som emneord.

Forklaringen på 2. og 3. tilgangs store forskel mellem de to indekseringstrin i forhold til ”Andre”, må søges i forskellige fænomener. Enklest at forklare er 3. tilgang, hvor optællingen fandt ord, som bare en hurtig intellektuel vurdering fik de fleste til at sortere fra, men til gengæld også fandt begreber, der næppe ellers, vurderet på de andre tilgange, ville være blevet fundet. I 2. tilgang er forskellen mere uklar, men kan have samme baggrund som diskuteret ovenfor: indekseren har været i tvivl om værdien af emnet, og har besluttet sig til at lade det fortabe sig i glemsel.

De relativt store forskelle mellem tilgangenes bidrag til aspektet ”Andre”, kombineret med et stort antal termer, førte til opdeling af aspektet. ”Andre” dæk-

<i>Andre aspekter</i>	<i>1. tilgang</i>	<i>2. tilgang</i>	<i>3. tilgang</i>	<i>4. tilgang</i>
<i>Aktiviteter</i>	1	1	76,9	4
<i>Dokumenttyper - materialer og indhold</i>	4	4	10	4
<i>Fag og discipliner</i>	3	1	3	4
<i>Viden og information</i>	2	4	1	
<i>Steder</i>		2	2	1
<i>Personer</i>	3	3	3	
<i>Naturlige processer og begivenheder</i>	1	1	1	2
<i>Produkter og serviceleverandører</i>	2	1	3	3

Tabel 5. Aspektet ”Andre”, opdelt på baggrund af kategorier i ASIS&T. Antal unikke termer.

ker over et bredt spektrum af begreber, som med Svenonius's ord er svære at klassificere ind i de andre aspekter. Der er tildelt i alt 103 termer med en del overlap. "Andre" er fordelt på mange delaspekter, som det kan ses i tabel 5.

De fleste delaspekter er repræsenteret i alle tilgange og rummer indenfor hver tilgang såvel specifikke som generelle termer. Flest af de specifikke findes i 3. tilgang, hvor fx mange dokumenttyper er nævnt, de generelle er i overvægt i 4. tilgang, med emneord som fx "libraries", "information science" og "knowledge organization".

Indeksering med én tilgang pr. indeksør

Først blev materialet analyseret med hensyn ukontrollerede emneord. Denne analyse kan derfor sammenlignes med den første undersøgelse. Indeksørerne brugte i gennemsnit flere ukontrollerede emneord pr. aspekt, op mod en fordobling. Det er fristende at antage, at det skyldes bedre tid til fordybelse, når kun én tilgang anvendes. Eller at indeksørerne er uøvede, og begår den typiske begynderfejl at overindeksere (Lancaster, 2003); tallene alene kan ikke forklare det.

Aspekter	1. tilgang		2. tilgang		3. tilgang		4. tilgang	
	4	1	4	1	4	1	4	1
	tilgange	tilgang	tilgange	tilgang	tilgange	tilgang	tilgange	tilgang
Historie	2,5	2,6	1,5	2,8	1,1	2,4	1,5	2,3
Indeksering	1,2	1,5	1,2	2,2	1,3	2,1	0,9	1,4
Abstract	1,0	0,9	0,7	1,1	0,4	0,5	0,4	0,9
Alfabet	0,1	0,4	0,4	0,8	0,6	1,0	0,1	0,6
Andre	0,2	1,6	0,2	1,8	0,9	1,9	0,4	1,8
I alt	5,0	7,0	4,0	8,7	4,3	7,9	3,3	7,0

Tabel 6. Antal ukontrollerede emneord i gennemsnit pr. indeksør. Sammenligning af 1. og 2. undersøgelse, henholdsvis 4 og 1 tilgang pr. indeksør

Den første undersøgelse viste forskel i vægtning af aspekterne i de enkelte tilgange. Med forbehold for det lavere antal indeksører pr. tilgang i anden omgang, fremtræder et noget andet billede (tabel 7). Historieaspektet og indekseringsaspektet nærmer sig

her komplet konsistens. "Abstract" ligger væsentligt højere i de 3 sidste tilgange og "Alfabet" er stærkere repræsenteret i 1. tilgang. Påfaldende mange har valgt at tildele termer indenfor "Andre"; det er især her det store antal ekstra emneord fremkommer.

Aspekter	1. tilgang		2. tilgang		3. tilgang		4. tilgang	
	4	1	4	1	4	1	4	1
	tilgange %	tilgang %	tilgange %	tilgang %	tilgange %	tilgang %	tilgange %	tilgang %
Historie	95,4	100	76,9	100	76,9	100	80	93,3
Indeksering	100	100	96,9	100	93,8	100	63,1	100
Abstract	93,8	85,5	61,5	81,3	29,2	50,0	33,8	68,8
Alfabet	7,7	37,5	32,3	43,8	72,3	93,8	10,8	62,5
Andre	12,3	62,5	10,8	56,3	38,5	68,8	24,6	56,3

Tabel 7. Andel af indeksører, som har aspektet med som ukontrollerede emneord i de to undersøgelser.

Emne og emneord

Også for det nye materiale blev der lavet en analyse af sammenhængen mellem beskrivelse af ”emne” og de resulterende emneord. Første undersøgelse viste overraskende forskelle mellem de to trin: nogle emner forsvandt i processen, nye aspekter dukkede frem som emneord. Det samme gjorde sig gældende i anden undersøgelse, endda med store udsving,

som fx ”Alfabet” i 2. tilgang (tabel 8). En antagelse på forhånd var, at når indekseren kunne koncentrere sig om én tilgang, ville der blive bedre overensstemmelse mellem indekseringsprocessens trin, men der er stadig forskel. En anden antagelse var, at der ville blive fundet flere emner, men de eneste steder, hvor tallene antyder, at det forholder sig sådan, er i 1. tilgang.

	1. tilgang		2. tilgang		3. tilgang		4. tilgang	
	Emne	emneord	Emne	emneord	Emne	emneord	Emne	emneord
Aspekter	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang	4 tilgange/ 1 tilgang
Historie	98,5 100	95,4 100	84,6 93,8	76,9 100	81,5 87,5	76,9 100	86,2 93,8	80,0 93,3
Indeksering	96,9 100	100 100	93,8 100	96,9 100	95,4 100	93,8 100	87,7 100	63,1 100
Abstract	80,0 93,8	93,8 85,7	67,7 75,0	61,5 81,3	47,7 68,8	29,2 30,0	53,8 87,5	33,8 66,7
Alfabet	6,2 35,7	7,7 35,7	44,6 6,25	32,3 43,8	86,2 68,8	72,3 90,0	9,2 33,3	10,8 60,0
Andre	7,7 43,9	12,3 64,3	26,2 25,0	10,8 56,3	63,1 43,8	38,5 70,0	26,2 20,0	24,6 53,3

Tabel 8. Andel af indekserer, som har aspektet med som emne og som ukontrollerede emneord.

Tallene viser i de fleste tilfælde højere konsistens på emneniveau end på termniveau. Det fund er i overensstemmelse med Preschels undersøgelse fra 1972, hvor hun forsøgte at adskille den konceptuelle analyse fra emnebeskrivelsen, og fandt, at indeksererne var mere enige om dokumentets emne end om det sproglige udtryk i ukontrolleret sprog (Leonard, 1977; Lancaster, 2003).

Kontrollerede emneord

I den anden empiriske undersøgelse indgik en tesaurus, for at se om et kontrolleret vokabular ville øge konsistensen. LISA tesaurus blev valgt, fordi den dækker fagområdet, og fordi den indekserede tekst burde kunne beskrives forholdsvis dybt med de deskriptorer, som er til rådighed.

Den første analyse i denne sammenhæng er en sammenligning af det antal kontrollerede og ukontrollerede emneord, indeksererne har tildelt.

Aspekter	1. tilgang		2. tilgang		3. tilgang		4. tilgang	
	Ukontrol. emneord	Kontrol. emneord	Ukontrol. emneord	Kontrol. emneord	Ukontrol. emneord	Kontrol. emneord	Ukontrol. emneord	Kontrol. emneord
Historie	2,6	1,6	2,8	1,4	2,4	0,9	2,3	1,4
Indeksering	1,5	2,3	2,2	2,6	2,1	2,6	1,4	2,4
Abstract	0,9	0,9	1,1	0,8	0,5	0,6	0,9	0,6
Alfabet	0,4	0,5	0,8	0,6	1,0	1,0	0,6	0,7
Andre	1,6	2,1	1,8	1,4	1,9	1,0	1,8	2,3
I alt pr. indeksør	7,0	7,4	8,7	6,8	7,9	6,1	7,0	7,4

Tabel 9. Antal ukontrollerede og kontrollerede emneord, i gennemsnit pr. indekser.

Antallet af emneord er faldet lidt i 2. og 3. tilgang, men ellers kan vokabularets begrænsende udvalg af termer ikke aflæses klart af tabellen.

Tesaurusen har generelt ført til let forhøjet konsistens, i forhold til de ukontrollerede emneord (tabel 10). Når effekten ikke er tydeligere, kan det skyldes indeksørernes manglende erfaring. *"One of the major advantages claimed for a controlled vocabulary is that it will improve consistency in the representation of subject matter. However, the relationship between vocabulary control and indexer consistency*

is not as straightforward as it might seem at first sight." (Lancaster, 2003, s. 73). Lancaster giver et eksempel på en gruppe utrænede indeksører og deres problemer med at anvende kontrolleret vokabular, især indenfor et fagområde, de ikke er fortrolige med. Gruppen i denne undersøgelse har et grundlæggende kendskab til faget, men ikke til brug af tesaurus. Det viste sig fx i en tilbøjelighed til at tilføje flere hierarkiske niveauer af samme begreb. En tesaurus er kompliceret at bruge, så en emneordsliste eller et andet enklere redskab, ville muligvis have givet mere ensartede indekseringer.

Aspekter	1. tilgang		2. tilgang		3. tilgang		4. tilgang	
	Ukontrol. emneord	Kontrol. emneord	Ukontrol. emneord	Kontrol. emneord	Ukontrol. emneord	Kontrol. emneord	Ukontrol. emneord	Kontrol. emneord
Historie	100	93,8	100	87,5	76,9	85,0	93,3	87,5
Indeksering	100	100	100	100	93,8	100	100	93,8
Abstract	85,5	87,5	81,3	81,3	29,2	43,8	68,8	68,8
Alfabet	37,5	31,5	43,8	43,8	72,3	81,3	62,5	0,5
Andre	62,5	68,8	56,3	68,8	38,5	51,0	56,3	68,8

Tabel 10. Andel af indeksører som har aspektet med som ukontrollerede og kontrollerede emneord.

Samlet tyder det på, at indeksørerne er i stand til at følge overordnede principper eller tilgange til indekseringsprocessen og at det kan føre til en vis konsistens. Først og fremmest på emne- og aspektniveau, i mindre grad på emneordsniveau. *"... it may be easier for two or more indexers to agree on a single term per document than on several. Or, the main topic of many documents may often be clear enough that choosing the term for that concept is easy, especially when such choices are severely limited by the indexing vocabulary."* (Sievrt & Andrews, 1991, side 5) I denne undersøgelse ser det ud til at passe: "indexing" og "history" holder en høj grad af konsistens i alle tilgange, mens mindre emner ofte kun medtages af en enkelt indeksør.

Undersøgelsen har vist områder, som bør undersøges mere i dybden. Analyse af de enkelte indeksørers beskrivelse af den samlede indekseringsproces kan vise, hvordan de har arbejdet med tilgange, herunder om de faktisk har fulgt Wilsons anvisninger, og om teoretiske beskrivelser af indekseringsprocessen, som fx Mais model, stemmer med indeksering i praksis. Kvaliteten af emnebeskrivelserne, fx med hensyn til hvilke typer af emneord de forskellige

tilgange er resulteret i, er et andet område, som materialet giver mulighed for at studere.

Konklusion

Det centrale studieobjekt i artiklen er, hvorvidt de holdninger og overvejelser – tilgange - som ligger til grund for emnebestemmelse og emnebeskrivelse, resulterer i konstaterbare forskelle. Formålet er at belyse indekseringsprocessen generelt.

Med udgangspunkt i de fire tilgange, som Patrick Wilson diskuterer, har to grupper indeksører indekseret samme tekst. Indeksørerne har et grundlæggende kendskab til faget, uden dog at være erfarne indeksører. Første gruppe benyttede hver især alle fire tilgange, den anden gruppe kun én tilgang hver. På den måde kunne Wilsons diskussion af, hvorvidt flere indeksører ville nå frem til samme emnebeskrivelse ud fra samme tilgang, eller om én indeksør ville få samme resultat med flere tilgange, afprøves.

Indekseringerne blev optalt og sammenlignet ved hjælp af konsistensmål. På trods af at nogle indeksører tilsyneladende har haft vanskeligheder med at følge Wilsons anvisninger fuldt ud, især i den 3. til-

gang, hvor frekvens er nøglen til emnet, viste det sig, at der kommer forskellige resultater ud af tilgangene. Forskellen gælder både de indekserede emner/aspekter af teksten og emnebeskrivelserne, og uanset om indeksørerne har arbejdet ud fra en eller fire tilgange.

Indeksørerne tildelte mange og meget forskellige emneord, og såvel specifikke som generelle termer fandtes i alle tilgangene. Undersøgelsen viste lav konsistens på termniveau, især med ukontrollerede emneord. Konsistensen blev en smule højere med en tesaurus, men indekseringerne bar også præg af, at en tesaurus er et vanskeligt værktøj at anvende uden træning. Når termerne blev grupperet efter aspekter af teksten, steg konsistensen mærkbart, ligesom der generelt var højere konsistens på *emneniveau* end på *emneordsniveau*.

Resultaterne tyder på, at såfremt indeksørerne har en overordnet tilgang til formålet med indekseringen, som de kan arbejde bevidst ud fra, vil det være muligt at opnå høj grad af ensartethed i indekseringsprocessens første trin.

Note

1. Francis J. Wittys (1973). "*The Beginnings of Indexing and Abstracting: Some notes towards a History of Indexing and Abstraction in Antiquity and the Middle Ages*"

Referencer

Beghtol, C (1986). Bibliographic classification theory and text linguistics: aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of Documentation*, 42, 84-113

Bliss, HE (1929). *The organization of knowledge and the system of science*. New York: Henry Holt and Company

Bloomfield, Masse (2001). Indexing: Neglected and Poorly Understood. *Cataloging & Classification Quarterly*, 33 (1), 63-75

Chu, CM. & A O'Brian (1993). Subject Analysis: The first Critical Stages in Indexing. *Journal of Information Science*, (6), 439-454

DS/INF100 (1994). Information og dokumentation: vejledning i indeksering: metoder til dokumentanalyse, emnebestemmelse og valg af emneord. Kbh.: Dansk Standardiseringsråd, DS-Information

Farrow, JF (1994). Indexing as a Cognitive Process. *Encyclopedia of Library and Information Science*, 53 (16), 155-171

Hjørland, Birger (1992). The Concept of "Subject" in Information Science. *Journal of Documentation*. 48 (2): 172-200

Lancaster, FW (2003). *Indexing and Abstracting in Theory and Practice*. London: Facet Publishing

Langridge, DW (1989). *Subject Analysis: Principles and Procedures*. London: Bowker-Saur

Leonard, LE (1977). *Inter-indexer consistency studies. 1954-1975: A review of the literature and summary of study results*. University of Illinois Graduate School of Library Science Occasional Papers, 131, 51pp.

LibraryThing. Lokaliseret 16.11.2008 på WWW: <http://www.librarything.com/>

Mai, Jens-Erik (2000). Deconstructing the Indexing Process. *Advances in Librarianship*. 23: 269-298

Mai, Jens-Erik (2001). Semiotics and indexing: an analysis of the subject indexing process. *Journal of Documentation*, 57, 591-622

Mai, Jens-Erik (2004). Analysis in Indexing: document and domain centered approaches. *Information Processing and Management*, 41, 599-611

Miksa, F (1983). *The Subject in the Dictionary Catalog from Cutter to the Present*. Chicago: American Library Association

Preschel, Barbara M (1972). *Indexer Consistency in Perception of Concepts and in Choice of Terminology: Final Report*. Washington, D.C.: U.S. Office of Education, p. 32

Ranganathan, SR (1950) *Colon classification*. Madras : Madras Library Association

Ricoeur, Paul (2002). *En hermeneutisk brobygger: tekster af Paul Ricoeur*. Red. af Mads Hermansen & Jacob Dahl Rendtorff. Århus, Klim

Rowley, Jennifer (1994). The Controlled Versus Natural Indexing Languages Debate Revisited: a Perspective on Information Retrieval Practice and Research. *Journal of Information Science*. 20 (2): 108-119

Sauperl, Alenka (2004). Catalogers' Common Ground and Shared Knowledge. *Journal of American Society for Information Science and Technology*, 55(1); 55-63

Sievert, MC & Andrews, MJ (1991). Indexing Consistency in Information Science Abstracts. *Journal of the American Society for Information Science*, 42, 1-6

Svenonius, E (1986). Unanswered Questions in the Design of Controlled Vocabularies. *Journal of the American Society for Information Science*, 37, 331-340.

Svenonius, Elaine (2001). *The Intellectual Foundation of Information Organization*. Cambridge, Mass.: MIT

Todd, Ross J (1992). *Academic indexing: what's it all about?* The Indexer, 18 (2), 101-104

Wilson, Patrick (1968). *Two Kinds of Power : an essay on bibliographical control*. Berkeley: University of California Press

Witty, Francis J (1973). The Beginnings of Indexing and Abstracting: Some Notes towards a History of Indexing and Abstraction in Antiquity and the Middle Ages. *The Indexer*, Vol. 8, No. 4