

Systemudvikling i et in-house miljø: Folketingets emneordssystem som case-studie

Af Peter Ingwersen

1. Introduktion

I årene fra 1989 gennemførte Folketinget en større omlægning af dets informationssystemer og -rutiner. I denne proces deltog jeg i perioden 1989-91 som ekstern informationskonsulent i forbindelse med opbygningen af Folketingets emneordstesaurus samt ved udarbejdelsen af databaseorganisation, indekseringsregler og brugergrænsefladedesign til Folketingets Udvalgsaktsystem (FTU). Dette arbejde udførtes som indtægtsdækket virksomhed under Danmarks Biblioteksskoles konsulentafdeling. Arbejdet foregik i et intensivt samarbejde mellem Folketingets udvalgssekretærer og systemkonsulent, bibliotekar Ingelise Romme, Datacentralen, som udførte - og stadig viderefører - den egentlige systemudvikling. Da den samlede systemløsning nu er færdigudviklet og taget i brug, har jeg fundet det relevant at beskrive de informationsanalytiske aspekter ved konsulentarbejdet.

Da det ofte er forbundet med store vanskeligheder, at finde litteratur på dansk om fremgangsmåden og især tidsforbruget i forbindelse med det grundlæggende

analysearbejde, har jeg fundet det naturligt at beskrive og analysere dette udviklingsarbejde nærmere. I det følgende vil jeg derfor koncentrere mig om de væsentligste aspekter forbundet med de analyser, som går forud for designet af en bruger-orienteret, flad thesaurus og brugergrænsefladen (systemanalysen), selve konstruktions- og designprincipperne samt brugsreglerne (detaljeret design), og tidsforbruget. Nærværende artikel er praksis-orienteret og retter sig mod bibliotekarer og andre informationsspecialister, som kommer i berøring med udvikling af emnesøgningssystemer i lokale, mellemstore og store miljøer. Samtidig kan adskillige af de beskrevne principper sagtens overføres til mindre systemløsninger.

Artiklen starter med at give et indblik i de problemer, som Folketingets primære brugere, dvs. MF-ere, udvalgssekretærer og medlemssekretærer har, i forbindelse med *genfindning og håndtering* af især udvalgsakter. Som man kan forvente, er hovedproblemet emneindekseringen og søgning af aktstykker og andre samhørende dokumenter.

Dernæst opstilles de *udviklingstrin*, som arbejdet kom til bestå af. I det store og hele forløber trinene som angivet i traditionelle indføringer i systemudvikling: den såkaldte *vandfalds-model* (Alnor, 1991). Dog blev modellen udbygget med større bruger-afhængighed end det normalt er tilfældet. Ligeledes optræder elementer fra den såkaldte *multi-view* udviklingsmodel (Wood-Harper et al., 1985). Jeg vil kort angive disse trin og det anvendte tidsforbrug, og iøvrigt betragte udviklingstrinene som den ramme, hvorunder resten af artiklen skal ses.

Selve *systemanalysen* gennemgås i det følgende afsnit. Den består hovedsagelig af en interviewundersøgelse af primær-brugernes søgeadfærd og -ønsker. Hovedelementerne i undersøgelsen opridses kort og de vigtigste resultater anskueliggøres. Det er her især spørgsmål vedrørende emneordsproblematikken og søgning som står i fokus.

Det følgende afsnit behandler udarbejdelsen af selve *tesaurussen*. Denne er af almen karakter, idet den dækker alle emneområder, som Folketinget kan tænkes at beskæftige sig med. Sammen med Folketingets arbejdsprocesser afspejler tesaurussen så at sige tingets *arbejdsdomæne*. I dette tilfælde forstås en flad tesaurus som en emneordskonstruktion, som i princippet medtager synonyme, nær-synonyme og, i visse tilfælde, tæt beslægtede termer til en foretrukket term. Der findes således ingen bredere eller snævre termer - ej heller delhelhedsrelationer eller relaterede termer. Vedligeholdelsesaspektet er den helt afgørende årsag til dette forhold. Tesaurussen er ment som grundlag for indekseringsprocessen, og som et indirekte hjælperedskab under søgning. Afsnittet beskriver kort tesaurussens dynamiske natur og principperne for dens brug og vedligeholdelse.

Det sidste afsnit analyserer kort de grundlæggende *grænsefladers* design-principper. Den ene grænseflade vedrører *registreringen* af udvalgsakter. Heri indgår visse principper for data-genbrug. Det andet grænseafsnit udgør den avancerede brugergrænseflade til brug ved *søgning*. En forenklet søgegrænseflade til registrerings-systemet, en midlertidig løsning, berøres kort i 3. afsnit. Artiklen afsluttes med en kort diskussion af resultatet og anvendelsesmulighederne.

2. Folketingets informationssystemer og -problemer

I slutningen af 80-erne besad Folketinget groft sagt to informationssystemer: *Folketingets Registreringssystem*

(FTR), som indeholdt en elektronisk registrering af lovforslag fremsat i folketingsalen af regering og oppositionspartier. Det andet system var et manuelt journaliseringssystem i gængs forstand, som registrerede de *udvalgsakter*, som behandledes i de dengang 23 stående folketingsudvalg. Der ønskedes en løsning - som illustreret fig. 1 - hvor hver MF-er kunne få adgang til ét samlet informationssystem, som ville gøre FTR, udvalgsakterne (FTU) og eksterne kilder som Retsinformation tilgængelige, samtidig med at systemet kunne anvendes af MF-erne og sekretærerne til tekstbehandling.

Som det er alment bekendt, fremsættes et lovforslag i salen, hvorefter det går i udvalg, dvs. det sendes til det udvalg som normalt tager sig af forslaget 'emne' eller 'sagsområde'. Man kan sige, at der her sker en reel klassifikation af forslaget, idet det normalt ikke tilsendes to (eller flere) udvalg. Derimod kan ét udvalg godt indkalde flere fagministre i forbindelse med behandlingen af et bestemt forslag. Efter endt udvalgsbehandling, sendes forslaget i modificeret form tilbage i salen til 2. behandling. Det kan nu returnere til udvalget, som genbehandler det eventuelt igen ændrede forslag, f.eks. via fremlæggelsen af nye ændringsforslag. Hvis der er politisk flertal for forslaget i udvalget, skrives der betænkning, der skabes bemærkninger til forslaget og det fremlægges afslutningsvis i salen. Udvalgsarbejdet danner derfor ryggraden i lovgivningsprocessen. Et vedtaget forslag vil altså have knyttet forskellige aktstykker til sig. Disse akter kaldes normalt for 'Forslagsdel'. Tilbage i udvalget eksisterer utallige andre akter, som kan hidrøre fra partier, udvalgsmedlemmer og -formand, berørte ministre (ministerier), samt kan være tilsendt udefra i forbindelse med høringer af interesseorganisationer og enkeltpersoner. Disse udgør FTU's 'Alm. del'. Mængden af udvalgsakter i 1990 var ca. 10.000 med en tekstmængde på 50.000 A-4 sider og er stadig stigende, dels i takt med antallet af fremsatte forslag, dels i takt med det stigende antal afholdte minister-samråd og især mængden af udvalgsspørgsmål, som stilles i almindelighed af udvalgsmedlemmerne. Spørgsmål og ministersvar udgør ca. 8000 årligt. Disse spørgsmål, stillet i udvalgene, må ikke forveksles med de såkaldte 'onsdagsspørgsmål', som stilles til ministre i salen. Af disse sidste er der nogle tusinde pr. år.

Som denne korte skitse antyder, foregår folketingsarbejdet som en løbende og ofte kompliceret proces med mange justeringer og indbyrdes afhængigheder mellem

diverse arbejdsresultater, dvs. af og mellem aktstykkerne. Ethvert udvalgsspørgsmål kræver naturligvis et svar - som kan resultere i flere nye spørgsmål, som alle er beslægtede på forskellig måde. Der findes derfor flere dimensioner i arbejdet med at holde styr på aktstykkerne:

en tidsdimension:

- fremsatte forslag, som fører til udvalgsbehandling, bl.a. høringer, og dermed forandringer i tekster og i emneaspekter for både forslag og underliggende aktstykker.
- Emnekernen er imidlertid bibeholdt og vil kunne genbruges.

en direkte afhængighedsdimension:

- mellem samrådsspørgsmål og svar ...
 - mellem spørgsmål, svar og nye spørgsmål ...
- Emnekernen opretholdes ikke altid i disse kæder af spørgsmål og svar.

en typologisk dimension:

- alle dokumenttyper har både overlappende og forskellige egenskaber, som naturligt peger på skabelsen af forskellige skræddersyede inddata- og søgeformater med mulighed for genbrug af overlappende data.

2.1 Emneordsproblemet

FTR-systemet lå på daværende tidspunkt tilgængeligt i BRS-Search. Der forgik en ad-hoc-lignende tildeling af emneord. Alle datatyper, emnedata såvel som beskrivende data, kunne søges i ét samlet register via en enkel grænseflade ved direkte brug af de Booleske operatører. Emnedata blev anvendt som baggrund for at skabe det trykte register til Folketingsårbogen. På grund af den skabte flertydighed, f.eks. at forslag fremsat vedr. "Storebæltsbroen" kunne findes under 'Storebæltsbroen', 'Storebæltsforbindelsen', 'Forbindelsen over Storebælt', 'Vestbroen', etc., var et samlet overblik alene i den enkelte årbog vanskeligt at opnå og online genfindingen i FTR besværliggjort. Det var derfor indlysende, at Folketingets Præsidium under formanden Svend Jakobsen tog initiativ til den *samlede løsning*, fig. 1, som i høj grad skulle inddrage et tidssvarende registrerings-system for udvalgsakterne. FTR og udvalgsaktsystemet burde derfor benytte samme typer emneord, dels p.g.a. de nævnte afhængigheds-dimensioner, dels for at lette primær-brugernes søgeadfærd. Det ansås ligeledes for naturligt at gøre udvalgsakterne søgbare for alle primær-brugerne, uanset under hvilket udvalg de rent klassifikatorisk var blevet anbragt. Specielt kan man her tænke på udvalgsmedlemmer i f.eks. Miljø- og Erhvervsud-

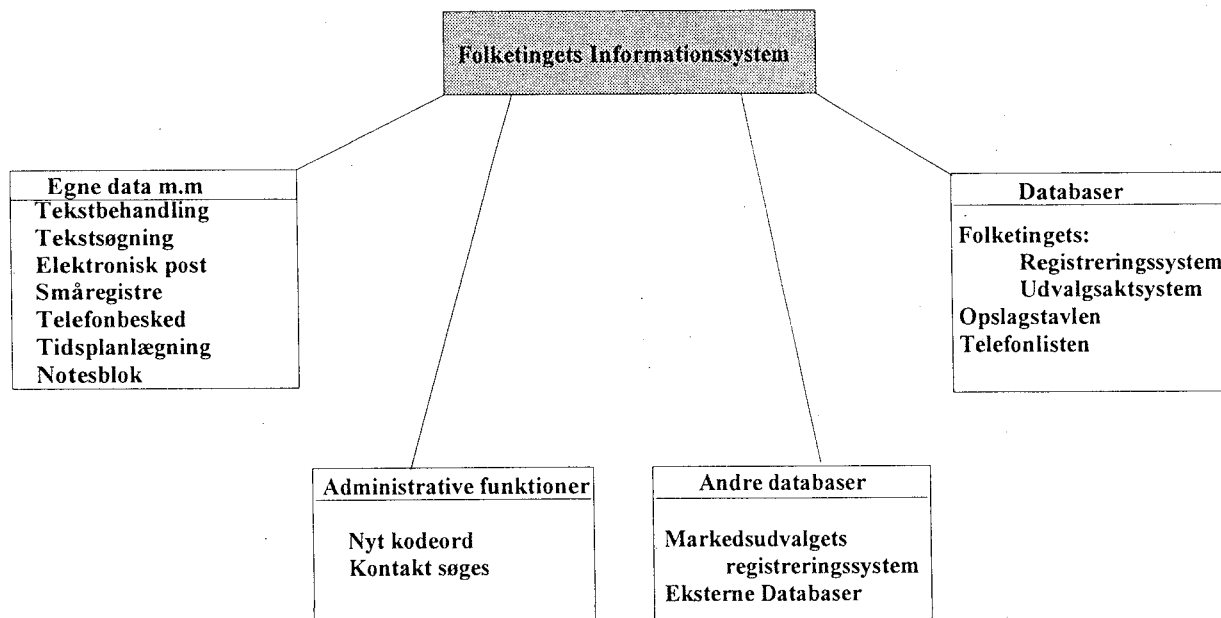


Fig. 1. Model af Folketingets Informationssystem, 1994, som det ses fra en bruger.

valgene samt Skatte- og Afgiftsudvalget, som ønsker adgang til hinandens ofte reelt emnemæssigt overlappende forslag eller aktstykker. Tænker man f.eks. på 'CO2-afgiften', befinder dette emne sig i rummet mellem de tre udvalg. Samme fænomen kan oftere og oftere iagttages imellem adskillige andre udvalg, f.eks. mellem Energi-, Markeds-, Trafik, Landbrug og Fiskeri- samt Miljøudvalget. Vort industrielle stade og vor kulturelt betingede demokratiform fremtvinger en forøget kompleksitet i lovgivningsarbejdet, som derfor kræver en *større grad af samtidighed og tilgængelighed* af beslutningsgrundlaget. Den enkelte beslutningstager i ét udvalg får ofte brug for adgang til flere andre udvalgs aktstykker. Hertil kræves en relativ specifik emneregistrering, som samtidig er kontrolleret som følge af mængden af potentielt set relevante dokumenter. Det kræves også, at informationssystemet opdateres løbende. Andre parlamenter, f.eks. det hollandske og engelske, sørger ikke for denne samtidighed og fleksibilitet i ajourføringen. Folketingets Bibliotek under informationschef Elise Holt blev naturligvis inddraget i dette arbejde. Her fandtes en fyldig men noget utidssvarende emneordliste, som oprindeligt var beregnet på bibliotekets interne behandling af dokumenter af en noget anden natur end korte aktstykker. I baggrunden spøjte endvidere en fremtidig overgang til registrering af og søgning i fuld-tekstversioner af akterne. Dette aspekt er blevet til virkelighed i 1993 - se 'Tekstfiler', fig. 2 nedenfor.

2.2. Emnesøgningsproblemet og arbejdsprocessen i domænet

Et problem, som man ofte støder på hos de fremtidige brugere ved overgang til elektronisk informationsstyring og -søgning, er deres manglende erfaring med informationssøgning som sådan. Problemet består ikke i teknologiangst, men knytter sig til den *anderledes søgemåde* i databaser - set i forhold til søgning i papirbaserede registre. Et beslægtet problem for beslutningstagerne, dvs. Folketingets Præsidium, bestod i at kontrollere de med systemudviklingen forbundne omkostninger, set i forhold til det forventede saglige udbytte, f.eks. i udvalgsarbejdet. Undertegnede blev derfor inviteret til at forklare 'emnesøgningens kunst' og de mulige løsninger for Præsidiets medlemmer. Med eksempler fra de daværende årbøgers registre var det en forholdsvis nem opgave at vise de fremkomne utilstrækkeligheder ved søgninger i årbogsregistre og lignende. Det indlysende svar er naturligvis at skabe en

kontrolleret emneordliste med tilhørende synonymer dækkende Folketingets arbejdsdomæne. Fremsatte forslag om opførelsen af Storebæltsbroen's Vestbro, samt de afledte aktstykker, burde f.eks. stå under 'Storebæltsforbindelsen - Vestbroen' i registret til årbogen. Problemet for beslutningstagerne opstår, når edb-løsningen kommer ind i billedet, idet alle kan se for sig de utallige aktstykker, som, forbundet med store ressource-tildelinger og omkostninger, alle vil have tildelt samme emneord: 'Storebæltsforbindelsen'. En stor og uproduktiv genfindingsprocent vil blive resultatet, hvis der kun søges på denne ene term. Man skal her så vidt muligt kunne anskueliggøre hvad der sker, når man som fremtidig bruger er nødt til, samtidigt, at *søge med flere søgeord* - en søgemåde, som de fleste mennesker i dagligdagen faktisk sjældent stifter bekendtskab med!

Beslutningstagerne skal med andre ord overbevises om mindst to grundlæggende faktorer, som begge er omkostningstunge: 1) Ved registreringen af dokumenterne, dvs. her både forslag og udvalgsakter, skal der tildeles *adskillige emneord* pr. dokument-enhed fra en kontrolleret liste, som naturligvis skal kunne opdateres løbende; andre væsentlige datatyper, inkl. ukontrollerede emneord til brug i søgningen, skal ligeledes tildeles dokumentet. 2) Det skal *fremgå eksplicit* ved udformningen af søgefladen og databasestrukturen, hvordan man kan kombinere disse søgedata uden for store søgetekniske problemer (anvendelsen af den Booleske logik). En væsentlig grund til at arbejdet overhovedet blev accepteret lå øjensynlig i det faktum, at *beslutningstagerne selv deltog i den politiske arbejdsproces* i Tinget, og derfor klart erkendte muligheden for *genbrug af data* i forbindelse med denne proces. Den politiske proces, og det dermed forbundne registreringsarbejde, består netop af de ovennævnte velkendte tids- og afhængighedsdimensioner. De omtalte kæder af eksempelvis spørgsmål-svar gør det indlysende, eller sandsynliggør, at omkostningsbesparelser kan finde sted, samtidig med at de enkelte dokumenter, f.eks. i form af spørgsmål, tilgængeliggøres. En yderligere årsag til beslutningstagernes ja til projektet skal måske ses i det faktum, at man klart ønskede at beholde de såkaldte 'bilag', samtidig med at hvert udvalgsdokument skal ses som en selvstændig enhed. Et forslag om at overgå til sagsstyring, hvor samhørende akter ville udgøre en "sag", blev forkastet af traditions- og proceduremæssige årsager. Et 'bilag' i folketinget er et A-4 ark indeholdende op til flere uafhængige spørgsmål til en bestemt minister. Både spørgsmål og svar, samt de

pågældende bilag har en selvstændig nummerering. Det er bilagene, der arkiveres fysisk. Ved en overgang til edb kunne man derfor forestille sig at opløse 'bilagsformen', da denne papirafhængige enhed ikke er nødvendig i elektroniske systemer. Ved trods alt at beholde bilagsformen understregedes imidlertid kompleksiteten yderligere - og de dermed forbundne ikke uvæsentlige omkostninger for beslutningstagerne. Konkluderende kan man derfor fremhæve, at det var en fordel for alle parter, at beslutningstagerne selv i detaljen var vidende om hvad de daglige arbejdsprocesser indebar af såvel vanskeligheder som muligheder. En sådan væsentlig og kompetent indsigt er imidlertid sjælden - ja unik - på lederplan i moderne institutioner og erhvervsliv.

3. FASER og UDVIKLINGSTRIN i det SAMLEDE PROJEKT for FTU

I den følgende beskrivelse af udviklingsarbejdet berøres alene systemet for Folketingets Udvalgsakter (FTU). Der skelnes imellem 'implementerings-faser' og 'systemudviklings-trin'. Faserne relaterer sig primært til Datacentralens arbejde med at få systemløsningerne op at stå. Trinene refererer til de enkelte analyse- og designarbejder, som ligger forud for implementeringen.

I forbindelse med faseopdelingen skal man være opmærksom på, at det endelige system kom til at bestå af to selvstændige hovedkomponenter - fig. 2: en inddaterings-komponent eller journaliseringssystemet, med såvel sine egne inddateringsformater, som sin egen simple søgegrænseflade til brug for sagsmedarbejderne - indeksererne. Dette system blev implementeret i en relationsdatabase-struktur med en 'Journalnitflade', fig. 2, tv. De enkelte udvalgsakter registreres i denne base, mens selve deres fulde tekst indskrives og lagres i 'tekstfil'-basen WPS+. Den anden hovedkomponent udgøres af et søgesystem byggende på inverterede datafiler (BRS-Search), her med en avanceret søgegrænseflade for primær-brugeren, fig. 2, til højre. Fuldtæksten af de i Folketinget skabte akter er overført til 'tekstfil'-basen. Med andre ord: inddateringen foregår til ét system, hvorefter filerne efter validering overføres til det andet system - dvs. til søgesystemet. Adskillelsen mellem online registrering og avanceret online søgning af udvalgsakter ansås hurtigt for fordelagtig af hensyn til system-performance - specielt når flere journaliseringer og søgninger forløber samtidig. Under stort arbejdspress i tinget bør man naturligvis undgå flaskehalse og tomgang i registrerings-processen. Erfaringsmæssigt vil sådanne problemer nemt opstå, hvis én og samme

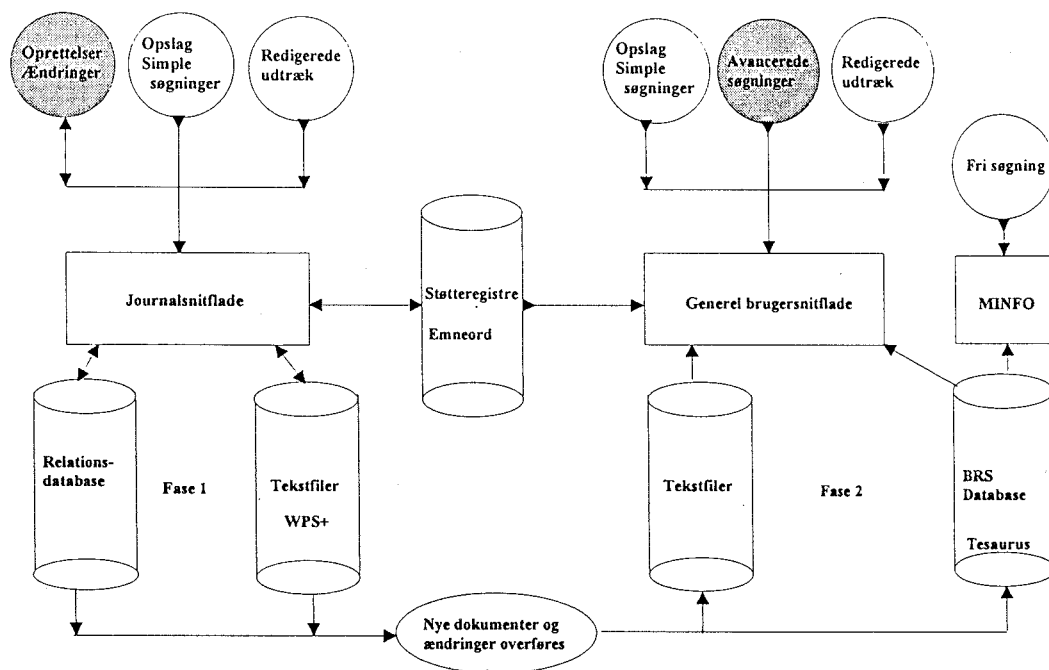


Fig. 2. Systemløsningen for Folketingets Udvalgsaktsystem (FTU) med Fase 1 i form af Journaliserings-systemet til venstre, og med Fase 2, BRS-søgesystemet, til højre.

computer på samme tid både skal honorere inddatering og søgning på fler-brugerniveau. Resultatet blev derfor et to-faset udviklings-arbejde: i *Fase 1*, 1989-1991, udvikledes journaliseringssystemet (All-in-1) med tilhørende inddata- og (simpel) søgegrænseflade samt

indlagt emneordliste som direkte støttefunktion; i *Fase 2*, -> 1993, skabtes søgesystemet i BRS-Search med den avancerede søgeflade og tesaurussen indbygget som indirekte søgesupport. En fuldstændig fri og kommandostyret søgemåde i BRS er mulig via MINFO-systemdelen.

3.1 Systemudviklingens indhold og tidsforbrug

SYSTEMUDVIKLING

Trin

1. System-analyse

Analyse af brugerbehov og -adfærd, arbejdsgange

- *Spørgeskemadesign*

- *Interviews *)*

- *Analyse og rapport*

2. Overordnet system-design

Kravspecifikation - overordnet

Foretaget af Styringsgruppen og Datacentralen

- *DB's bidrag til krav- og grænsefladeforslag*

3. Systemevaluering

Komparativ vurdering af edb-løsninger

foretaget af Datacentralen og Tinget. Resultater i

Fase 1 og Fase 2 løsningen.

4. Detaljeret system-design

- *Avanceret søgegrænseflade til Fase 2, ved DB*

- *Simpel søgegrænseflade til Fase 1, ved DB*

- *Tesaurus-konstruktion, sagsområder, vedligeholdelse, ved DB*

Registreringsformater ved Datacentralen

- *DB's bidrag hertil*

5. Implementering

Af journaliseringsdel (Fase 1), Datacentralen

Af søgedel i BRS-Search inkl. tesaurus (Fase 2)

- *Skabelse af regelsæt til indeksering og brug af tesaurus, ved DB*

6. Evaluering og tests - revision

- *Modifikation af tesaurus, sagsområder, regler*

- *Foreløbig ressourcetest af tidsforbrug ved inddatering, ved DB*

Evaluering og tuning af systemløsninger, DC

TIDSFORBRUG

Arbejdstimer (hele)

DB andel

90 timer (total)

35 "

15 "

40 "

30 timer

-

-

30 "

0 timer

440 timer

100 "

70 "

230 "

-

40 "

170 timer

-

-

170 "

130 timer

90 "

40 "

-

Total: 860 timer

- : Timeforbrug uoplyst

* : Interviews udført af udvalgssekretærer

Fig. 3. Trin i udviklingen af Folketingets Udvalgsaktsystem med angivelse af effektivt timeforbrug for de arbejder, som Danmarks Biblioteksskole udførte.

I nærværende fremstilling koncentrerer vi os om de udviklingstrin i form af analyser og designløsninger, som danner baggrund for Fase 1, samt om Fase 2's brugersnitflade til avanceret søgning, inkl. tesaurusanvendelsen. Danmarks Biblioteksskoles (DB) bidrag hertil består af ovenstående arbejder i *kursiv*, med angivelse af den reelt forbrugte tid, Fig. 3.

Projektet leledes af en politisk styringsgruppe nedsat blandt Folketingets partier af Præsidiet. Under denne gruppe varetog en egentlig faglig styringsgruppe bestående af udvalgssekretærer, sekretariatsledere og informationschefen det løbende arbejde med udviklingen igennem faser og trin.

Overfladisk set viser fig. 3 en systemudvikling i et trindelt forløb, som er typisk for den omtalte 'vandfaldsmodel': hvert trin bør være en forudsætning for det næste. Dog kan der optræde tilbagekoblinger, f.eks. som et resultat af systemevaluering, trin 3, som kommer til at påvirke og forandre specifikationer bagud i trin 2. Faktisk var det dét der skete ved faseopdelingen, hvorved det mere simple søgeflade-design måtte skabes i trin 4, men på baggrund af modificerede krav i trin 2. I sammenhæng hermed kan vi iagttage, at brugergrænseflade-udformningen havde en stor vægt, idet den så at sige blev bestemmende for de underliggende base- og formatstrukturer. I sig selv bestemtes den i høj grad af resultater fra bruger- og arbejdsprocesanalysen, trin 1. Med andre ord, betingelserne i arbejdsdomænet og primær-brugernes informationsadfærd ansås for afgørende ved udformningen af systemløsningerne. Derved ligger den anvendte udviklingsmodel tættere op ad den såkaldte 'multi-view metode', som er karakteristisk ved at betragte arbejdsopgaver og adfærd som de grundlæggende design-faktorer og styrende for grænsefladeudformning og funktionalitet. Afhængig af løsningen for snitfladerne, fører denne model først i anden omgang frem til udformningen af basestrukturer, feltbeskrivelser og datamanipulation.

Ser vi på *tidsforbruget*, er dette alene oplyst for DB's informations-arbejde. Tidsforbruget er målt i hele, effektive arbejdstimer, optalt og registreret under udførelsen. Det inkluderer mødevirksomhed og diskussioner med den faglige styringsgruppe og DC's systemkonsulent. Man kan altså ikke blot dividere timetallet med 37 for at få antallet af arbejdsuger, idet en normal arbejdsuge normalt vil indeholde perioder, hvor der ikke arbejdes effektivt. Her tænkes på spise-

pauser og andre afbrydelser. Realistisk set kan disse sættes til ca. 5 timer pr. uge. Skal man derfor prøve at anslå det reelle antal effektive arbejdsuger og evt. måneder, bør tallene snarere divideres med maksimalt 32.

Man bemærker, at *brugeranalysen*, se nedenfor i afsnit 4, ialt tog ca. 90 arbejdstimer at udføre, på trods af at selve interviewene udførtes af Folketingets egne udvalgssekretærer - og ikke af undertegnede. Ligeledes kan man observere, at *tesaurusproduktionen*, inkl. indekseringsregelsæt og vedligeholdelses-manual, kostede ialt små 500 timer. Heraf udgør de 230 timer selve konstruktionsarbejdet (trin 4), mens regelsættene tog 170 timer (trin 5) og revisionen ca. 90 timer (trin 6). *Brugergrænseflade-design* udgjorde ca. 170 timer ialt for begge søgefladetyper (trin 4).

4. Bruger- og domæneanalysen

Bruger- og domæneanalysen foregik som en interviewundersøgelse på baggrund af et struktureret spørgeskema bestående af 8 hoved- og 24 afhængige spørgsmål. Ialt interviewedes 32 udvalgte 'respondenter', fordelt på 18 folketingsmedlemmer (MF) sammen med deres sekretærer, 8 medarbejdere fra partiernes gruppesekretariater, samt 6 (grupper af) medarbejdere fra Folketingets administration. Det ses, at i flere tilfælde bestod den udspurgte 'enhed' af to eller flere personer som følge af den teamwork-adfærd, som er sædvane i Tingets daglige arbejde - f.eks. mellem MF-er og sekretær. Selve interviewene udførtes af udvalgssekretærer tilknyttet den faglige styregruppe. Derved sikredes en hurtigere gennemførelse af undersøgelsen og en større grad af tillid mellem udspørger og udspurgte, end hvis den for respondenterne ukendte konsulent foretog interviewene.

Spørgeskemaet afspejlede undersøgelsens dobbelte formål: at få kendskab til forskellige brugergruppers anvendelse af bilag og udvalgsakter, samt deres måde at søge på og brug af søgeord og -begreber. Der anvendtes overvejende lukkede spørgsmål, som regel med foruddefinerede svarmuligheder. Skemaet opdeltes i fire dele:

1) *Almene kendetegn:*

De adspurgtes *arbejdstilhørsforhold* (MF-er, Udv.-sekr., etc.); deres nuværende *systemkendskab*, f.eks. til FTR, Årbog eller udvalgsakter; deres *tidsramme* for retrospektiv søgning; *brugen af udvalgsbilag*, f.eks. fra andre end egne udvalg, til

anvendelse i møder, udarbejdelse af avisindlæg eller spørgsmål til ministre; søges der efter specifikke dokumenter eller/og emner? *Forældelsen* af udvalgsakter, f.eks. efter ministerskift, valg, ved nyt folketingsår, flere år - hvor mange?

2) *Informations-kvalitet - i sidste konkrete søgetilfælde:*

Det konkrete *emne*; den søgte *dokumenttype*, f.eks. lovforslag eller spørgsmål, samt hvilke *datatyper* den adspurgte *kendte*, f.eks. afsender, titel, nummer, udvalgsnavn, minister, ordlyd - eller blot emnet. *Hvem søgte information* i det aktuelle tilfælde? - den adspurgte, eller via sekretærer, biblioteket, etc., samt *kvaliteten* af den fundne information, f.eks. relevans, brugbarhed, aktualitet, tilstrækkelighed, fejlagtighed.

3) *Informations-kvalitet - i almindelighed:*

Samme spørgsmål som i Del 2), men besvaret som den adspurgtes opfattelse af de daværende systemer i al almindelighed. Her spørges også om den *tidsramme* hvorunder bilag opbevares hos den adspurgte i et eget arkiv.

4) *Emneordstyper og -niveauer:*

De adspurgtes opfattelse af *synonym-former*, f.eks. om hvorvidt 'Storebæltsbroen' opfattes som lig med 'Storebæltsforbindelsen', lig med 'Fast Storebæltsforbindelse'. Forventning til de *sproglige udtryk og niveauer*, f.eks. om man ønsker opslag på ord som: Brugerbetaling, Egen- eller Selvbetaling, Kartoffelkur, Skattefusk, Skatteål, Overskudsdeling, Federøvstillæg eller Lokalløn, EF (nu EU) eller De Europæiske Fællesskaber, DIS eller Dansk Internationalt Skibsregister.

I skemaets første del undersøges hvorvidt de nuværende EDB-søgesystemer anvendes, altså primært FTR, samt Biblioteket og Årbogen. Her ønskes viden om de bagvedliggende *informationsbehov*, altså grundene til søgning og brugen af funden information. Oplysninger herom har betydning for fleksibiliteten i fremvisningen af søgeresultater samt muligheden for genbrug af funden tekst. Hvor lang tid tilbage der søges kan sammenholdes med *tidsrammen* for opbevaring i personlige arkiver - Del 3. Oplysning herom indebærer, at designeren kan fastsætte en tidshorisont, som er umiddelbar (default) søgbar, f.eks. det sidste folketingsår el.lign.

Skemaets anden del bygger på 'critical incident'-metoden (Martyn & Lancaster, 1981), hvor det sidst aktuelle søgeproblem beskrives. Heroverfor står skemaets tredje del, som anskuer hele adfærden i 'al almindelighed'. Man vil således kunne observere forskellige *adfærdsmønstre og opfattelser af kvaliteten*, alt efter hvad den adspurgte konkret har foretaget sig, parret med han/hendes generelle erfaring. Er der f.eks. overensstemmelse mellem den generelle opfattelse og de konkrete søgninger på tværs af respondenterne, har det pågældende mønster stor slagkraft. Ligeledes ønskes viden om brugernes adfærdsmønstre i mere detaljeret form under søgning, hvilke beskrivende eller emnedat typer de kender, og om de søger efter ét eller flere akter. Dette har betydning for hvor *specifik indekseringen* af de enkelte udvalgsakter må forventes at skulle være.

Den fjerde del af interviewet koncentreredes udelukkende om *emneordenes optræden* set fra brugernes synspunkt, dvs. brugernes forventning til tesaurusstrukturen. Hvis f.eks. beslægtede termer som Vestbro og Storebæltsbro opfattes som synonyme, bør den fremtidige tesaurus også indeholde 'beslægtede termer' som indgangsord til ret overordnede foretrukne termer. Er det ikke tilfældet, bør disse begreber adskilles i form af forskellige foretrukne termer. En anden dimension i et arbejdsdomæne som Folketingets er det sproglige *niveau* eller *form*. MF-erne er alle storforbrugere af pressens omtale af politiske begivenheder, og de skaber og formidler selv en mængde budskaber via TV og avisindlæg. De påvirker og påvirkes politisk og sprogligt i disse processer. Alt efter partitilhørsforhold og øjeblikkelig stemning forekommer jævnligt visse politiske slangudtryk i debatten - som dermed sandsynligvis bør indgå som ord i en tesaurus. De i undersøgelsen valgte udtryk udsprang dels fra visse dagblades stadige anvendelse i overskrifter, dels fra begrebsdannelser undfanget i selve Folketinget under debatter. Andre igen var 'neutrale' eller objektivt beskrivende udtryk. Til den første kategori tilhører udtryk som 'skatteål' og 'federøvstillæg', mens udtrykket 'kartoffelkuren' tilhører Tingets egne opfindelser (efter sigende undfanget af Statsminister Schlüter i forbindelse med fremlæggelsesdebatten i salen).

4.1 De væsentligste resultater

Brugen af udvalgsbilag. Eksisterende udvalgsakter anvendes hovedsagelig ved behandling af indholdsbeslægtede nye forslag, til udvalgsmøder og ved udarbej-

delse af spørgsmål til ministre. Bilag fra andre udvalg end de(t) MF-eren selv er medlem, af anvendes sjældent. Dette betyder, at designet skal sørge for at søgeren let kan *isolere* og fastholde sine 'egne' udvalgs bilag. Det skal være muligt at finde et forslag (i 'forslags-del') som søgeudgangspunkt, for derefter at kunne anvende emneordene til at søge efter andre udvalgsakter eller forslag, som er beslægtede, se fig. 7. Samtidig anvender medlemssekretærene meget ofte bilagsakterne i arbejdet for MF-eren. Over halvdelen af medlemmerne bruger akterne til noget 'andet' end de i skemaet nævnte muligheder, f.eks. i kontakten med vælgerne, med kolleger og med partiet.

Forventet forældelse og opbevaring. For udvalgsakter tilhørende både 'Forslagsdel' og 'Alm. del' gælder, at over halvdelen af respondenterne anser disse for forældede efter ca. to år, svarende til den gennemsnitlige valgerperiode i Danmark. En trediedel af MF-erne mener endda, at forældelsen indtræder efter folketingsårets afslutning. Tilsvarende opbevares i snit begge aktyper godt ét år i egne arkiver. Dette vil konkret betyde, at den umiddelbare (default) adgang via EDB til FTU højst bør udgøre de seneste to års aktstykker, snarere mindre, f.eks. et folketingsår alene. Der er forskellige årsager til opbevaring af bilag, f.eks. i henhold til emne (20%) eller i relation til egne ordførerskaber (20%). 25% af MF-erne gemmer selv akterne, mens 75% overlader dette til sekretæren eller gruppesekretariatet. Systematikken i opbevaringen er forskellig og afspejler de for brugeren væsentlige genfindings-facetter: Det hidtidige journalsystem, folketingsår og udvalg, samt udvalg kombineret med forslag.

Vi kan således konstatere, at *udvalget, kronologien, forslag og emne* har stor betydning for søgningen. Det er ligeledes væsentligt for brugerne at kunne finde *ét bestemt dokument* (90%), hvor de beskrivende data er kendt (54%) eller hvor alene emneområdet er kendt (36%). Dette gælder både ved konkrete søgninger og ved besvarelserne om søgning i al almindelighed. Som man iøvrigt kunne forvente, søgte MF-erne (og deres sekretærer) oftest i eget arkiv, mens sagsmedarbejderne tog udgangspunkt i FTR-systemet. Designet af søgesnitfladen bør derfor muliggøre en klar skelnen mellem søgning på flere emne- eller tekstord og søgning på beskrivende data, som afsender, bilagsnummer, etc. *Emnesøgningen* skal kunne gøres *specifik og snæver*, da de fleste helst ser ét eller meget få aktstykke som slutresultat. Ved fuld-tekst søgning vil dette forhold

udgøre et problem, dels hvis emneindekseringen er sparsom og bred, dels hvis denne ikke kan adskilles fra fuld-tekstsøgningen. I Fase-2 løsningen, afsnit 6 nedenfor, er denne mulighed indbygget.

Disse resultater lægger op til, at søgegrænsefladen skal være avanceret og form- og menubaseret, d.v.s. være opdelt i bokse til de vigtigste datatyper.

Et interessant aspekt var *graden af tilfredshed* eller brugbarhed af det fremfundne materiale - inden indførelsen af FTU på edb-form. Af besvarelserne vedr. de konkrete (sidste gangs) søgninger fremgår, at ca. 80% fandt resultatet relevant eller tilfredsstillende. I de 20% utilfredsstillende svar var det hovedsagelig FTR-søgninger, som gav anledning hertil, f.eks. ved brugen af den Boolske logik. I svarene for søgning generelt var det derimod ekstremt få søgninger 'i al almindelighed', som var utilfredsstillende. Vi kan her konstatere den sædvanlige metodiske brist ved at spørge på respondenternes brede og generelle 'gennemsnits-tilfredshed'. Svarene er ofte uanvendelige, i modsætning til når der spørges på konkrete sidstegangs-søgninger (critical incident-metoden).

Resultatet af *emneords-analysen* gav følgende vedrørende *synonymkontrol*. Ca 60% af alle adspurgte anså Virksomheder=Firmaer, mens 37% mente nej. MF-erne var her mere sikre på synonymitet end sagsmedarbejderne. Hypotesen var, at man ville anse de to udtryk for synonyme. Hypotesen er således ikke ganske underbygget. Uenigheden vedr. Storebæltsbro=Storebæltsforbindelse var endnu mere utalt: 55% ja - og 45% nej. Dette skyldes naturligvis den just vedtagne dobbeltløsning, med såvel bro som tunnel. Derimod var der en udtalt enighed om, at Fast Storebæltsforbindelse=Storebæltsforbindelse var synonyme udtryk (80%). Storebæltsbroen=Vestbroen gav som ventet 75% nej, mens 12% var usikre.

Det blev derfor klart, at der så vidt muligt kun skulle skabes synonymitet imellem reelle synonyme, mens del-begreber, som Vestbroen, måtte anses for selvstændige emneord. Samtidig vil visse uægte synonyme aften Storebæltsbroen=Storebæltsforbindelsen, kunne være gavnlige i søgemæssig sammenhæng. På sin vis hænger dette forhold sammen med den velkendte og naturlige, men fagligt set 'uægte' relation Hval <-> Fisk.

I relation til *emneordstyper og -niveau* var hypoteserne følgende: begreber opfundet i Tinget, som 'Brugerbetalning' og 'Kartoffelkur', forventes at kunne anvendes. Ordene anvendes ligeledes af befolkningen og pressen. 'Egenbetaling' og 'Selvbetalning' er begge opfundet som ideologiske modudtryk (af KF og SF henholdsvis). De anvendes sjældent af pressen, men kunne tænkes at finde anerkendelse som opslagsord. 'Overskudsdeling' er et eksempel på et anerkendt, 'neutralt' udtryk, som forventes at få stor opbakning. Med ord som 'Skattefusk', 'Skatteål' og 'Fедterøvstillæg' forsøges fundet en nedre, æstetisk præget grænse for opslagsordenes form. Begrebet 'Skattesnyd' fandtes iøvrigt i forvejen som anerkendt emneord i FTR.

Resultatet af undersøgelsen, Tabel 1, var relativt set i overensstemmelse med forventningerne.

Forventes opslag på:	JA	NEJ	Usikker
BRUGERBETALING	100	0	0
KARTOFFELKUR	100	0	0
OVERSKUDSDELING	100	0	0
STATENS KUNSTFOND	100	0	0
EGENBETALING	31	59	10 1)
SELVBETALING	3	90	7
SKATTEFUSK	16	84	0
SKATTEÅL	10	87	3 2)
FEDTERØVSTILLÆG	13	84	3 3)
EF	94	-	-
DIS / Fuld navneform	26/29	-	-
Begge navneformer	45	-	-

- 1): MF-ere: 68% NEJ
 2): MF-ere: 100% NEJ
 3): Alle grupper enige

Tabel 1. Forventninger til typer og niveau for fremtidige opslagsord, procent-fordeling.

Vi kan konstatere, at usikkerheden er lille blandt de adspurgte. Kun i forbindelse med ordet 'Egenbetaling' spores en vaklen (10%). Som noterne udtrykker, er MF-erne mere sikre på deres opfattelse end sagsmedarbejderne og sekretærerne. For det pågældende ord er tvivlen størst blandt de sidstnævnte grupper (23%). Det samme forhold gælder begrebet 'Skatteål'. Igen er det de sidstnævnte grupper som er mere vaklende end tabellens gennemsnitstal fortæller. Faktisk var 31% af

disse grupper enten i tvivl om anvendelsen, eller forventede at kunne anvende ordet. For alle andre ord er resultatet en klar udmelding, som også bekræftedes af kommentarer fra respondenterne:

Tingets egne populæruddtryk kan anvendes, især hvis de er skabt som neutrale eller distancerende begreber (kartoffelkur). Mere ideologisk prægede eller dagbladsjournalistiske slagord finder ikke rigtig genklang blandt MF-erne, mens de medarbejdere, som ofte optræder som formidlere mellem omverden og MF-eren, i højere grad vil forvente at kunne anvende sådanne ord. Der ønskes i almindelighed anvendt normaldanske ord, ikke juridisk jargon eller økonomisk, dvs. videnskabelig præget, sprogbrug.

Entydige forkortelser forventes anvendt. Ikke overraskende var de fleste bevidste om homograf-problemet med 'DIS', som også kan stå for rejsebureauet og studenterorganisationen. I sådanne tilfælde forventes også opslag på den fulde navneform.

Konkluderende kan man hævde, at folketingsmedlemmerne har en klar idé om og forventning til emneordenes niveau og form. Sagsmedarbejderne og medlemssekretærerne har, sikkert som følge af deres daglige berøringsflader, en noget mere nuanceret forventning om at kunne finde folkelige populæruddtryk anvendt som opslagsord.

5. Tesauruskonstruktionen

Dette afsnit beskriver kort principperne for valg af emneord og opbygningen af den flade tesaurus. Ligeledes fremhæves de væsentligste aspekter ved dens vedligeholdelse, og hovedprincipperne ved emneindekseringen af aktstykker understreges.

Tesaurusens størrelse blev fastsat til ikke over 3.000 foretrukne termer (KE-ord). Hertil kommer synonymmer og tæt beslægtede termer, som udgør tesaurusens registerord. Disse vil også kunne anvendes som ukontrollerede ord (UK-ord), hvis det viser sig hensigtsmæssigt. Med en sådan størrelse opnås to ting: 1) For udvalgssekretærerne, som dagligt kommer til at arbejde med emneordslisten, er antallet til at overkomme hukommelsesmæssigt. 2) Det vil være muligt at anvende *to-ledede begreber* som emneord. På dansk har vi mange sammensatte ord. Halverer man antallet af emneord til ca. 1500, som samtidig skal dække det

meste af samfundslivet, opnår man kun at kunne anvende ét-ledede ord. Begrebet 'olieboreplatforme' vil eksempelvis så skulle anbringes under tre ord: Platforme; Boring; Olie. Det er imidlertid ikke sikkert, at alle sådanne ordtyper vil kunne anbringes i en mindre liste, idet specifikke ord, som f.eks. 'boring', sandsynligvis ikke finder plads. Der bliver derfor tale om at beskrive emner med ret brede begreber. Visse begreber vil slet ikke eksistere. Valget på ca. 3.000 emneord gør det muligt at finde et mellemlige i specifikationsmæssig henseende. Der bliver plads til normale ét-ledede ord, som Olie eller Benzin, men også til sammensatte ord, som f.eks. 'Boreplatforme' eller 'Storebæltsforbindelsen'. En yderligere grund til dette valg ligger i, at BRS-Search både kan trunkere til højre og venstre. Man vil altså kunne regne med at få et søgeresultat, hvis man f.eks. søger på '*platforme' (=olie-, bore-, olieplatforme, etc.).

I realiteten endte *emneordstesaurussen* med at udgøre 2.600 foretrukne termer og ca. 4.500 yderligere op-

slagstermer i form af synonymer og andre beslægtede ord. Hertil kom danske institutionsnavne, ca. 2.500, med ca. 5.000 tilknyttede navneformer. Denne navneformskontrol blev udført internt i Folketinget på et senere tidspunkt.

Der findes 'Scope-notes' hvor det er nødvendigt, mest i form af se-også henvisninger til beslægtede under- eller sideordnede begreber, som en indeksør måske hellere vil bruge for at specificere et aktstykke. Se Fig. 4.

Som det fremgår af figuren, indeholder de beslægtede ord også andre stavemåder afpasset efter ny retskrivning (f.eks. Afbyureaukratisering). Navneform i ental er så vidt muligt anvendt. Hvor emneordet, set fra et registersynspunkt, burde være i flertal, f.eks. i tilfældene 'Afgift' eller 'Skat', findes flertalsformen som synonym. Opslag på disse former vil derfor automatisk lede frem til den foretrukne term, hvis synonymkontrollen er slået til under søgning.

ABORT	Abortindgreb Abortloven Fosterdrab Fosterfordrivelse Graviditetsafbrydelse Legal abort Svangerskabsafbrydelse	AFBUREAUKRATISERING Afbyureaukratisering
ADGANG		AFFALD Skarn Skrald
ADMINISTRATION	Forvaltning	AFFALDSDEPOTER (s.o. Fyldpladser, Lossepladser) Affaldsdepot Affaldslagre
ADOPTION	Adoptionsloven	AFFALDSFORBRÆNDING Affaldsafbrænding
ADVOKATER	Jurister Sagførere	AFFALDSPOLITIK
AFBETALING	Afdrag Amortisering	AFGIFT (s.o. Forbrugsafgifter, enkelte afgifter) Afgifter Gebyr
AFBRYDELSE	Afbrud Standsnings Stop	AFGIFTSFRIT SALG Afgiftsfri salg

Fig. 4. Udsnit af Folketingets Tesaurs (emneordsdel) med foretrukne termer (store bogstaver), scope-notes i () og beslægtede, synonyme begreber indrykket med små bogstaver.

5.1 Tesaurusopbygningen

Til at opbygge tesaurussen anvendtes fire hovedkilder:

- Sagsområderne i de enkelte udvalg
- Lovkatalogen og dens beskrivelse i pressen i forbindelse med åbningstaler
- Registre i flere årgange af Folketingsårbogen
- Status' registerord

Dertil kom anvendelsen af Synonymordbogen. Der var i emneordsdelen tale om tre emneordstyper: neutralt beskrivende ord, f.eks. 'Afgift'; lokalitetsord, f.eks. by-, regions- og landenavne; handlingsprægede ord, f.eks. 'Lempelse' eller 'Ophævelse'. Handlingsordene var bestemt til at udtrykke den proces ved et objekt, som et dokument indholdsmæssigt omhandlede, f.eks. at et hørings svar argumenterer for en 'lempelse af formueskat' eller en 'ophævelse af renteaftgiften'. Set fra et søgesynspunkt var dette en udmærket idé, dog kun under visse forudsætninger. *Handlingerne*, som aktstykket behandler, skal være entydige. Hvis et spørgsmål til en minister f.eks. behandler sammenhængen mellem 'ophævelse' af 'renteaftgift' og 'nedsættelse' af 'rentefradrag', ses det klart, at en søgning på 'ophævelse OG rentefradrag' tillige vil kunne genfinde det nævnte spørgsmål og evt. andre aktstykker, som handler om noget helt andet end det søgte. Desuden vil udvalgsbilag, som knytter sig til et forslag, citere forslaget titel, som normalt indeholder de relevante handlingsord, f.eks. Ophævelse af lov.... Ved indførelse af fuld-tekst søgning af internt skabte udvalgsakter, vil det ligeledes være muligt at søge på denne ordtype i selve teksten. Tilbage af relevante udvalgsakter findes alene de udefra kommende dokumenter. Handlingsordene finder derfor kun meget lidt anvendelse i det nuværende system.

Sagsområderne i udvalgene udgør så at sige *kernen* i emneordsproduktionen. Forud for skabelsen af FTU havde kun et fåtal af udvalgene en sagsområde-styring af akterne. Ved at sætte dette i system i hvert udvalg, skabtes derfor *udgangspunktet* for tesaurussen. Forslagsdelen i FTU har ikke brug for sagsområdeopdeling, da hvert forslag virker styrende for de tilknyttede bilag og aktstykker. For de andre akter, i 'Alm. Del', skabtes et antal hinanden udelukkende (eksklusive) sagsområder af de ansvarlige udvalgssekretærer. Sagsområderne består af strenge af fremtidige foretrukne emneord. Fig. 5 viser sagsområderne for Skatte- og afgiftsudvalget.

SKATTE- OG AFGIFTSUDVALGET; SAU Indeholder følgende SAGSOMRÅDER: Administration (personale, bevilling m.v.) Internationale sager Løbende oversigter Rejser; Besøg Diverse ----- Arveafgift; Gaveafgift Forbrugsafgifter; Renteafgift Klageinstanser Lejeværdi; Ejendomsvurdering Moms; Arbejdsmarkedsbidrag Pension; Beskatning; Realrenteafgift Personskat; Virksomhedsskat Selskabsskat; Skattemæssig afskrivning Selvangivelse; Forskudsregistrering Tipning; Lotto Told Værdipapirbeskatning

Figur 5: Sagsområderne i Skatte- og afgiftsudvalget. (1992).

For alle udvalg findes et antal fællesområder, f.eks. af administrativ art. Desuden kan samme sagsområde optræde i flere udvalg. F.eks. optræder 'Tipning; Lotto' både i Skatte- og afgiftsudvalget og i Kulturudvalget. Efter nogle høringsrunder, som resulterede i konsistens af termer og områder, samt i regler for opretning, nedlæggelse eller ændring af sagsområder, var der derfor skabt en kerne på ca 350 egentlige emneord. Herudfra er det alene et intensivt arbejde, at få skabt de resterende 2200 ord med tilhørende beslægtede udtryk. Her kommer Lovkatalogen og dens pressedækning, samt Status, ind som domæner relevante kilder. Ud fra sagsområderne og de andre hovedkilder kan konstruktøren associere sig frem til andre relevante begreber og termer, som endeligt valideres af udvalgssekretærerne.

Tesaurussen opbyggedes i Polydoc Mikro, som er et informationsøgeprogram med inverteret filstruktur, indeholdende en stærk rapportgenerator. Fordelen ved denne opbygningsmåde ligger i, at hver foretrukne term, og de enkelte beslægtede termer, udgør en selvstændig post i systemet. Den anvendte feltstruktur i basen gør det muligt at sammenkæde den foretrukne term med sine beslægtede udtryk, *samtidig med* at man kan hæfte relevante udvalg og andre oplysninger på den enkelte kontrollerede term. Man har således under opbygningen hele tiden en ajourført liste af allerede tildelte emneord med beslægtede ord for de enkelte ud-

valg. Derved undgås dobbeltindførsler og konsistensen bevares. Man kan senere udskrive hele tesaurusens vokabular, med eller uden beslægtede termer, og udskrive del-tesaurusser for de enkelte udvalg. Rapportgeneratoren kan indstilles til utallige udskrivningsformater, bl.a. også til BRS-Search' tesaurus-struktur.

Tillige får konstruktøren efterhånden et overblik over tyngden af emneord i de enkelte udvalg. Det er ganske interessant at observere hvilke udvalg, som fremstår som *tunge* udvalg i henseende til antallet af tildelte emneord. Jo flere kontrollerede emneord, jo større er

sandsynligheden for en præcisering af indholdet i de pågældende udvalgs akter. Udvalg med stor gennemslagskraft i den politiske proces vil tiltrække store mængder aktstykker. Det er derfor vigtigt, at de enkelte udvalgs emneordstildeling foregår i et gensidigt samarbejde mellem pågældende udvalgssekretær, i egenskab af domæneekspert, og konstruktøren. Finansudvalget er det altoverskyggende forum i Folketinget, og anvender naturligvis hele tesaurusens ordforråd, inkl. de kontrollerede navneformer. Men det ses tydeligt af Tabel 2, at Markedsudvalget også indtager en særstilling i henseende til områdedækning.

Udvalg Frekvensorden -----	Totale termer -----	Egne termer -----	Over- lap -----	Ratio* Unikhed -----	Sags- områder -----
Finansudvalget	2600	0	2600	0,0	-
Markedsudvalget	1180	130	1050	12,3	25
Erhvervsudvalget	930	65	865	7,5	15
Landbrug og Fiskeri	535	45	490	9,2	18
Skatte- og Afgiftsudv.	495	75	420	17,8	12
Retsudvalget	475	150	325	46,1	16
Miljø og Planlægning	440	45	395	11,4	14
Arbejdsmarkedsudvalget	340	35	305	11,5	10
Socialudvalget	325	25	300	8,3	8
Sundhedsudvalget	300	40	260	15,4	12

*: Formel: $b / (a-b)\%$; a=Total, b=Egne termer, (a-b)=Overlap

Tabel 2. Folketingsudvalg sorteret efter totalt antal tildelte emneord ved tesaurusens start (1992), med antallet af "egne" udvalgstermer i kolonne 2 og antallet af overlappende termer mellem udvalget og andre udvalg i 3. kolonne. Ratioen er forholdet mellem Egne termer og Overlap-termer og angiver graden af unikhed emnemæssigt set.

Det totale antal emneord og mængden af sagsområder, Tabel 2, kan sandsynligvis ses som en indikator for den *komplexitet* som kendetegner de enkelte udvalgs faglige domæne. Vi ser, at Markeds- og Erhvervsudvalgene begge kendetegnes ved stor kompleksitet ved tildelingen af emneord i tesaurusen. Kolonne 2 indikerer hvor *specifikt* udvalgets område kan indekseres. Rets- og Markedsudvalgene ses her at have en høj grad af specificitet. Hver især udgør deres specielle termer ca. 5-8% af det totale antal emneord på 2600 termer. Overlappet af emneord indikerer den *gensidige indflydelse* mellem det enkelte udvalg og andre udvalg. Et begreb som 'CO-2 afgift' er som nævnt relevant i flere udvalg, udover Skatte- og afgiftsudvalget, og befinder sig derfor blandt overlappingsstermerne. Overlappet vil

kunne bestemmes år for år ved hjælp af en matrix, hvor hvert udvalg ses i forhold til alle andre udvalg, og på baggrund af den konkrete anvendelse af termerne på de behandlede udvalgsakter. I den foreliggende form angiver tabellen, at Markeds- og Erhvervsudvalgene begge rækker langt ind over andre udvalgs emneområder og vice versa. *Ratioen*, kolonne 4, er forholdet mellem udvalgets egne specifikke termer og antallet af overlappende termer i udvalgets emneforråd. Jo større procent-tal, jo mere *unik* eller markant og domænespecifikt er udvalget sandsynligvis, sammenlignet med andre udvalg. Man kan konstatere, at Retsudvalget har en sådan markant position i denne henseende, p.g.a. de mange specifikke termer og det forholdsvis lille overlap til andre områder. Mens Markedsudvalget således

indtager en førerposition i alle andre sammenhænge, indtager det her en mellemposition. Finansudvalgets særstatus ses af det totale overlap med alle andre udvalg og dets ratio på værdien 0. Det interessante ved disse og andre *bibliometriske* beregninger, f.eks. over antallet af behandlede udvalgsakter pr. udvalg og pr. emne- eller sagsområde pr. givne tidsenhed, ligger i mulighederne for deres praktiske anvendelse - f.eks. ved ressourcetildeling samt naturligvis i forbindelse med fremtidige politologiske undersøgelser. Tesaurusens *dynamiske karakter* kan således afspejle nogle facetter af det politiske liv. Dynamikken indgår som det væsentligste i vedligeholdelsen af tesaurusen.

5.2 Vedligeholdelsen

Som vist på Fig. 2, kan der foretages visse udtræk fra journaliseringssystemet. Til at vedligeholde tesaurusen anvendes nogle udtræksformer, som bygger på frekvensanalyse af kontrollerede (KE) og ukontrollerede (UK) termer tildelt alle akter indenfor en given tidsperiode, f.eks. igennem et år. Et antal programmer sørger for at udskrive de anvendte KE-ord i faldende frekvensorden, samt i alfabetisk orden med påsat anvendelsesfrekvens. Udskrifterne vil vise hvilke emneord som anvendes meget, ofte og meget lidt. Sammenkørt med selve tesaurus-listen af emneord, vil en udskrift vise, hvilke KE-ord der ikke er anvendt overhovedet indenfor tidsrummet. Det er nu op til en ad-hoc styringsgruppe af domæneeksperter med erfaring i lovgivningsprocessen at afgøre, om de ikke anvendte KE-ord bør udgå af tesaurusen. Samtidig vil man kunne afgøre om der mangler mere præcise (underordnede) ord i tilfælde af storforbrug af enkelte KE-ord.

Samme metode anvendes for de ukontrollerede emneord, som er blevet tildelt akterne i den givne periode. Højfrekvente UK-ord kan blive valgt som nye relevante KE-ord og blive optaget i tesaurusen. Sådanne udskrifter vil naturligvis også afsløre oversete fejlstavninger, samt give et aktuelt billede af de ovennævnte emnefordelinger af udvalgsakter i Folketingets arbejdsproces.

Et resultat af disse løbende vedligeholdelsesprocesser er, at tesaurusen modificeres og som sådan vil blive større som årene går. Man vil derfor kunne tale om en 'retrospektiv' tesaurus, som historisk set afspejler de anvendte emner i de foregående folketingsår, samt en aktuell tesaurus, som afspejler den nuværende brug.

5.3 Indeksring af udvalgsakter

Registreringen og emneindekseringen af udvalgsakterne foretages i Folketinget *decentralt* af de enkelte udvalgssekretærer. Dette er hensigtsmæssigt af flere grunde. Dels går alle aktstykker igennem sekretærens hænder, dels skrives mange udvalgsakter direkte af sekretæren, i samarbejde med udvalgets formand. Det væsentlige i processen består i, at sekretæren kender til de enkelte dokumenter, som passerer udvalget. Emneindekseringen er derfor bedst tjent med, at han eller hun udfører den, samtidig med at resten af journaliseringen eller den direkte udformning i fuld tekst foregår. Tidsmålinger, udført i Trin 6, fig. 3, viser klart, at denne registreringsmåde er tidsbesparende i forhold til alternativet: at emneindekseringen foregår centralt. I den foretrukne metode spares genlæsnings-tiden, som ville være nødvendig for hver udvalgsakt i den centrale instans.

Som nævnt tidligere handler registreringen om to hovedtyper. De med 'Forslag' forbundne akter, og de andre, dvs. dokumenter tilhørende 'Alm. Del'. For 'Forslags-delen's vedkommende registreres de enkelte forslag først til FTR-systemet. Hertil anvendes tesaurusens ordforråd, således at der tildeles op til 6 KE-ord og 6 UK-ord pr. forslag. De med forslaget forbundne udvalgsakter registreres nu i FTU ved at overføre (genbruge) forslagsnumre og andre beskrivende data, samt emnedata. Der kan nu rettes og tilføjes emneord af begge typer så de passer til det aktuelle samhørende dokument. Man kan sige, at *emnekernen* og relevante beskrivende data genbruges, mens bilagsnumre og andre nye data tilføjes, samt evt. nye emneaspekter.

I forbindelse med udvalgsakter tilhørende 'Alm. Del' skal registreringen ske ved en oprettelse, fig. 6.

Oprette spørgsmål		
Folketingsår:	1993-94	Løbenr. : 5451-1
Udvalg:	Markedsudvalget	
Alm.del, sagsomr.:	fiskeri	
Kontrollerede emneord:	fiskerizoner	
Ukontrollerede emneord:	fiskekvoter 93 havfiskeri	
Er emneordene godkendt [J/N]:	N	
Klynge:		
Titel:	spm. XXXX om fiskekvoter 93, havfiskeri, til udenrigsministeren	
Bemærkning:		
Sekretæroppl.:		
Tryk RETUR når indtastningen er færdig		

Fig. 6. Skærmudtræk af registreringsformat til 'Alm. Del'- Spørgsmål, FTU 1994. Med tilladelse af DC.

Fig. 6 viser et registreringsformat, specielt udformet til at registrere Spørgsmål i et udvalg. Andre specielle formater er designet til andre aktyper, men principperne er de samme. Folketingsår og Bilagsnummer udskrives automatisk af systemet. Udvalgsnavnet og det pågældende sagsområde, 'fiskeri', påføres manuelt øverst. Dernæst tilføjes op til 6 KE- og UK-ord: Fiskerizoner; Fiskekvoter 93, Havfiskeri. Systemet sørger nu for at genbruge UK-ordene ved konstruktion af titlen, idet UK-termerne automatisk indsættes i titelbeskrivelsens ordlyd, mellem de 'formatspecifikke' ord "spm. XXXX om"...."til". XXXX tildeles automatisk. Det er altså alene den pågældende 'minister', som skal indtastes i titlen.

'Klynge-feltet' blev konstrueret sammen med 10 predefinerede klynger under designprocessen, for at søgeren kunne få ideer til at bevæge sig rundt i systemet (browsing) og forbedre sin genfindning. Hver klynge var af tværfaglig natur og bestod af op til 25 af de mest

anvendte KE-ord, og blev repræsenteret ved et klyngebegreb, f.eks. Forskning eller Miljø. Da klyngebegrebet, i modsætning til i BASIS-systemet for folkebibliotekerne, var tænkt tildelt hver post i alle udvalg, ville det være muligt at ajourføre klyngerne maskinelt, f.eks. én gang årligt. Feltet anvendes imidlertid sjældent af ressourcemæssige årsager, da den fulde tekst i en udvalgsakt via søgning i sig selv sørger for en højere genfindingsprocent. Egentlig tværgående browsing og navigering er således desværre ikke mulig.

Som hjælp til emneregistreringen findes tesaurusen i trykt form og online, i form af en alfabetisk emneordsliste i Støtteregeisterne, fig. 2. Er indekseren i tvivl om begreber, stavemåde og lignende kan listen lukkes op i et vindue og relevante emneord 'klippes og klistres' over på registreringsformatet. Valideringen af inddateringen foretages på forskellig måde. Under registrering af emneord vil systemet automatisk validere KE-ordene. Hvis der indskrives en term, som ikke

findes på emneordlisten, vil systemet meddele dette og spørge om termen skal indgå som 'nyt' emneord. Hvis ja, kan man efter en given periode beslutte, om termen faktisk bør indgå, f.eks. fordi andre sekretærer også har foreslået samme term. En anden validering foregår ved, at den foregående dags log over registreringer, specielt af UK-ordene, udtrækkes og gennemses. Fejlskrivninger kan konstateres i felter, hvor en term kun er anvendt én gang. Sandsynligheden for at indtaste forkert i flere omgange er naturligvis tilstede, men må anses for minimal.

Oplysningerne i felterne lagres i journalbasens relations-database, mens den fulde tekst lagres i tekstfiler. Det hele overføres jævnligt til BRS-Search, fig. 2 til højre, og bliver tilgængeliggjort via den noget mere avancerede søgeflade i dette system.

6. Søgefaciliteter i FTU

Ifølge domæneanalysens resultater måtte søgegrænsefladen designes på en sådan måde, at det blev gennemskueligt dels hvordan der søges, dels hvordan resultaterne af en søgning umiddelbart kan fortolkes og finde anvendelse i folketingsarbejdet. Søgning direkte ved de Boolske operatører (OG, ELLER, IKKE) skulle undgås. IKKE-operatøren blev af operationelle grunde udelukket fra den avancerede søgegrænseflade. Så mange oplysninger som muligt burde derfor kunne vælges fra eksisterende menuer og nemt indsættes i et *struktureret søgeformat* - og burde ikke skulle huskes og genkaldes af søgeren. Søgegrænsefladen blev derfor designet med flere hjælpefaciliteter i form af vinduer, indeholdende diverse lister over udvalg, folketingsår, og emneord. Den Boolske logik blev indbygget i søgeformatet og tesaurusens relationer udnyttet automatisk under søgning. Nedenfor vil vi kort anskueliggøre resultatet af designet, som det foreligger i sin nuværende form, fig. 7.

Søge i		Total
År: 91 902 901		30.123
Udv.: AMU FIU MKU LFU SOU SUU		
Bilagsdel: Forslag		Antal
Forslagnumre: (901)L 4 (902)B 3 (902)L 123 B		20.519
Søg i alle opl. pensionister		732
og boligstøtte		629
og _____		2.168
_____ eller _____		
_____ eller _____		327
og boligstøtte		2.168
_____ eller _____		
og _____		
Documenttype: H _____		3.421
Af under: C Holdet		87
Omdelingsdato: Fra d. 1/12-90 til d. 27/ 2-92		20.582
Sekretæroppl.: gs/fp		61
	Antal fundne dok. -->	3
Udfyld de ønskede søgefelter (brug evt. SØG) og tryk på RETUR		

Figur 7: Struktureret søgeformat i FTU's avancerede brugergrænseflade. Udtræk 1994. Med tilladelse af DC.

I forbindelse med det frosne 'billede', fig. 7, har søgeren forud valgt folketingsår og -samling(er), samt relevante udvalg fra en predefineret liste. Søgeren skal vælge mindst ét udvalg og én samling. Alt efter kombinationen af et/flere udvalg og en/flere samlinger eller år, fremkommer en (af fire) søgeflader, hvorpå man vælger om man decideret vil søge i 'spørgsmål-svar', eller i bestemte bilagsnumre. Vælges disse muligheder ikke, vil den efterfølgende søgning foregå blandt alle typer af udvalgsakter.

På den omtalte søgeflade, som altså optræder *direkte forud* for billedet, fig. 7, får søgeren opgivet antallet af poster (udvalgsakter), som findes i de valgte udvalg i de valgte år/samlinger. Desuden skal søgeren vælge søgemetode på dette sted i processen. Der sker således en *samtidig søgning* i BRS' Search når disse valg udføres. I vort søgetilfælde har søgeren forud valgt 6 udvalg (AMU=Arbejdsmarkedsudv.; FIU=Finans-; MKU=Markeds-; LFU=Landbrugs- og fiskeri-; SOU=Social-; SUU=Sundheds-). Desuden er årene 1990, 1. og 2. samling, og 1991 valgt. Disse kriterier udgør et *søgeunivers* på ialt 30.123 poster. Denne oplysning står altså allerede på skærbilledet forud for fig. 7, og tjener til at vejlede søgeren, f.eks. ved at han/hun allerede her indsnævret til færre år eller udvalg. Valget af *søgemetode* foregår også på denne forudliggende skærmside, og er bestemmende for, hvad søgeren får at se direkte på billedet, fig. 7.

Der er 3 *søgemetoder*, som alle vedrører svarkolonnen til højre på fig. 7: 1) *Fuldt svar*-metoden, som vil angive alle mellemresultater *direkte* for hvert nyt søgeargument der indtastes - når det pågældende 'felt' forlades. På skærmen, fig. 7, vil søgeren altså umiddelbart kunne se, at der er 20.519 udvalgsdokumenter i 'Forslags-delen' af FTU indenfor det allerede valgte søgeunivers. 2) *Mellem-svar*-metoden, som giver alle de mellemliggende svar, men først *efter* at alle søge-felter er udfyldt, og den endelige søgning er foretaget med et slutresultat nederst til højre. Kolonnen til højre vil da se ud som på fig. 7. 3) *Slut-svar*-metoden, som alene angiver det endelige slutresultat af søgningen - men ingen mellem-svar. På skærmen, fig. 7, angives kun slutresultatet på 3 fundne dokumenter.

Metode 1, *Fuldt svar-metoden*, er praktisk hvis man ikke er sikre på brugbarheden af sine indførsler i søgningen. Man kan direkte gå 'tilbage' og forandre indholdet i et søgefelt, hvis f.eks. et søgeargument giver

meget få svar. Fuldt svar-metoden er avanceret i sin brugervenlighed. De to andre online søgemetoder er af mere traditionel natur.

I eksemplet forudsætter vi altså, at søgeren har valgt de omtalte folketingsår og udvalgene, som angivet på figuren. Desuden forudsættes, at Fuldt svar-metoden er valgt som søgemåde. Søgeren véd med andre ord, at der findes ialt 30.123 udvalgsdokumenter i søgeuniverset. I feltet 'Bilagsdel' vælges nu 'Forslag' - i modsætning til alternativet 'Alm. Del' - som direkte giver 20.519 poster. Da vi er i forslagsdelen, kan man nu indtaste kendte forslagsnumre, hvis det ønskes. Når søgeren forlader dette felt, for at bringe cursoren ned til 'Søg i...'-feltet, oplyser systemet, at der findes 732 udvalgsakter tilknyttet de indtastede forslagsnumre. Dette antal poster udgør alene et mellemsæt, ikke et foreløbigt slutresultat.

De næste felter vedrører *emnedata*. På fig. 7 findes 7 linjer under etiketten 'Søg i *alle opl.*' Dette er default-søgemåden i posternes forskellige felter som indebærer, at alle termer der skrives i de næste 7 linjer vil blive søgt i *alle* BRS' søgbare felter. De 7 linjer er *på forhånd strukturerede* i forhold til OG og ELLER operatorene i den Booleske logik. Det er klart, at størsteparten af termene angivet i disse felter har emnekarakter og derfor naturligt bør findes i følgende felter: Titel-, Fuldtekst-, Kontrollerede og Ukontrollerede emneordsfelter. F.eks. i forbindelse med indskrivningen af termen 'pensionister', kan man ved et enkelt greb (ved at trykke på en SØG-tast) gå ind i BRS' alfabetiske ordliste og kontrollere stavemåde og form. Vi skal her bemærke, at søgningen på termene sker med automatisk *tilsluttet synonymkontrol* via tesaurusen. Denne form for 'skjult' 'synonym'-kontrolleret søgning kan genfinde et forholdsvis stort antal poster, da søgning i fuldteksten er inkluderet. For eksempel vil antal fundne poster for termen 'boligstøtte', ialt 2.168 dokumenter, automatisk indeholde dokumenter fundet via tesaurusens (nær)synonymer 'boligsikring' og 'boligydelse', hvor disse tre termer end findes i de ovennævnte BRS felter. Med hensyn til den fastlagte lodrette struktur for de Booleske operatører OG og ELLER kan denne facon siges at være en smagssag. Som søgeeksemplet viser, skal søgeren gentage termen 'boligstøtte', for at søgelogikken bibeholdes, dvs. (pensionister ELLER efterlønsmodtagere) OG boligstøtte. Lidt aparte måske, men hvis man tænkte sig en bruger, som søgte på (pensionister OG boligtilskud)

ELLER (efterlønsmodtagere OG boligstøtte), er den viste, fastlagte struktur udmærket. Personlig ville jeg nok foretrække ELLER-operatoren fastlagt vandret og OG-operatoren i det lodrette plan.

Efter at data evt. er indtastet i Dokumenttype, Afsender og Omdelingsdato angiver systemet antal poster for disse mellemsæt, og det *endelige søgeresultat* fremkommer = 3 poster. Det skal bemærkes, at hvis søgeren enten staver forkeret eller skriver ord og ordstreng, som ikke findes i BRS' samlede ordbog, vil en meddelelse angive dette faktum. Man kan ikke forlade feltet uden at lade det være tomt eller udfyldt med et genkendt ord. Søgelogikken er nem at overskue for den professionelle søger: alle søgeargumenter vandret afsøges med ELLER-operatoren, mens alle lodrette felter anvender OG-operatoren, med undtagelse af emnedatafeltet, hvor strukturen som sagt brydes. Hele søgningen med mellemsæt foregår som nævnt indenfor det allerede afgrænsede søgeunivers, her på 30.123 poster.

Fordelen ved denne grænsefladeopbygning er, at det er muligt at *indsnævre søgningen* til alene at omfatte emneord. Ved at trykke på en funktionstast, skifter Søg i 'alle opl.' til Søg i 'emneord'. Dette kan enten gøres *inden* emnedata-linjerne udfyldes eller *efter* en søgning i 'alle opl.' allerede er foretaget, som vist på fig. 7. Tallene i højre kolonne vil nu skifte til et mindre antal poster, da søgningerne med termerne kun foretages blandt KE- og UK-ord. I nogle tilfælde vil antallet derfor blive 0. Denne operation tager en vis tid at udføre, på grund af datamatens kapacitet og som følge af, at alle emnedatasøgninger skal foretages på ny. Man vil i løbet af processen kunne slå op i tesaurusen online og overføre relevante emneord til søgeformatet.

Uanset om der søges i 'alle opl.' eller i 'emneord' kan der venstre- og højretrunkeres, dog indenfor visse kapacitetsgrænser. Klippe- og klistre-funktionen kan anvendes til at overføre termer fra én linje til en anden og fra fremviste ordlister til søgeformatet.

7. Konkluderende bemærkninger

FTU er tilgængelig for sine primær-brugere, men ikke for offentligheden. Det vides ikke om, og evt. hvornår, en sådan tilgængelighed bliver iværksat. En del af indholdet i Folketingets informationssystem er tilgængeligt via Retsinformation, som også opererer på BRS-

Search, men med andre former for emnedata. De interne udvalgsakter, f.eks. spørgsmål, findes som sagt søgbare i fuld tekst. Derimod findes svarene, udformet i respektive ministerier, endnu ikke i elektronisk fuld tekst-form. Dertil kræves overførsel via telekommunikation, en udbygning, som endnu ikke er gennemført.

Den producerede almindanske emneordstesaurus er sandsynligvis anvendelig i flere andre sammenhænge udover i Folketingets Informationssystem, f.eks. i fagministerier eller styrelser. Den kan rekvireres, eventuelt i modificeret form, ved henvendelse til forfatteren.

Der er ingen tvivl om, at FTR/FTU tilsammen udgør en rig kilde til såvel ren informationsindhentning, som til samfundsvidenskabelige undersøgelser af forløbet af den danske politiske proces. På grund af emneordsstrukturen og navneformskontrollen for institutioner og virksomheder findes her et oplagt potentiale for veltilrettelagte undersøgelser af politologisk art.

Indenfor rammerne af den teknologi, som fandtes i slutningen af 80-erne m.h.t. terminalnetværk, programmel og maskinkraft, er den implementerede løsning et stærkt arbejdsredskab i Folketingets daglige arbejde. Der er så vidt muligt taget hensyn til de flaskehalsproblemer, som erfaringen viser der opstår, f.eks. i forbindelse med den årlige finanslovsvedtagelse og i begyndelsen af hver folketingssamling. Netop det domæneafhængige erfaringsgrundlag er en væsentlig faktor i moderne systemudvikling. Det snævre samarbejde mellem systemdesignerne, -udviklerne og de erfarne repræsentanter for primærbrugerne er i det foreliggende systemarbejde lykkedes i et stort omfang.

Selve FTU's funktionsmåde er søgt gjort brugervenlig, både på registrerings- og søgesiden. Alt i alt tvinger den viste implementering af søgegrænsefladen brugeren til at være mere bevidst om sine søgekriterier, end man sædvanligvis ser i kommandostyrede søgesystemer. På grund af datamængden er specifikation i søgninger en absolut nødvendighed. Den i afsnit 6 omtalte fleksibilitet under emnesøgning har naturligvis sine omkostninger. Systemet bliver langsommere end godt er. Man må derfor håbe, at datakraften forøges mærkbart indenfor en kort årrække. Desuden vil det måske være en idé, at forsøge at indføre PC-ere og en Windows-baseret BRS-løsning, eventuelt i etaper begyndende med registreringsfunktionen. Set fra et informationsvidenskabe-

ligt synspunkt har udviklingen af dette fungerende in-house system været tankevækkende og dets videre udvikling bliver interessant at følge og værd et nærmere studium.

Referencer

Alnor, Karl og Sørensen, Jens Otto. Introduktion til Systemanalyse. 2. udg. Århus: Tenakel, 1991.

Martyn, J. and Lancaster, F.W. Investigative Methods in Library and Information Science. Arlington, Virginia: Information Resources Press, 1981 (or later eds.).

Wood-Harper, A.T.; Antill, Lyn; Avison; D.E. Information Systems Definition: The Multiview Approach. Oxford: Blackwell Scientific Publ., 1985.