

UNIVERSITETET I OSLO
Institutt for rettsinformatikk

Rettskilder og hyperstrukturer

Om alternative
grensesnitt til rettslige
informasjonssystemer

Kandnr.: 395

Veileder: Jon Bing

Lev.frist: 25.04.2006 (V-06)

17.856 ord

April 2006



Innhold

1	Innledning	1
1.1	Juristens oppgaver	2
1.2	Rettslige informasjonssystemer	4
1.3	Relevans	6
1.4	Data, informasjon og rettskildefaktorer	7
2	Søkeprosessen	10
2.1	Indeksering	12
2.2	Utforming av søkeargument	14
2.3	Siling	15
2.3.1	Boolsk sammenlikning	15
2.3.2	Vektorbasert sammenlikning	16
2.4	Kriterier for søkeeffektivitet	17
2.5	Fritekst-søk	19
3	Rettskildefaktorer i rettslige informasjonssystemer	20
3.1	Fra tekst til tall – et glimt fra virkeligheten	21
3.2	En rettskildefaktor blir et dokument	21
3.3	Søking	24
3.4	Bristende forutsetninger	25
3.5	Rettskildefaktorer danner strukturer	26
3.5.1	Lover som hypertekst	26
3.6	Henvisninger	27
4	 Navigasjonsverktøy	31
5	Format, metode, teknologi	34
5.0.1	ZigZag og ZZStructure	35
5.0.2	RDF	35
5.0.3	Emnekart	36
5.1	Hvorfor Emnekart?	37

5.2	Kort om Emnekart	38
5.2.1	Emner	40
5.2.2	Assosiasjoner	41
5.2.3	Forekomster	43
5.3	Muligheter	44
6	Ontologisk analyse av rettskildefaktorene	46
6.1	Tilgjengelige data	47
6.2	Ontologi-begrepet	48
6.3	Analyse av dokumentbasen NL	49
6.3.1	Emner	49
6.3.2	Forekomster	51
6.3.3	Assosiasjoner	52
6.4	Analyse av dokumentbasen HRSIV	54
6.4.1	Emner	54
6.4.2	Forekomster	55
6.4.3	Assosiasjoner	56
6.4.4	Perspektiv	57
7	Fra lov til emnekart. Et forsøk.	58
8	Representasjon og navigasjon	62
8.1	Representasjon	63
8.2	Navigasjon	64
8.3	Visualiseringsmetoder	64
8.3.1	Grafer og trær	65
8.3.2	Kart	67
9	Resultatene og veien videre	70
A	En avgjørelse i XML-format	75
B	En avgjørelse i Lovdatas Bokus-format	77
C	Bustadsoppføringslova i LTM-format	79

Kapittel 1

Innledning

— *Hva skriver du om?*
— *Om hyperstrukturer og grensesnitt.*
— *Hyper hva?!*

Tema for denne avhandlingen er *Rettskilder og hyperstrukturer*. Når det gjelder rettskilder, er disse kjent stoff som sitter i benmargen hos alle med juridisk skoloring av en viss varighet. Situasjonen er verre når det kommer til hyperstrukturer. Meget få skjønnte hva en hyperstruktur var for noe, og av de var det ikke mange som kunne forstå hvilken interesse kunne disse rare strukturene ha for en jurist¹

Denne avhandling skrives ved Det juridiske fakultet, men vil i sitt innhold være svært tverrfaglig og berøre så vidt forskjellige disipliner som jus, informatikk, semiotikk, logikk og kunnskapsforvaltning. Det forutsettes ikke at leseren har noen kjennskap til andre områder enn jus og det vil bli gitt en løpende innføring og forklaring der disse er på sin plass. Mye av det jeg kommer til å skrive om vil dessuten være nytt stoff for meg, og derfor vil skriveprosessen samtidig være en læringsprosess, som jeg skal forsøke å dokumentere så godt det lar seg gjøre.

¹Henvisningene vil spille en sentral rolle videre, derfor viser jeg direkte til avsnitt 3.5 på side 26 for en forklaring av hva jeg mener med hyperstrukturer.

Et innledende avsnitt passer godt for å fortelle litt om tanken som har vært drivkraften bak dette arbeidet.

Søking etter rettskilder foregår i stor utstrekning ved hjelp av EDB-baserte rettslige informasjonssystemer. Disse systemene bygger på nokså gammel teknologi for tekstanalyse. Som alt annet, har den både sterke og svake sider, og den er langt fra perfekt. Den bakenforliggende teknologien legger dessuten sterke føringer på hvordan grensesnittet utformes² og dermed også på hvordan selve søkeprosessen foregår.

Tanken er at *det er mulig å søke annerledes*. Det krever at man må begynne å tenke annerledes. Nye søkemetoder lar seg heller ikke realisere, hvis det ikke samtidig innføres nye typer søkegrensesnitt.

Det finnes en rell mulighet for å søke annerledes, og denne muligheten var inspirasjonskilden til denne oppgaven. Muligheten ligger i hyperstrukturer, som også gir en mulighet til å bytte ut det gamle søkefeltet med andre typer søkegrensesnitt, som er både kraftigere og mer fleksible.

Det blir nok for mye å kreve at en normalt oppegående norsk jurist skal kjenne til de tekniske detaljene bak eksisterende systemer for tekstsøking. Å ha en viss oversikt over og formening om hvordan søkeprosessen foregår er imidlertid helt vitalt for å få klarhet i årsakene til at disse systemene ikke kan fungere godt nok, og at det kan være vel verdt innsatsen å se på alternativene. I de første kapitlene vil jeg derfor presentere prinsippene bak tekstsøking i informasjonssystemer. Sammen med denne presentasjonen vil jeg introdusere og avklare en del viktige begreper som vil spille en sentral rolle videre i oppgaven.

1.1 Juristens oppgaver

La meg først bruke noen ord til å minne om oppgavene til en jurist.

Jurister skal løse rettslige problemer. Det kuriøse er at rettslige

²Hva skulle årsaken ellers være til at alle søkesystemer ser omtrent like ut?

problemer eksisterer ikke for andre enn jurister selv: når *NN* står overfor et problem som han *tror* en jurist kan hjelpe til å løse, henvender vedkommende seg til en jurist, legger fakta på bordet og gjør rede for hva slags *faktisk* problem som har oppstått og som ønskes løst. Juristen analyserer fakta og ved hjelp av sin bakgrunnskunnskap, erfaring og kreativitet oppstiller en *rettslig* problemstilling ved å skrelle vekk det som ikke er juridisk relevant og forsøker deretter å besvare den oppsatte problemstillingen ved å prøve å finne en rettslig norm som kan anvendes på det faktiske forholdet. Rettslige normer har et hvis-ledd (antecedent) og et så-ledd (konsekvent). Juristen prøver å argumentere for at det finnes en rettslig norm hvis antecedent stemmer overens med problemets faktum, og hvis konsekvent vil følgelig gi anvisning på følgen av at rettsnormen kommer til anvendelse.

Et problem er “rettslig” hvis man ved hjelp av rettslig argumentasjon kan finne en løsning på problemet.[3]

Det rettslige problemet kan ikke løses etter rettsanvenderens eget forgoftbefinnende – juridiske spilleregler må følges og juristen må argumentere for en gitt rettslig norm ut fra tilgjengelige og relevante rettskilder og følge den juridiske metoden i sin argumentasjon. Styrken av argumentasjonen vil avhenge av hvor godt rettskildemessig fundament juristen klarer å bygge sin drøftelse på.

De fleste spesialoppgaver ved det juridiske fakultet inneholder noen ord om rettskilder. Jeg ønsker ikke å bryte med denne tradisjonen, men jeg kommer heller ikke til å gjenta det kjente og banale, derimot kommer jeg til å se nærmere på selve prosessen som finner sted når man søker etter relevante rettskilder samt fremheve og sette navn på enkelte aspekter ved denne prosessen.

1.2 Rettslige informasjonssystemer

En sannhet som overrasker mange lekmenn er at rettsregler står slett ikke i den tykke røde boken som juridiske studenter på første avdeling alltid bærer med seg. Hadde rettsregler vært noe man bare kunne finne, ville jurister fort bli overflødige, eller deres rolle ville være redusert til å være veivisere til steder hvor rettsregler kunne finnes. En sjelden gang kan man finne svar på spørsmålet i en lovbestemmelse hvis forståelse ikke gir noen som helst grunn til tvil, men det vanlige er at man bare finner materiale for vurderinger og argumentasjon. Dette materiale for rettslig argumentasjon er hva man mener med rettskilder.

Rettskilder finner juristen i samlinger som jeg vil kalle rettslige informasjonssystemer. Ordet "informasjonssystem" leder tanken hen mot ny teknologi, Internett, Lovdata, men det er ikke slik at et slikt system nødvendigvis er EDB-basert. "Informasjonsteknologi" blir i stor grad brukt som et synonym for EDB-teknologi, og det er nok denne bruken som er årsaken til at alle sammensatte ord som inneholder "informasjon" umiddelbart forbindes med datamaskiner. Med et informasjonssystem forstås en mekanisme for overføring av informasjon fra senderen til mottakeren; når det er rettslig informasjon som overføres, kalles hele systemet et rettslig informasjonssystem.[3, s. 15] Begrepet er teknologinøytralt, og den trykte samlingen Norges Lover er således et informasjonssystem for overføring av informasjon om gjeldende lover fra lovgiver (sender) til rettsanvender (mottaker). Når det er presisert at rettslige informasjonssystemer ikke behøver å være EDB-baserte, har jeg likevel min oppmerksomhet rettet mot de nye mulighetene som nettopp EDB-baserte rettslige informasjonssystemer gir. For å illustrere likhetene og forskjellene vil jeg ofte vende blikket bakover mot bøker o.l.

Det har alltid vært vanlig å dele rettskildene opp i forskjellige kategorier. Echkoﬀ[9, s. 23] gjør også dette og hans inndeling passer fint for meg i mitt arbeid. De forskjellige typene rettskilder har tradisjonelt vært å finne i forskjellige informasjonssystemer: lovteksten finner man i lovsam-

lingen, høyesterettsdommer i Rettsidende osv. For å få et tilfredsstillende grunnlag til å løse et rettslig problem, må derfor et tilstrekkelig antall rettslige informasjonssystemer benyttes. Lovsamlingen har man stort sett alltid for hånden, og man har ofte et mer eller mindre omfattende bibliotek med juridisk litteratur tilgjengelig. Når jeg skriver dette, kan jeg i tillegg til lovsamlingen se Skoghøys Tvistemål, Hagstrøms Obligasjonsrett, Knophs Oversikt over Norges Rett og Prosessnøkkelen; jeg har dessuten NOU-ene til den nylig vedtatte tvisteloven stående i hyllen. Litteraturen rundt meg gjør meg nok så godt rustet til å løse et problem innenfor obligasjonsretten eller sivilprosess. Skulle jeg derimot stå overfor et strafferettslig spørsmål, måtte jeg nok oppsøke et bibliotek eller spørre en kollega. En kollega fungerer dermed også som et informasjonssystem, men det er nyttig å huske at det vedkommende måtte si vil veie lite i juridisk argumentasjon. Brukerens samling av forskjellige informasjonssystemer vil danne et slags "brukerkonstruert" informasjonssystem.

Informasjonssystemets tilgjengelighet er en viktig og ofte avgjørende faktor for om informasjonssystemet blir benyttet. Med tilgjengelighet forstås her tilgang på rettslig informasjon.[3, s. 85] Økt bruk av ressurser, først og fremst i form av tid, penger, gir tilgang på økt mengde (presumptivt) relevant informasjon. Mitt strafferettslige spørsmål kan tjene som et eksempel: jeg har verken dommer eller litteratur for hånden og må oppsøke et bibliotek for å få tilgang til disse. Både samlinger av dommer og litteratur er rettslige informasjonssystemer, ja, selve biblioteket er jo også et informasjonssystem. Hvis spørsmålet bare er av typen "kjekt å vite", er det trolig at jeg ikke finner det bryet verdt å gå til biblioteket; om jeg går eller ikke vil altså bero på om nytten av økt informasjon er større enn kostnadene forbundet med å skaffe seg denne informasjonen. Av hensyn til rettssikkerhet og rettferdighet er det viktig at juridiske avgjørelser treffes på så godt grunnlag som mulig, men hensyn til bruk av ressurser setter grenser for hvor langt man kan gå for å oppnå målene for rettssikkerhet og rettferdighet. — Moderne EDB-baserte rettslige informasjonssystemer, f.eks. WESTLAW og Lovdata, tilbyr nå

en søkbar, elektronisk tilgjengelig, oppdatert rettslig informasjonstjeneste med tilgang til mange databaser med juridisk materiale, og iallfall spørsmålet om tilgang til samlingen av en gitt type rettskilder er ikke lenger like aktuell når juristen har tilgang til det EDB-baserte systemet, som blir et slags multi-informasjonssystem.

Med innføring av og overgang til EDB-baserte rettslige informasjonssystemer, endrer spørsmålet karakter fra å være et spørsmål *om* noe er tilgjengelig til å være et spørsmål om *hvor* i informasjonssystemet det er. Mengden av det rettslige materialet som må undersøkes for å finne det som eventuelt måtte være relevant vokser raskt og ubønnhørlig. Martin Støren har gjort en empirisk undersøkelse av veksten i Norsk Lovtidend i perioden 1990–2001.[16] Tiden det tar å finne relevant rettskilder setter skranker for hvor lang tid en jurist er villig til å bruke på sitt søk — å finne det relevante forblir en utfordring, selv med ny teknologi.

1.3 Relevans

Jeg har nevnt ordet “relevant” i forbindelse med søking etter rettskilder flere ganger. I rettskildelæren er relevans en del av den kjente sekvensen ‘relevans, slutning, vekt’. Eckhoff anser som relevant det som “*ifølge rettskildeprinsippene er tillatt å ta i betraktning*” når man skal avgjøre et rettsspørsmål.[9, s. 22] Dette blir et for bredt relevansbegrep, ettersom mesteparten av innholdet i de rettslige informasjonssystemene er “relevante”, hvis man skal se på om det er tillatt å ta disse i betraktning i rettslig argumentasjon; relevans i denne betydning vil være bestemt av rettslige meta-normer og vil sjeldent by på tvil. For en jurist som står overfor et konkret problem har relevans en annen og snevrere betydning — juristen vil anse som relevante de rettskildefaktorene, som kan bidra med rettslige argumenter, det er disse han er ute etter og i dette perspektiv vil relevans være direkte avhengig av problemet juristen står overfor. I en slik situasjon vil en rettskildefaktor være relevant hvis:

- (i) Brukerens løsning av det rettslige problem ville blitt annerledes hvis brukeren ikke hadde kjent innholdet i faktoren, dvs. at det av faktoren kan utledes minst ett argument;
- (ii) rettslige meta-normer krever at brukeren skal undersøke hvorvidt faktoren tilhører gruppe (i); eller
- (iii) brukeren finner det selv nødvendig å vurdere hvorvidt faktoren tilhører gruppe (i). [3, s. 47]

Noe forenklet kan man si at de dokumenter er relevante, som brukeren kan finne nyttige for argumentasjonen eller som kan påvirke denne.

1.4 Data, informasjon og rettskildefaktorer

“Data” og “informasjon” er to ord som brukes mye og ofte som om begge betyr det samme. Jeg skal bruke disse to på samme måte som Matthijssen[14, s. 22-23] og Bing[3, s. 66-67]: “data” om en formalisert representasjon av visse opplysninger, mens “informasjon” får et semantisk innhold og vil gjelde den verdi eller mening data kan ha for en bruker, som forutsetningsvis er et menneske. Informasjonssystemet gir brukeren tilgang til data (i form av tekst), og data kan i brukeren hode bli omgjort til informasjon, hvis brukeren har de nødvendige forutsetninger for å forstå budskapet som data er ment å bære. Forholdet mellom data og informasjon er av samme type som forholdet mellom ord og mening.

Som nevnt i avsnitt 1.2 på side 4, er et informasjonssystem en mekanisme for overføring av informasjon fra senderen til mottakeren. I forrige avsnitt ble informasjon definert til å være et resultat av et menneskes forståelse av data, informasjon er noe som oppstår som følge av en kognitiv prosess og kan følgelig ikke overføres *per se*. Derfor må informasjon uttrykkes på en måte som er egnet for overføringen fra senderen til mottakeren. Overføringen skjer ved at senderen koder informasjon som data, sender data til mottakeren og denne kan få informasjon ved å dekode de

mottatte data. Denne kodingen av informasjon til data og dekodningen av data til informasjon er noe vi alle gjør hele tiden når vi snakker, skriver, hører og leser: tanker blir formulert i setninger som overføres ved lydbølger eller på papir.

I rettslige informasjonssystemer overføres data i form av *dokumenter*. Et dokument er et problematisk begrep. Forvaltningslovens § 3 definerer et dokument som “*en logisk avgrenset informasjonsmengde som er lagret på et medium for senere lesing, lytting, framføring eller overføring*”. Denne definisjonen får kritikk av Bing.[4] Når jeg videre snakker om dokumenter, vil jeg stort sett³ mene den *representasjon* som en rettskildedefaktor får i et informasjonssystem.[3, s. 42] Med en slik definisjon unngår jeg problemer som de forskjellige elementene i forvaltningslovens definisjon måtte medføre.

En og samme rettskildedefaktor kan i forskjellige systemer være representert på forskjellige måter, avhengig av hvordan systemet er realisert, hva systemets formål er og hvilken brukergruppe det tar sikte på å betjene: i et informasjonssystem kan en dom kan være representert bare ved et sammendrag med en henvisning til et annet informasjonssystem hvor dommen kan finnes i fulltekst. Tre typiske former for dokumenter kan nevnes[3, s. 122–135]: stikkordkarakteristikker, sammendrag, autentisk tekst. Disse tre formene ivaretar forskjellige funksjoner som rettslige informasjonssystemer kan tenkes å ha: søkefunksjonen, relevansvurdering og kildefunksjonen. Søkefunksjonen innebærer at brukeren av et system kan instruere systemet om å finne relevante dokumenter, i relevansvurdering bestemmer brukeren om de dokumentene som systemet fant virkelig er relevante i forhold til den oppgaven som brukeren har foran seg, og kildefunksjonen går ut på at brukeren skal kunne bruke rettskildefaktorene som dokumentet gir uttrykk for i rettslig argumentasjon. Et par av de forskjellige funksjonene kan kort illustreres ved en tenkt situasjon med en jurist som har en referanse til en dom i Rettstidende⁴. Han begynner ikke

³Dvs. med mindre noe annet klart fremgår av sammenhengen.

⁴Eksempelet viser ikke søkefunksjonen, altså hvordan juristen fant frem til dommen.

å bla gjennom alle årgangene av Rettstidende, men tar riktig årgang og slår opp på ønsket side. Årgang og sidetall er en referanse til dommen. I begynnelsen av dommen vil det være et sammendrag. Sammendraget tar kort tid å lese og leseren vil meget fort finne ut om dommen virkelig er relevant (relevansvurdering), men rettskildemessig veier sammendraget lite og har status som juridisk teori og brukes ikke direkte i juridisk argumentasjon, juristen må derfor bygge på dommen som ligger til grunn for sammendraget (kildefunksjonen).

Ordet "rettskildefaktor" vil i min fremstilling få en noe snevrere betydning enn i vanlig juridisk sjargong. Echhoff[9, s. 23] deler rettskildene i flere kategorier rettskildefaktorer. Noen vil være representert ved tekst, mens andre ikke kan komme fysisk til uttrykk. Eksempler på rettskilder fra disse to gruppene er lovtekst og reelle hensyn. Lovteksten publiseres i Lovtidend, mens reelle hensyn er vurderinger, de er "*produkter av rettsanvenderens egne overveielser*".[9, s. 24] Slike overveielser, vurderinger og andre ikke-tekstlige rettskilder (f.eks. sedvane) egner seg ikke for behandling og jeg vil derfor holde oppmerksomheten rettet bare mot de rettskilder som har fått et skriftlig uttrykk. Jeg kommer altså til å forbeholde betegnelsen rettskildefaktorer bare de rettskilder som har fått et skriftlig uttrykk og dessuten *bare* om dette skriftlige uttrykket. Årsaken til disse avgrensningene er at avhandlingen dreier seg om EDB-baserte rettslige informasjonssystemer, og disse systemene inneholder bare skriftlige rettskilder og har for tiden bare mekanismer for å håndtere disse kildene *qua* data, ikke *qua* informasjon. Bing bruker "rettskildefaktor" i denne snevrere betydning i sin doktoravhandling.[3, s. 26-28]

Kapittel 2

Søkeprosessen

Relevante rettskilder må søkes og bli funnet for å kunne bli brukt i juridisk argumentasjon. Å søke betyr ikke nødvendigvis å finne det man er ute etter, og det er heller ikke alltid slik at man vet hva man søker etter. Man skiller normalt mellom to typer søk: faktasøk og emnesøk (også kalt interessesøk). Ved faktasøk leter man etter et svar på et konkret spørsmål, mens ved emnesøk forsøker man å finne dokumenter som behandler et gitt emne. Et søk på personens navn i Telefonkatalogen vil gi telefonnummer og evt. adresse og er faktasøk, mens å søke på personens navn i Googles søkesystem er emnesøk. Det er verken nødvendig eller mulig å trekke en skarp grense mellom faktasøk og emnesøk, det er vurderingen av hva som er relevant som er forskjellig: ved faktasøking kan man nøyaktig angi hvilke betingelser et dokument må oppfylle for å være relevant.

I juridisk søking er det emnesøk som dominerer. Dette skyldes at svar på juridiske problemer ikke er gitt på forhånd, man finner råmateriale for argumenter som kan støtte opp et synspunkt.

Formålet med et søk i et informasjonssystem er å lokalisere dokumenter hvis meningsinnhold helt eller delvis dekker brukerens informasjonsbehov eller å slå fast at slike ikke finnes. Systemet må altså inneholde dataanalysemekanismer som er i stand til å lokalisere dokumenter som inne-

holder informasjon som er relevant for denne brukeren der og da; for en jurist som skal løse et rettslig problem betyr det at systemet må lokalisere rettskildedefaktorer som kan bidra med rettslige argumenter ved løsning av et rettslig problem.

Som regel må det gjøres forberedende arbeid hvis et informasjonssystem skal tilby en søkemulighet. Uten et slikt forberedende arbeid ville alternativet være å arbeide seg gjennom tekstene til samtlige dokumenter i informasjonssystemet hver gang søket skulle gjøres, og dette ville kreve mye tid. Hvis et slikt sekvensielt fulltekst-søk gjøres manuelt, er det bare et alternativ dersom informasjonssystemet inneholder relativt få relativt korte dokumenter. Når tekstmengden blir stor, er et automatisert sekvensielt fulltekst-søk heller intet godt alternativ i praksis, for også dette vil fort bli svært ressurs- og tidkrevende.

For bøker vil forberedende arbeid typisk være utarbeidelse av registre (f.eks. stikkordregister eller systematisk register); i EDB-baserte systemer skjer det liknende prosesser der det etableres datastrukturer som er spesielt egnet for søking.

Brukerens informasjonsbehov, dvs. hva brukeren ønsker å finne informasjon om, eksisterer bare i brukerenes hode og må uttrykkes på en måte som informasjonssystemet kan godta og "forstå". Informasjonssystemets oppbygning setter skranker for hvordan dette informasjonsbehovet kan uttrykkes: et stikkordregister vil ikke hjelpe å lokalisere hvor i boken en dom er omtalt, da må man bruke et domsregister. Innholdsfortegnelse kan også betraktes som en type register og forhåpentligvis gir en god oversikt over innholdet i en bok. Med datamaskinbaserte systemer er mulighetene større, og det er større fleksibilitet for så vidt gjelder måter for å henvende seg til et elektronisk register.

Ved søk i et informasjonssystem kan det pekes på tre prosesser: den forberedende *indeksering* av det søkbare materiale, brukerenes utforming av *søkeargument* og systemets *siling* eller utvalg av relevante dokumenter.[14, s. 25] Register i bøker utarbeides ved indeksering, leseren utformer et

søkeargument (oppslagsord) som kan brukes for å gjøre et oppslag i registeret og ved siling prøver leseren å finne søkeargumentet i registeret. Disse tre trinn i en søkeprosess kan identifiseres både i digitale og trykte informasjonssystemer, men jeg ønsker primært å belyse søkeprosessen i EDB-baserte informasjonssystemer.

2.1 Indeksering

Indeksering er navnet på prosessen der et dokument tilordnes stikkord (termer) som beskriver dokumentets innhold. Stikkordet (termen) kan være et enkelt ord eller et uttrykk. Et register i en bok er et resultat av en indeksering og registeret kan inneholde emner, navn, steder og annet som måtte få en omtale i boken. Registeret har dessuten henvisninger til hvor i boken en indeksert term er omtalt.

Indeksering kan være manuell eller automatisk. Ved manuell indeksering er det typisk forfatteren som ved hjelp av stikkord angir hvilke emner teksten dreier seg samt hvor i teksten disse får en omtale. Manuell indeksering er *intellektuell*. Den som leser teksten forstår innholdet og kan beskrive teksten med et stikkord som ikke må fremgå av selve teksten, men kan være underforstått eller komme av sammenhengen. Ulempen med manuell indeksering er at prosessen krever mye menneskelig arbeid (i all-fall hvis store tekstmengder skal indekseres). Automatisk datamaskinbasert indeksering skjer svært fort, men ulempen er at det per i dag er ingen muligheter for å "forstå" teksten; prosessen går derfor ut på å lage en liste over alle ord i teksten med en referanse til hvor ordene forekommer. Det legges til grunn at ordene som forekommer i dokumentet gir en beskrivelse av innholdet i dokumentet. Denne listen med nøkkelord og referanser til dokumenter kalles en invertert fil eller en invertert indeks. Invertert betyr omvendt og beskriver den omvendte relasjonen mellom nøkkelord og opprinnelig tekst: innholdet i et dokument *handler om* nøkkelord og nøkkelord *beskriver* innholdet i dokumentet. Lovdata bruker et søkesys-

tem som heter SIFT¹. SIFT, Google, FAST og andre søkesystemer bygger på inverterte filer ved søk i systemenes databaser. Ved å bruke inverterte filer kan systemet raskt finne ut om og i tilfelle i hvilke dokumenter et bestemt ord forekommer. For å øke søkehastigheten er inverterte filer ofte implementert som B-trær. Et B-tre er en hierarkisk datastruktur der elementene er ordnet på en måte som tillater svært raske søk etter elementer selv om antallet elementer er stort.

Hvilke nøkkelord/stikkord som brukes for å beskrive innholdet kan være gitt på forhånd, f.eks. i en tesaurus², men indeksereren kan også tenkes å ha helt frie hender når det gjelder valg av nøkkelord. Fritt bruk av nøkkelord i indekseringsprosessen kan bidra til fleksibilitet og nøyaktighet i registrene i den forstand at nøkkelordene på best mulig måte beskriver innholdet, men fritt bruk av nøkkelord kan fort gå ut over konsistens. Med konsistens forstås her at det finnes regler eller konvensjoner for når et ord skal brukes, og at i en gitt situasjon vil bare et gitt ord brukes og at et gitt ord bare brukes i en gitt situasjon. Ved manuell indeksering er det ikke bare slik at forskjellige personer kan velge å beskrive innhold i samme dokument med forskjellige stikkord, men også at samme indekserer kan velge å bruke andre termer når samme dokument skal indekseres på nytt. Bruk av en forhåndsdefinert liste av nøkkelord kan bidra til høyere grad av konsistens i registrene, men denne metoden kan i sin tur gå ut over fleksibiliteten. Spørsmålet om konsistens melder seg ikke ved automatisert indeksering, der samme dokument vil bli tildelt de samme nøkkelordene hver gang, uansett når og hvor indeksering finner sted.

Ved automatisert indeksering av alle ord i dokumentet vil man se at ord kan forekomme i forskjellige former, f.eks. 'lov', 'loven', 'lover'. Disse handler om det samme og man kan derfor ønske at søk etter ett av ordene også vil gi treff på dokumenter der ordet forekommer i andre former.

¹SIFT står for *Søking i FriTekst*.

²Tesaurus betyr skattkammer på gresk, men her er det snakk om en kontrollert vokabular der termene er klassifisert i forhold til hverandre. En tesaurus er altså mer enn bare en liste over tillatte nøkkelord.

Man kan oppnå dette ved å reduserte ordet til dets basisform og bruke denne i indekseringen. Denne reduksjonen av ordet til dets basisform kalles stemming.

2.2 Utforming av søkeargument

Som tidligere nevnt, eksisterer en brukers informasjonsbehov bare i brukerens hode, det er bare brukeren som kan vite hva han vil, og for at informasjonssystemet skal gi brukeren dokumenter som dekker behovet (relevante dokumenter), er det en nødvendig forutsetning at brukeren klarer å uttrykke sitt informasjonsbehov på en måte som informasjonssystemet kan behandle.

Når en bokleser benytter registre, må spørsmålet uttrykkes på en måte som er egnet for oppslag i registeret, og leseren velger et emne han så prøver å finne i det systematiske registeret.

Jeg kommer til å bruke ordet "søkeargument" om resultatet av formuleringen av søkeforespørselen. Søkeargument kan bestå av ett eller flere *søkeord* som eventuelt kombineres ved hjelp av *søkeoperatorer*. Informasjonssystemet må inneholde utvalgsmekanismer til å plukke ut dokumenter hvis informasjonsinnhold er i samsvar med det brukeren har bedt om ved hjelp av søkeargumentet.

Indeksering er en prosess som ikke påvirkes av brukeren. Utforming av et søkeargument er motsatt en prosess som ikke kan påvirkes av informasjonssystemet, informasjonssystemet kan bare gi en viss veiledning til brukeren om regler for hvordan søkeargumentet kan bygges.

Som nevnt innledningsvis, vil informasjonssystemets oppbygning legge føringer på hvordan søkeargumentet kan utformes. Ofte benyttes det et formalisert spørrespråk der søkeord kan kombineres ved hjelp av operatører *AND*, *OR* og *NOT* som utfører snitt-, union- og negasjonoperasjoner på dokumentmengder. Spørrespråket kan også ha nærhetsoperatører som lar en bruker angi at søkeordene må forekomme i samme setning eller sam-

me avsnitt. For dokumenter som inneholder strukturerte data³, kan søket begrenses bare til noen av feltene.

Det er helt vanlig at en søkeprosess forgår i flere trinn. Brukeren lager et søkeargument, lar informasjonssystemet plukke ut dokumenter som stemmer overens med søkeargumentet og så avhengig av søkeresultatet forandrer brukeren på søkeargumentet ved å tilføye, fjerne eller endre på søkeordene. Søking er på denne måten en iterativ prosess.

2.3 Siling

Siling er mitt forslag til det engelske *matching* og betegner prosessen som finner sted når informasjonssystemet foretar et utvalg av dokumenter i informasjonssystemet basert på kriteriene i søkeargumentet. Det er verdt å merke seg at siling skjer som oftest ikke direkte på dokumentene i dokument-samlingen, men på registrene som ble etablert i indekseringsprosessen, f.eks. inverterte filer. Vi ser at silingsmekanismen jobber ikke direkte med dokumenter og brukerens spørsmål, men med en tilnærmet formell representasjon av dokumentenes innhold i form av indekseringstermer i registre og en formell representasjon av brukerens spørsmål som tar form av et søkeargument.

Siling kan utføres på flere måter, her vil jeg gi en kort beskrivelse av de to vanligste silingsmetodene: boolsk og vektorbasert sammenlikning.

2.3.1 Boolsk sammenlikning

Et boolsk søk er et "enten eller" søk med to mulige utfall, og ved boolsk sammenlikning skilles dokumenter i systemet i to grupper: kanskje relevante og kanskje ikke relevante.

³Strukturerte data har spesielle merkelapper på alle dataelementer, slik at disse er identifiserbare for datamaskiner og på brukerens forespørsel kan systemet gi brukeren innholdet av dataelementet med en gitt merkelapp.

Når det bare er to utfall i boolsk sammenlikning, betyr at det ikke er mulig at et dokument er mer relevant enn et annet, og det er heller ikke mulig å rangere disse etter relevans ved visning av søkeresultatet. I praksis tilbyr søkesystemer som baserer seg på boolsk sammenlikning rangering av søkeresultatene, rangeringen kan f.eks. være bygget på hvor mange av søkeord finnes i dokumentet, men også mer avanserte rangeringsalgoritmene benyttes.

2.3.2 Vektorbasert sammenlikning

Vektorbasert sammenlikning er mer fleksibel enn boolsk sammenlikning bl.a. ved at metoden tillater bruk av dokumenter som et søkeargument og dermed søk etter “liknende” dokumenter⁴.

Ved vektorbasert sammenlikning blir både søkeargument og de tilgjengelige dokumentene representert i form av vektorer. De relevante dokumentene vil ha vektorer som ligger i nærheten av eller peker i samme retning som vektoren til søkeargumentet.

Sammenlikningen foregår i et rom som har like mange dimensjoner som det er indekseringstermer, gjerne mange tusen. Det er ikke mulig å forestille seg et rom med 50.000 dimensjoner, men det er ikke vanskelig å behandle slike rom matematisk.

Søkeordene som dimensjonene i vektorrommet representerer behøver ikke å være “ord” slik vi vanligvis forstår ordet, men kan f.eks. være henvisninger. Det har vært gjort forsøk på å ordne liknende dokumenter automatisk i grupper ved å analysere vektorene til henvisningene i dokumentene. Forsøket er interessant og resultatet er publisert i [17].

⁴Systemer som baserer seg på boolsk sammenlikning kan i praksis også tilby søk etter “liknende” dokumenter, men da vil rangeringsalgoritmen være dominerende, ikke den boolske søkealgoritmen.

2.4 Kriterier for søkeeffektivitet

Formålet med et søk er å lokalisere dokumenter hvis meningsinnhold helt eller delvis dekker brukerens informasjonsbehov, altså relevante dokumenter. I forhold til juristens oppgaver er relevansbegrepet forklart i avsnitt 1.3 på side 6.

Hvorvidt et dokument er relevant eller ikke avhenger av person og oppgaven vedkommende står overfor. En dokumentetsamling som søket skal foregå i vil inneholde dokumenter som er relevante og dokumenter som ikke er relevante, fordelingen mellom disse kategoriene vil stadig variere. Brukeren vil at søkeresultatet skal gi ham flest mulig relevante dokumenter og færrest mulig ikke relevante. *Gjenfinningsgrad* og *presisjon* kan brukes som kriterier på hvor godt et søkesystem klarer å nå opp til målene.

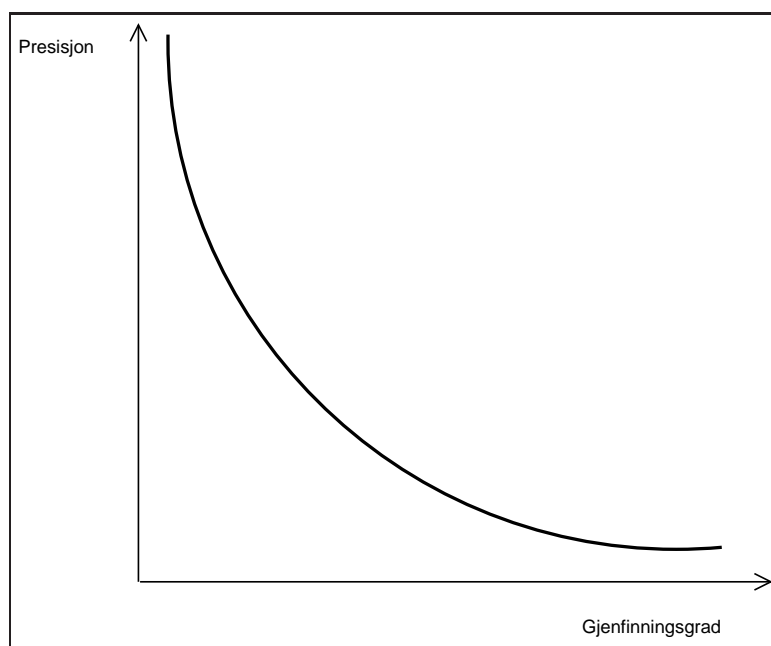
Hvis vi for et gitt søketilfelle betegner alle relevante dokumenter i et informasjonssystem med R og de dokumentene som systemet anser som relevante som F , vil presisjon og gjenfinningsgrad bli definert på følgende måte:

$$\text{Presisjon} = \frac{|R \cap F|}{|R|}$$

$$\text{Gjenfinningsgrad} = \frac{|R \cap F|}{|F|}$$

Presisjon forteller hvor mye støy av ikke relevante dokumenter søkeresultatet inneholder og er et uttrykk for hvor stor er sannsynligheten for at et funnet dokument er relevant, mens gjenfinningsgrad gir et uttrykk for mengden av relevante dokumenter som systemet ikke klarte å lokalisere og kan ses på som sannsynligheten for at et relevant dokument blir funnet.

Ved evaluering av presisjon og gjenfinningsgrad er man avhengig av



Figur 2.1: Sammenhengen mellom et søkesystems presisjon og gjenfinningsgrad.

å vite hvilke dokumenter som er relevante. Relevans er et begrep som er person- og situasjonsavhengig og ved evaluering går noen eksperter gjennom dokumentutvalget og velger ut de som etter deres mening er relevante ved løsning av et konkret spørsmål.

Presisjon og gjenfinningsgrad brukes ved måling av gjenfinningskvalitet. Ved mer generell evaluering av søkesystemene kan man også benytte andre kriterier: støymengde (altså dokumenter som ikke er relevante), dekningsgrad, brukerinnsats, svartid og presentasjon av søkeresultatet.

Det er en sammenheng mellom presisjon og gjenfinningsgrad og når man klarer å oppnå høyere presisjon vil gjenfinningsgraden synke, se til illustrasjon figur 2.1.

2.5 Fritekst-søk

Den mest brukte teknikken for å finne frem i digitale tekster er i dag *fritekst-søk* i fulltekst. Fritekst-søk dreier seg om å finne dokumentene som omtaler det brukeren søker etter. Så lenge datamaskiner ikke kan forstå tekst, må man nøye seg med en analyse av teksten. Men det skjer ikke en eksakt mønstergjenkjenning, det skjer en mer tilnærmet sammenlikning der det tas hensyn til språkets syntaktiske egenskaper.

Ethvert søk har den svakhet at søkeresultatet alltid vil ha irrelevante treff og noen relevante dokumenter ikke vil bli funnet i det hele tatt. Målet for et søkesystem er å øke både presisjon og gjenfinningsgrad, men resultatet vil aldri bli 100% presisjon og 100% gjenfinningsgrad så lenge datamaskiner ikke kan "forstå" tekst. Ved en evaluering av et fulltekst-søkesystem (IBMs STAIRS) som inneholdt 350.000 sider med tekst, ble konklusjonen at bare 20% av relevante dokumenter ble funnet, mens brukere trodde at gjenfinningsgraden var på over 75%.^[5]

I menneskelige språk vil meningen til et ord avhenge av konteksten. Homonymi (dvs. at et ord kan bety forskjellige ting) og synonymi (at forskjellige ord kan bety liknende ting) er utfordringer som datamaskiner ikke klarer å løse alene. Et eksempel fra avtaleretten kan tjene som en illustrasjon. Avtalelovens fulle navn er "Lov om avslutning av avtaler, om fullmakt og om ugyldige viljeserklæringer". I § 1 benyttes ordet "rettshandel" som et synonym til ordet "viljeserklæring", og i tillegg har man "dispositivt utsagn" som en slags fellesbetegnelse på private partsytringer som får rettsvirkninger etter sitt innhold⁵.

Homonymer reduserer et søkesystems presisjon, synonymer reduserer gjenfinningsgraden.

⁵Uttrykket "dispositivt utsagn" nevnes ikke i avtaleloven, men ser ut til å være mye brukt i juridisk teori og rettspraksis, og er bl.a. brukt i artikkel 12 i den norske oversettelsen av FN-konvensjonen om kontrakter for internasjonale løserekjøp.

Kapittel 3

Rettskildefaktorer i rettslige informasjonssystemer

Rettskildefaktorer produseres kontinuerlig og etter hvert blir samlet rettslige informasjonssystemer.

Bak rettslige informasjonssystemer (tradisjonelle papirbaserte så vel som digitale) står det en redaksjon som bestemmer hvilke rettskildefaktorer skal være med i systemet og som dessuten bearbeider rettskildefaktoren for å gjøre den best mulig egnet for utnyttelse i informasjonssystemet.[3, s. 41]. Rettstidende inneholder ikke alle dommer som avsies i Norge, men bare de dommer som avsies av Høyesterett. Ved publisering utstyres avgjørelsene dessuten med et sammendrag som svært kort angir sakens faktiske bakgrunn, rettslige høydepunkter og resultatet. Norges Lover (som gis ut av Stiftelsen Lovdata) og Gyldendals Norsk Lovkommentar inneholder begge gjeldende lover, men Norsk Lovkommentar har i tillegg kommentarene til lovene og de enkelte lovbestemmelsene.

3.1 Fra tekst til tall – et glimt fra virkeligheten

Det er en fordel for leseren at forfatter ved abstrakte drøftelser klarer å relatere disse til noe leseren vet fra før. Jeg skal derfor så vidt det lar seg gjøre prøve å knytte det jeg skriver om til et system som de fleste har kjennskap til og en viss erfaring med.

Jeg vil bruke Lovdatas WWW-baserte rettslige informasjonssystem¹ som et eksempel på hvordan rettskildesøk skjer i dag. Et annet system kunne tjent som illustrasjon, men for norske lesere vil Lovdata være et kjent eksempel, men også et eksempel på et system der grensesnittet roper om forbedringer.

3.2 En rettskildefaktor blir et dokument

I dette avsnittet vil jeg kort skissere prosessen som rettskildefaktorene går gjennom før de gjøres tilgjengelige i Lovdatas system. Morten Daae har skrevet ganske grundig om Lovdatas historie og system i sin avhandling[8] og særlig interesserte henvises til denne for en mer utførlig presentasjon av Lovdata og Lovdatas system.

Jeg vil illustrere prosessen med en tilfeldig kjennelse fra en tilfeldig lagmannsrett.

Daglig får Lovdata tilsendt nye avgjørelser fra ting- og lagmannsrettene fra Domstolsadministrasjonen via LOVISA, domstolenes elektroniske saksbehandlingssystem. Avgjørelsene fra Høyesterett kommer også inn daglig, men via et annet system². Fra LOVISA kommer det to filer for hver avgjørelse. Den ene er en XML-fil³ med strukturerte opplysninger om partene, advokater, dommere og et sammendrag av avgjørelsen. Den andre

¹<http://websir.lovdato.no/>

²Lovdata har siden 1987 mottatt avgjørelser fortløpende direkte fra Høyesterett.

³XML står for *Extensible Markup Language* og er en standard for å strukturere eller beskrive data i henhold til en egendefinert syntaks. XML brukes ofte for å utveksle data mellom forskjellige datasystemer.

filen er i "vanlig" Microsoft Word-format og inneholder avgjørelsen i fulltekst.

XML-filen inneholder *meta-data* i et strukturert format som er velegnet for maskinell behandling. Meta-data defineres ofte som "data om data". En slik definisjon sier ikke mye, og en mer funksjonell beskrivelse er at meta-data er strukturerte data med informasjon om innholdet i et dokument, f.eks. forfatter, tittel, dato, sammendrag. Disse opplysningene er representert på en måte som gjør at en datamaskin kan behandle disse som opplysninger om dokumentet. En datamaskin vil ikke forstå hva "forfatter" betyr, men dersom dokumenter er utstyrt med et meta-data-felt "forfatter", kan datamaskinen gi brukeren innholdet i feltet, og det blir så opp til brukeren å finne ut hva navnet på feltet betyr og hvordan innholdet i feltet skal forstås. Hvordan XML-filen som Lovdata mottar ser ut kan man se i vedlegg A på side 75.

Den andre filen er et dokument i Microsoft Word-format. Den kan leses på skjerm og skrives ut, men formatet er ikke godt egnet for automatisk behandling. Derfor konverteres filen til HTML⁴.

Lovdata konverterer Microsoft Word-filen til HTML-format, som sammen med XML-filen blir brukt til å opprette en fil i et spesielt Lovdata-format kalt Bokus⁵. Konverteringsprosessen fra XML og HTML til Bokus skjer automatisk, men resultatet blir kontrollert, og det skjer en manuell gjennomgang av teksten bl.a. for å sikre at formateringen er i orden og i overensstemmelse med innkommet original. Formatet Bokus likner mye på HTML i den forstand at det primært spesifiserer hvordan den oppmerkede teksten skal se ut: det er egne koder for avsnitt, kursiv, fet skrift, anførselstegn osv. I alt har Bokus seks typer koder[8, s. 37]:

1. Faste felter i begynnelsen av dokumentet⁶

⁴HTML står for *HyperText Markup Language* og er et markeringsspråk som brukes for å lage WWW-sider.

⁵Bokus er et format utviklet av Lovdata med Lovdatas egne koder for strukturering av dokumenter.

⁶Feltene brukes for lagring av meta-data.

2. Overskrifter/paragraf
3. Marg, innrykk og luft i brødtekst
4. Skrifttype i brødtekst
5. Tabeller
6. Spesialtegn

Så langt har det ikke skjedd noe med innholdet i avgjørelsen. Enkelte saker må anonymiseres, både når det gjelder navn og steder. Lovdata har automatiserte rutiner for gjenkjenning og anonymisering av person- og stedsnavn, men resultatet av automatisert anonymisering blir grundig kontrollert manuelt. Ved anonymisering blir personnavn erstattet med bokstaver i begynnelsen av alfabetet og andre navn blir erstattet med bokstaver som kommer mot slutten av alfabetet.

Avgjørelsene inneholder henvisninger bl.a. til lover, forskrifter og andre avgjørelser. Neste steg i bearbeidelsesprosessen er standardisering av lovhenvvisninger. Det kjøres et spesielt program som leser teksten og finner de delene av teksten som ser ut som lovhenvvisninger — disse tekstelementene vises på skjermen, henvisningene utheves og det gis forslag til standardhenvvisninger. Står det “kjøpsloven” i teksten, vil systemet spørre om det er snakk om kjøpsloven av 1988 eller den gamle kjøpsloven av 1907.

Lover, forskrifter, artikler og andre dokumenter går gjennom en liknende behandling. Juridiske artikler krever spesielt mye manuelt arbeid. Ofte foreligger artikkelteksten bare i trykt form og må skannes inn for at artikkelen skal kunne legges i Lovdatas elektroniske artikkeldatabase. Men det er spesielt henvisningene som krever mye innsats. Henvisningene til Høyesterettsdommer byr sjeldent på problemer, de begynner med “Rt” eller “Rettstidende” etterfulgt av årstall og sidetall. Artikkelforfattere nøyer seg ikke med norske avgjørelser, de elsker å vise til andre artikler i diverse nasjonale og utenlandske tidsskrifter, utenlandsk rettspraksis og dessuten

utarbeider egne måter å forkorte lovtitler på (f.eks. kan "s" være brukt som en forkortelse for straffeloven og "dl" brukes som forkortelse for så vel dekningsloven som for domstolloven). I USA har man to publikasjoner, *The Bluebook – A Uniform System of Citation* og *The ALWD Citation Manual*, som begge inneholder retningslinjer for hvordan henvisningene i juridiske tekster bør utformes.

Det gjøres ikke noe mer med innholdet i dokumenter av Lovdatas redaksjon. Andre tilbydere av rettslige informasjonssystemer legger ofte en større mengde arbeid i redaksjonen av dokumenter, f.eks. har Gyldendal Rettsdatas Norsk Lovkommentar kommentarer og bemerkninger til de enkelte lover og bestemmelser.

Deretter skjer det en indeksering av dokumentet med tilhørende nødvendige oppdateringer av relevante databaser. En av databasene som oppdateres er en såkalt skyggebase. Skyggebasen inneholder en oversikt over henvisningene mellom dokumentene i de forskjellige basene. Ved å gjøre et oppslag i skyggebasen kan man finne ut hvilke dokumenter viser til dokumentet som brukes som søkeargument.

I Lovdatas system er avgjørelsen til slutt representert som et dokument inntatt i vedlegg B på side 78 (bare et utdrag er tatt med). Dokumentet inneholder avgjørelsens meta-data (type avgjørelse, instans, publiseringsdato, saksgang, partene, stikkord osv.), sammendraget og selve teksten. Meta-data ligger i dokumentets "hode" (feltene <h0> til <h99>).

3.3 Søking

Når en jurist skal søke etter relevante rettskildefaktorer i Lovdata, må han først velge en base. Lovdata har noen titalls baser med lover, forskrifter, avgjørelser, forarbeider m.m. og normalt vil brukere være interessert i dokumenter bare innenfor et utvalg av disse basene.

Velger man rettslige avgjørelser, vil man deretter få et WWW-skjema med felter for emnesøk i fulltekst, lovhenvvisninger m.m. Det er bare em-

nefeltene som i praksis blir benyttet i søk der man leter etter dokumenter om et bestemt emne. Det er 3 emnefelt (Emne 1, Emne 2, Emne 3) og i hver av disse skriver man inn stikkord som beskriver ett emne, man formulerer et søkeargument. Når man trykker på SØK vil systemet finne dokumenter som oppfyller kriteriene oppgitt i alle de utfylte feltene. Søkesystemet benytter av registrene som opprettes ved indeksering når dokumentene blir lagt inn i systemet.

Kjernen i søket etter dokumenter om et emne og en forutsetning for et godt søkeresultat er at brukeren selv finner ut hvilke ord som best beskriver emnet og derfor sannsynligvis vil forekomme i relevante dokumenter. Praktisk erfaring fra fagfeltet og resultater fra tidligere søk vil hjelpe brukeren i å bestemme hvilke søkeord som han kan bruke.

3.4 Bristende forutsetninger

For en jurist vil målet normalt være å finne flest mulig relevante rettskilder om et gitt emne. Teknikken som er beskrevet ovenfor med søk i dokumentsamlingen ved et brukerformulert søkeargument forutsetter at

- juristen klarer å identifisere søkeord som beskriver emnet han interesserer seg for og
- søkeord juristen benytter i sitt søkeargument faktisk forekommer i dokumentet.

Oppfylles ikke forutsetningene, vil noen av de relevante dokumentene ikke bli funnet. Forutsetninger blir aldri helt oppfylt for alle relevante dokumenter, jf. omtalen av IBM STAIRS-evalueringen i avsnitt 2.5 på side 19.

3.5 Rettskildefaktorer danner strukturer

Et rettsystem utgjør en helhet, men det består ikke av én enhet. Rettsystemet er som et lappeteppe eller et puslespill, og er bygget opp av forskjellige fragmenter, Eckhoff snakker f.eks. om lovbestemmelsenes fragmentariske karakter.[9, kap. 2 pkt. IV] Lover har en kompleks struktur og lovgiver bruker forskjellige teknikker for å gi en lov en hensiktsmessig lineær utforming.[10] Henvisninger på kryss og tvers av lovverket er en av slike teknikker.

Hvis man i tillegg til lovene trekker i bildet forarbeider, rettslige avgjørelser og juridisk teori og ser på relasjonene som eksisterer mellom rettskildene i disse forskjellige kategoriene vil en svært komplisert struktur komme frem, en *hyperstruktur*, der rettskildene kobles sammen i mangedimensjonelle datastrukturer⁷.

3.5.1 Lover som hypertekst

Lover er en type tekst som bare unntaksvis leses sekvensielt; det normale er derimot at man ikke bare hopper mellom forskjellige paragrafer, men også mellom lover. Slik lovene er bygget opp, er det mulig å forta entydige henvisninger til lovbestemmelsene med paragraf, ledd, litra, punktum. Slike henvisningene har en fantastisk egenskap — de er mediauavhengige!

Datamaskinbaserte systemer som tillater at teksten organiseres på flere måter og som tillater å “hoppe” i teksten har fått betegnelsen *hypertekst*. Begrepet ble i sin tid introdusert av Ted Nelsom om “*non-sequential writing*”⁸. Man kan også si at hypertekst er tekst som ikke er bundet av en lineær struktur eller hierarkiske menyer, tekst som kan ha koblinger til andre tekster⁹.

⁷Som en liten digresjon kan jeg nevne at en hyperstruktur i matematikken er et algebraisk struktur som støtter hyperoperasjoner og behandles i abstrakt algebra.

⁸<http://www.w3.org/Xanadu.html>

⁹Se definisjon av *World Wide Web* i Vedlegg 1 til St.meld. nr. 41 (1998-99) (Om elektronisk handel og forretningsdrift).

Slik rettskildefaktorer er definert ovenfor i avsnitt 1.4 på side 7, vil de alltid bestå av tekst. Kan man se på rettskildefaktorene som en form for hypertekst? En hypertekst forutsetter at det eksisterer uttrykkelige koblinger mellom tekster som lar seg følge. For å betrakte rettskildefaktorene som hypertekst, må man altså identifisere slike koblinger, disse må komme til uttrykk på et eller annet vis for at hypertekstsystemet skal kunne følge disse koblingene.

3.6 Henvisninger

Jeg vil begynne dette avsnittet med en sitat fra en artikkel om en fotnote, *The Footnote*[2]:

Here indeed is a footnote that has become more important than the text; that is often read separated from its text; that can stand alone. Nor is this footnote a trifle, or an insignificant bauble. It has inspired countless books and law review articles.

Fotnoter, referanser og sitater er alle sammen former av henvisninger. Bruk av henvisninger til andre verk har lange tradisjoner, i vitenskapelig litteratur er henvisninger sementen som lar en forsker bygge sitt arbeid på vitenskapens grunnmur. I juridisk teori finner man alle tre nevnte former for henvisninger, mens i norske rettslige avgjørelser ser man bare referanser og sitater. Lenker på WWW-sider er også henvisninger og de skiller seg ikke mye fra tradisjonelle henvisninger i bøker og artikler.

Tradisjonelle henvisninger har egenskaper som det kan være nyttig å merke seg.

For det første sier en henvisning i *seg selv* ikke noe ut over at det eksisterer en viss relasjon mellom innholdet i de to tekstene som knyttes sammen med henvisningen. Henvisningen sier ikke hva slags relasjon det er: en henvisning i en juridisk artikkel kan være en kritisk bemerkning til en tidligere artikkel, en henvisning mellom to lovbestemmelser kan

etablere et unntak fra en hovedregel, og Høyesterett kan velge å fravike en rettsoppfatning en tidligere dom bygger på og vise til den tidligere dommen. Hva slags relasjon det er snakk om fremgår ofte, men ikke alltid, av konteksten. Konteksten kan gi leseren en viss indikasjon på typen av forholdet som forfatteren mener eksisterer mellom innholdet i to tekster. Videre kan forfatteren bruke spesielle markører (typisk: "jf", "sml", "motsatt" o.l.) for å markere relasjonen, men også disse markørene er en del av konteksten.

Enkelte juridiske studenter har utviklet kompliserte måter for å innarbeide hjelpemidler til eksamen. Henvisningene spiller en sentral rolle i dette arbeidet, og det benyttes ofte forskjellige farger for å markere forskjellige typer relasjoner.

Med overgang til digitale informasjonssystemer er mulighetene åpne for utvikling av systemer der henvisninger får selvstendige egenskaper, f.eks. *type*, og at man kan si noe om forholdet mellom innholdet i to tekster bare ved å se på egenskapene til henvisningen. På denne måten kan man automatisk velge ut dokumenter som på en spesiell måte relaterer seg til et gitt dokument. For å illustrere hva jeg mener, gir jeg et lite utdrag over standardtermer som brukes i *Australian Case Citator*[7, s. 98-99] for oppgi hvordan en senere dom forholder seg til en tidligere avgjørelse:

Affirmed Used where the decision in the annotated case has been upheld on appeal

Applied Used where the principles of law enunciated in the annotated case have been applied to a new set of facts and circumstances

Approved Used where principles of law enunciated in the annotated case have, in the annotating case, been held to be good law

Reversed Used where the decision in the annotated case has been upset by a court on appeal

Listen over termer som brukes er lengre og inneholder i tillegg følgende: *Considered, Disapproved, Distinguished, Explained, Followed, Not*

Followed, Overruled, Referred to. På samme måte som “jf” og “sml” er også disse markørene bare en del av konteksten, men ved konsistent og utstrakt bruk vil slike markører være bundet så sterkt til henvisningen at henvisningen kan sies å få en type.

På WWW-sider lages det lenker til andre sider ved at noen ord gjøres klikkbare. Dersom disse ordene er valgt med omhu, kan man allerede ved å se på ordene avgjøre om det er interessant å følge lenken eller ikke. Klikkbare ord er også her en del av konteksten, og lenkens eneste funksjon er å bringe leseren frem til et nytt dokument. Det er dermed ingen funksjonell forskjell mellom dagens WWW-hyperlenker og ikke klikkbare henvisninger, forskjellen ligger bare i hvor mye tid det tar å følge henvisningen.

Henvisninger i bøker og andre trykte skrifter lar seg følge i én retning. Dette betyr at ved henvisning fra *dokument A* til *dokument B*, etableres det ikke noe referanse tilbake fra *dokument B* til *dokument A*. Noen vil sikkert spørre om det er en så stor begrensning. Jeg påstår at en slik egenskap ved henvisninger er en stor praktisk ulempe, og jeg er selvsagt ikke den som oppdaget dette først. I amerikansk rett spiller prejudikater en meget stor rolle, mye større enn i Norge, og der har det lenge vært utgitt en egen publikasjon med oversikt ikke bare over hvilke tidligere dommer en senere dom bygger på (eller fraviker), men også hvilke senere dommer som bygger på (eller fraviker) den tidligere dommen. Alt i 1873 ga Frank Shepard sin *Citation Index* som umiddelbart ble en stor suksess. Før Shepard’s måtte advokater selv vedlikeholde oversikter over senere dommer som omtalte tidligere. Nye avgjørelser treffes hele tiden, og det er derfor viktig at nye utgaver av slike publikasjoner kommer ut hyppig. Å vedlikeholde en publikasjon som Shepard’s er arbeidskrevende.

Et interessant eksempel fra Norge er *Lov og Dom*. Denne gir en oversikt over rettsavgjørelser med utgangspunkt i de enkelte lover og deres enkelte paragrafer. Per mars 1992 har *Lov og Dom* vært utgitt som tre samlinger: *Lov og Dom 1836–1968*, *Lov og Dom 1969–1988* og *Lov og Dom Rettens Gang 1933–1989*. Ved ett oppslag i hver av disse samlingene kan man finne

henvisninger til samtlige dommer fra de nevnte periodene som viser til den aktuelle lovparagraf.

Både *Shepard's* og *Lov og Dom* er hjelpemidler for å omgå henvisningenes relasjonelle enveisnatur.

Henvisningene har vært brukt som en indikator på et verks gjennomslagskraft og innflytelse. *Science Citation Index*¹⁰ bygger på tanken om at jo viktigere en artikkel er, desto flere henvisninger fra andre artikler vil den ha.

18. januar 2006 lanserte Lovdata en ny funksjon ved søk i rettsavgjørelser — rangering av søkeresultater etter popularitet¹¹:

Ved søk i rettsavgjørelser er det nå mulig å rangere dokumentene etter hvor mange andre dokumenter i systemet som omtaler dem. De mest omtalte og dermed gjerne de viktigste avgjørelsene blir liggende øverst i resultatlisten — lett tilgjengelige for brukerne.

Lovdatas ide er ikke ny, søkemotoren Google har hatt en teknologi som heter *PageRank* som også bygger på at antall henvisninger vil være økende med viktigheten til en side.[6]

Det er gjennom henvisninger at sammenhenger mellom tekster kommer til uttrykk og henvisninger får stadig større oppmerksomhet ettersom mengden av tekstmateriale øker. Jeg vil se på henvisningene som koblinger som knytter forskjellige rettskildedefaktorer sammen til hyperstrukturer. Dette gjør at rettskildedefaktorene kan betraktes som en form for hypertekst.

¹⁰<http://scientific.thomson.com/products/sci/>

¹¹<http://www.lovdata.no/nyhet/2006/20060118-1603.html>

Kapittel 4

Navigasjonsverktøy

Fulltekst-søk i fritekst har sine begrensninger, jf. avsnitt 2.5 på side 19, og jeg ønsker å se nærmere på søkestrategier som kan fungere som et supplement eller til og med som et alternativ til fritekst-søk.

Henvisningene etablerer som allerede nevnt, koblinger mellom dokumenter. Disse koblingene sier ikke mye, men de sier at det iallfall eksisterer en viss relasjon mellom innholdet dokumentene. Henvisningene gjøres av forfattere, dermed fanges opp også relasjoner som ikke kommer direkte til uttrykk i teksten, men følger av konteksten. En mulig søkestrategi kan derfor være å ta utgangspunkt i en relevant rettskildefaktor (f.eks. en dom eller en lovparagraf) som man på et eller annet vis har funnet frem til, og følge stiene som henvisningene til og fra rettskildefaktoren skaper. Lov og Dom er en henvisningsregister som lar jurister søke slik: man tar utgangspunkt i en lovbestemmelse og får en oversikt over dommer som viser til bestemmelsen. Noe publikasjon tilsvarende Lov og Dom for å finne dommer med utgangspunkt i andre dommer (altså norsk *Shephard's*) har man ikke. Det er interessant å vite hvilke senere dokumenter som viser til det foreliggende. Lovdatas system gir brukere en viss mulighet for å søke etter senere dokumenter som viser til det foreliggende vha. henvisninger: her og der vil det dukke opp knapper der det med rødt står *Henvisninger, Avgjørelser, Forbeider* eller *Artikler*. Ved å klikke på disse knappene vil man

få en oversikt over dokumenter som viser til det foreliggende dokumentet eller paragrafen. En henvisning mellom to dokumenter er en indikasjon på at det eksisterer en viss sammenheng mellom innholdet i disse dokumentene. Brukeren må selv finne ut om relasjonen er relevant for ham i hans situasjon.

Juristen finner altså noe som er relevant for hans situasjon f.eks. ved hjelp av fritekst-søk eller ved å ha en direkte referanse til et relevant dokument, og deretter benytter henvisningsveven som dokumentet inngår i til å navigere seg gjennom dokumentstrukturen for å sjekke om det er eventuelle andre dokumenter som er relevante. Lovdatas eksisterende knappe- og lenkeløsning egner seg dårlig for en slik søkestrategi, grensesnittet må tilpasses strategien. Derfor heter undertittelen på oppgaven *Om alternative grensesnitt til rettslige informasjonssystemer*. Jeg ønsker å se på muligheter for å visualisere henvisningsstrukturen og gjøre den navigerbar, slik at det kan lages et interaktivt verktøy der brukeren og systemet spiller sammen, et verktøy som kan kartlegge og visualisere rettslige hyperstrukturer.

I min avhandling vil jeg ta utgangspunkt i Lovdatas system, dokumentdatabaser og eksisterende grensesnitt og analysere muligheter for å forbedre søkemuligheten ved å at henvisningene bedre lar seg utnytte.

Det er ingen tvil om at det er store muligheter til å forbedre grensesnittet til henvisningsnavigasjon. For å finne ut om dokumentene som viser til et relevant dokument også er relevante må brukeren åpne hvert dokument, finne ut hva den dreier seg om ved å lese større eller mindre deler av dokumentet, lokalisere henvisningen som førte brukeren til dokumentet og finne ut i hvilken kontekst henvisningen forekommer i. På denne måten er det enkelt å rote seg helt bort fra det opprinnelige dokumentet og ikke finne veien tilbake. Brukeren av systemet må i tillegg til sitt juridiske spørsmål alltid tenke på hvor i henvisningsstrukturen han befinner seg, hvor han har vært, og hvor han bør komme innom. En grafisk visualisering som et nettverk der dokumenter er representert som noder som bindes sammen ved henvisninger, vil kunne danne en

slags kart over strukturen og dermed gjøre navigasjonen mulig og sørge for at brukeren vet hvor han er, hvor han har vært, og hvor han skal. Forskjellige typer dokumenter kan merkes på forskjellige måter i denne grafiske representasjonen og noder i grafen kan tenkes å ha forskjellige attributter som brukeren kan velge å se på uten å miste oversikten over hvor han er og hvor han skal. Et slikt navigasjonsverktøy kan *kanskje* bedre informasjonssystemets gjenfinningsgrad.

Kapittel 5

Format, metode, teknologi

Da jeg begynte å skrive var målet å se på eventuelle forbedringsmulighetene som visualisering av *henvisningsstrukturene*¹ som eksisterer i rettskildemateriale kunne gi. Visualiseringen skulle skje automatisk ved hjelp av egnet programvare og gi brukeren et kart over dokumentstrukturen, slik at skjulte sammenhenger mellom dokumenter kunne komme frem og bli oppdaget.

Det fantes selvsagt allerede interaktive verktøy med grafisk brukergrensesnitt² for navigasjon i nettverksstrukturer³. Noen disse verktøyene krevde at data som skulle visualiseres forelå i et spesielt proprietært format, mens andre bygde på åpne internasjonale standarder. Jeg ville unngå bruk av proprietære formater. Lovdata er organisert som en stiftelse for å være uavhengig; bruk av frie og åpne formater bidrar til at også den tekniske løsningen forblir uavhengig.

*GraphViz*⁴ er en programpakke for å visualisere grafer og ville for så vidt tjene visualiseringsformålet godt. Før man treffer en beslutning, er

¹Det var en enkel avgrensning gjort av hensyn til tid, for det eksisterer strukturer langs andre dimensjoner også.

²Egentlig en stor mengde slike verktøy, jf. oversikten på http://directory.google.com/Top/Science/Math/Combinatorics/Software/Graph_Drawing/.

³Et nettverk er et system av sammenkoblede komponenter, og rettslige hyperstrukturer kan dermed betraktes som en type nettverksstruktur.

⁴<http://www.graphviz.org/>

det viktig å undersøke nærmere om det kan være mer fordelaktig å velge en annen løsning. Jeg valgte derfor å tenke bredt og se på alternative tilnæringsmåter. Det var et godt trekk, for det viste seg at det fantes en teknologi som gjør er mulig å benytte samme datasett til mye mer enn å tegne grafer av sammenhenger i strukturene. Jeg tillater meg å bruke et par ord for å fortelle om søkeprosessen.

5.0.1 ZigZag og ZZStructure

I avsnitt 3.5.1 på side 26 vurderte jeg om lover kan betraktes som en hypertekst. Jeg startet min undersøkelse etter hvordan hyperstrukturene kunne representeres der, i ordet "hypertekst".

Ted Nelson fant på ordet "hypertekst" og skal ha mye av æren for at man blir mer og mer kritisk til at dagens elektroniske systemer i stor grad etterlikner tradisjonelle papirbaserte teknologier. Ted Nelsons visjon var en global dokumentsamling hvor millioner av forfattere kunne legge inn sine arbeider, og et system som kunne tillate sammenkoblinger av dokumenter på mange forskjellige måter. Han har jobbet med ideen over 40 år og et av resultatene av dette arbeidet er en datastruktur kalt ZigZag⁵ som tillater sammenkobling av elementene på vilkårlige måter i ubegrenset mange rom. ZigZag er generell og fleksibel, men ZigZag er fortsatt en prototyp og så vidt jeg vet ikke mye brukt i praksis, men benyttes stadig i forskjellige forskningsprosjekter.

5.0.2 RDF

En annet mulighet kunne være *Resource Description Framework* (RDF)⁶. RDF er en internasjonal standard til et rammeverk for å beskrive og utveksle meta-data, og den har solid fotfeste særlig i bibliotekverden. For data og relasjoner som er beskrevet ved hjelp av RDF finnes det grafiske

⁵<http://xanadu.com/zigzag/tutorial/ZZwelcome.html>

⁶<http://www.w3.org/RDF/>

visualiserings- og navigasjonsverktøy. Jeg sier mer om RDF i forbindelse med analyse av Emnekart.

5.0.3 Emnekart

Et tredje alternativ som jeg vurderte, var *Topic Maps*⁷ (“Emnekart” på norsk). Emnekart har på samme måte som RDF status som en internasjonal standard, og det finnes grafiske verktøy for å navigere i strukturene som er representert som et emnekart⁸.

Da jeg så på RDF og Emnekart, oppdaget jeg at begge to kunne brukes til mye mer enn å tjene som dataformat for visualisering av rettslige hyperstrukturer.

Både RDF og Emnekart er *datamodeller* for å beskrive innholdet og relasjoner i dataressurser. Emnekart og RDF lar en bruker kode hvordan data er ment å skulle forstås, og begge modellene gjør at datamaskiner kan “forstå” hva f.eks. et dokument snakker om og hvilken relasjon innholdet i dokumentet har til innholdet i andre dokumenter. Dessuten gjør modellene datamaskiner i stand til å si at to mengder med data har samme mening for et menneske, selv om datamengdene ikke er identiske. I avsnitt 3.6 på side 28 nevnte jeg at henvisningene kunne gis egenskaper som f.eks. type, både RDF og Emnekart tillater ikke bare beskrivelse av ting, men også relasjoner og disse kan gis selvstendige egenskaper, deriblant type.

Der er forskjeller mellom RDF og Emnekart, men det er også meget store likheter i anvendelsesområde: begge er laget for å beskrive informasjon om ting og relasjoner i en mengde dataressurser. RDF og Emnekart har begge status som internasjonale standarder og blir brukt i stadig større utstrekning.

⁷ISO/IEC 13250:2002

⁸Når jeg videre sier “Emnekart” med stor “E”, mener jeg datamodellen, når jeg sier “emnekart” med liten “e”, sikter jeg til dataformatet eller et konkret emnekart som er bygget på Emnekart-modellen.

5.1 Hvorfor Emnekart?

Når man har to tilsynelatende like gode løsninger, er det vanskelig å velge. Jeg valgte å bruke Emnekart som datamodell for mine undersøkelser. Valget har egentlig ikke vært vanskelig, for dersom det viser seg at RDF ville være et bedre alternativ, skal det være fullt mulig å gå over fra Emnekart til RDF[12] og det pågår arbeid for å bringe Emnekart og RDF nærmere hverandre⁹. I neste avsnitt presenterer jeg noen av grunner til at jeg valgte å se på Emnekart i denne omgang.

RDF bygger på formallogikk og er ment å være brukt av datamaskiner; Emnekart er ikke basert på formallogikk og er ment å være brukt av mennesker. Emnekart er mer menneskevennlig.

Flere offentlige og private prosjekter allerede nå har systemer som er drevet av emnekart: USIT ved Universitetet i Oslo, nettstedene *itu.no*, *forskning.no*, *kulturnett.no* og *udi.no* m.fl. I samarbeid med Statens forvaltningstjeneste har Finansdepartementet gjennomført et prøveprosjekt for å etablere en emnekartløsning for *ODIN*¹⁰. *Regjeringen.no* skal etter hvert erstatte nåværende *ODIN*, og Emnekart er en aktuell teknologi i prosjektet¹¹. Emnekart har vært foreslått til å være skjullet i et av forslagene til en prototyp på den nasjonale kunnskapsbasen¹².

Resultatet av prosjektet *Henrik Ibsens skrifter*¹³ skal etter hvert gjøres tilgjengelig på WWW. Emnekart-teknologi skal hjelpe brukere med å finne frem når nettstedet lanseres en gang i fremtiden, meningen er at brukere lettere skal finne relevant materiale, at systemet skal gi et overblikk over hvor brukeren befinner seg samt at materiale skal presenteres på en måte som vil gi brukeren motivasjon til å lete etter nytt stoff ved å utforske

⁹<http://www.w3.org/TR/2006/NOTE-rdf-tm-survey-20060210/>

¹⁰<http://odin.dep.no/>

¹¹Jf. programmet til Emnekart 2006-konferansen på <http://www.emnekart.no/2006/program/program.html>.

¹²Jf. Steve Peppers og Pål Steigans innspill på <http://odin.dep.no/odinarkiv/norsk/nhd/2002/enorge/024101-990100/dok-bn.html>.

¹³<http://www.ibsen.uio.no/his/>

relasjonene i materiale. *Henrik Ibsens skrifter* er et prosjekt som likner mye på det Lovdata gjorde i sin tid da norsk lovgivning skulle konverteres til digital form.

I Norge er Emnekart-miljøet meget kompetent og Emnekart får stadig større oppmerksomhet. Det offentlige var i denne omgang pioner da emnekart var tatt i bruk for å lette navigasjonen på diverse portaler. Private organisasjoner begynner å følge etter.

Når ingen sterke hensyn tilsier bruk av RDF eller en annen teknologi, og at Emnekart, etter hva jeg så langt kan se, egner seg svært godt for mitt prosjekt, velger jeg å bygge videre på Emnekart.

Jeg ønsker først å presentere Emnekart som metode og teknologi og deretter se på hva som må gjøres for å bruke emnekart som et hjelpemiddel til navigasjon i store dokumentsamlinger.

5.2 Kort om Emnekart

Ordet “emnekart” består av “emne” og “kart”. Disse to sier hva Emnekart dreier seg om – emnekart er et kart over emner. Disse emnene befinner seg i forskjellige ressurser og disse ressursene er typisk forskjellige dokumenter av ymse slag i en eller flere dokumentsamlinger.

Emnekart er en måte å beskrive emnene som er omtalt i dokumentsamlingen, relasjoner mellom emnene og referanser til hvor i dokumentsamlingen emnene er omtalt. Når jeg sier “beskriver”, mener jeg en formalbeskrivelse som også datamaskiner kan “forstå”, behandle og utveksle.

Emnekart er “bare” en modell. Hvordan man teknisk velger å implementere denne modellen er et annet spørsmål, og her har man flere alternativer: f.eks. HyTime¹⁴, XTM¹⁵, LTM¹⁶. Hvordan et emnekart er re-

¹⁴ISO/IEC 10744:1992

¹⁵XTM står for *XML Topic Maps* og er en XML-basert standard for å representere og utveksle emnekart. For detaljer om XTM se <http://www.topicmaps.org/xtm/1.0/>.

¹⁶LTM står for *Linear Topic Maps* og er et dataformat for å representere emnekart som

Abonnemet	39, 60, 123
Abstrakt mangelsbedømmelse	153–159
Action directe, se Direktekrav	
Actio Pauliana	825
Action quanti minoris	397
Addisjonsmetoden	239–241
Adekvans	kap. 20.3
– i og utenfor kontraktsforhold	528–529
– og casus mixtus	kap. 20.3.4
– som objektiv egenrisiko	kap. 20.6.5

Tabell 5.1: Utdrag fra et stikkordregister.

presentert i et system er uten betydning så lenge systemet lever et liv i isolasjon. Når systemet skal utveksle emnekart med andre systemer, må man bruke en representasjon som andre systemer vil forstå, f.eks. XTM. Informasjonsutveksling på tvers av systemer – ved at emnekartene lar seg flette sammen – er noe som er et av sentrale poengene med Emnekart, og standarden definerer to alternative måter for å representere og utveksle emnekart: HyTime og XTM.[1]

Det er lettere å lære noe nytt når det nye kan relateres til noe man allerede kjenner godt fra før. For å forstå hva Emnekart dreier seg om, kan man ta utgangspunkt i et konvensjonelt stikkordregister i en fagbok, se til illustrasjon et utdrag fra stikkordregisteret i Hagstrøms Obligasjonsrett i tabell 5.1.

Et stikkordregister består av en liste av emnenavn med tilhørende side- og kapittelhenvisninger der det står noe relevant om disse. Noen emner kan ha flere navn, f.eks. viser *Action directe* og *Direktekrav* til samme emne, men med forskjellige navn. Side- og kapiteltallene representerer *referanser* til *forekomster* av emnene i boken, dvs. hvor i boken emnene er omtalt.

Det er viktig å presisere at et emnekart er ikke en del av dokument-

er kompakt og lettleselig, for detaljer om LTM se <http://www.ontopia.net/download/1tm.html>. LTM er *ikke* et standard format for emnekart, men mange systemer støtter det likevel.

samlingen. Igjen kan parallellen trekkes til stikkordregisteret som ikke er en del av den teksten som registeret dekker.

Dersom det som har vært sagt så lang er forståelig, bør det ikke være en stor utfordring å forstå konseptet rundt Emnekart. Emnekart bygger videre på tanker fra et stikkordregister, men modellen er gjort generisk og noe utvidet, og egner seg like bra for tekst som for annet innhold. Grunnbegrepene i Emnekart er emner (*topics*), assosiasjoner (*associations*) og forekomster (*occurences*).[15] I tillegg har emnekart noe som stikkordregistre ikke har, nemlig perspektiv (*scope*) og temaidentitet (*subject identity*). Perspektiv er en mulig løsning på homonym-problemet og temaidentitet brukes enkelt sagt for å forsikre at to forfattere faktisk snakker om det samme når de bruker samme navn. Temaidentitet er en unik identifikator av et emne, temaidentitet gjør sammenfletting av emnekart mulig ved at forskjellige forfattere kan bruke samme temaidentitet når de snakker om samme tema i sine egne emnekart.

Jeg skal straks forklare hva de forskjellige elementene i et emnekart kan gjøre. For å gjøre det lettere å forstå begrepene, kan man si at emner og temaidentiteter svarer på spørsmål *hva* og *hvem*, assosiasjoner på spørsmål *hvordan* og *hvorfor* og forekomster på spørsmålet *hvor* (i ressurs-samlingen).

5.2.1 Emner

Emner er kjernen i Emnekart. Emner kan være hva som helst, alt kan defineres som et emne og dermed representeres i et emnekart. Emner kan f.eks. representere de ulike temaene som omtales i dokumentene, emner kan representere forfattere, steder, hendelser osv. Et tema er noe abstrakt, noe som kan omtales og forstås av et menneske. Et emne er et navngitt tema. Det abstrakte temaet blir konkretisert gjennom et emne og på denne måten blir det mulig å behandle temaet ved hjelp av emnekart. Det er viktig når man lager et emnekart at et emne representerer bare ett tema og

at et tema bare er representert ved ett emne¹⁷. Relasjonen mellom et emne og et tema tilsvarer den mellom et begrep og en ide hos Platon.

Emner i Emnekart kan ha egenskaper, de kan spille roller i assosiasjoner og de kan ha forekomster.[1, pkt. 3.23]

Et emne kan ha forskjellige navn, f.eks. på forskjellige språk eller i forskjellige miljøer. Det kan defineres forskjellige typer navn for forskjellige formål. Standarden nevner tre typer emnenavn, i tillegg til fullstendig navn, kan man definere navn som skal brukes i forbindelse med visning f.eks. på skjerm og navn som kan brukes ved sortering.

Emnenavn er gyldige innenfor et gitt *perspektiv*. Ved å knytte navn til perspektiv, kan datamaskiner skille mellom "bjørn" i perspektiv personnavn og "bjørn" i perspektiv dyr. Perspektiv løser dermed vanskeligheter forbundet med datamaskinell behandling av homonymi. Standarden tillater ikke at to emner i et emnekart har samme fullstendig navn innenfor samme perspektiv. Dersom det skjer, skal begge emnene anses for å representere samme tema og slås sammen.[1, pkt. 5.2.2]

5.2.2 Assosiasjoner

Vi mennesker tenker i assosiasjoner. Da jeg satt på bussen til byen i dag, så jeg en annen buss kjøre forbi med en reklame i bakvinduet. Det var reklame for et mobilselskap. Straks tenkte jeg på en venninne jeg ikke har sett på meget lenge. Hvorfor tenkte jeg på henne da jeg så reklamen? Hun jobber i selskapet som reklamen omtalte.

Temaer som omtales i dokumenter er i den virkelige verden ofte knyttet til andre temaer. I bøker og artikler brukes forskjellige former for henvisninger for å markere sammenhengene. I Emnekart uttrykkes relasjonene mellom emner ved hjelp av assosiasjoner, disse kan være av forskjellige typer. Det opp til forfatteren å bestemme hvilke typer

¹⁷Dette er for å bidra til det såkalte *samløkaliseringsprinsippet* som går ut på at det skal være mulig å finne alt om et emne på samme sted.

assosiasjoner som skal brukes i emnekartet og hvilke assosiasjoner som skal gjelde mellom to emner.

Dokumenter kan ha henvisninger og kryssreferanser til andre dokumentene i og utenfor dokumentsamlingen. Der er viktig å huske at emnekart lever et selvstendig liv i forhold til dokumentene og det kan være at en henvisning får en tilsvarende assosiasjon i et emnekart, men det er ingen selvfølge. Det vanlige er nok at henvisningene vil bli representert som assosiasjoner.

En merkbar begrensning ved henvisninger i bøker og WWW-hyperlenker er at de har bare en retning, og det finnes følgelig ingen muligheter til å komme tilbake gjennom samme dør som man brukte for å komme inn¹⁸. Assosiasjoner i Emnekart har ikke denne skavanken, og kan følges begge veier. Hvis *emne A* har en kobling til *emne B*, kan man følge koblingen tilbake fra til *emne A* fra *emne B*.

På samme måte som emnenavn, kan assosiasjoner gjøres gyldige innenfor bare et gitt perspektiv.

Emner spiller forskjellige *roller* i assosiasjoner. I assosiasjonen *søksmål* kan emner spille to roller: *saksøker* eller *saksøkt*. Både *saksøker* og *saksøkt* er emner av typen *person* og denne typen kan omfatte både fysiske og juridiske personer. Assosiasjonen *søksmål* uttrykkes på følgende måte i emnekart-formatet LTM[11]:

```
søksmål(saksøker : person, saksøkt : person)
```

Jeg tror det kan være fornuftig å utvide eksempelet ovenfor med emner, der Peder Ås er saksøker og Lars Holm saksøkt:

```
[på : person = "Peder Ås"]  
[lh : person = "Lars Holm"]  
søksmål(på, lh)
```

¹⁸Det er som å tenke på en dør som man kan gå gjennom, men når man har gått gjennom den og snur seg for å komme tilbake oppdager man at døren er borte.

Søksmål-assosiasjonen kan gjøres til et emne og beskrives nærmere:

```
[...]  
søksmål(på, lh) ~ søksmål-på-vs-lh
```

Nå er *søksmål-på-vs-lh* et emne som representerer søksmålet mellom Peder Ås og Lars Holm. Det bør merkes hvordan assosiasjonen mellom to emner ble selv gjort til et eget emne. Dette kalles reifikasjon (*reification*)¹⁹. Hele Emnekart-konseptet er bygget rundt emner og *alt* i emnekart kan defineres som et emne og på denne måten beskrives nærmere, f.eks. kan søksmålet ovenfor gis typen *nabokrangel*:

```
[...]  
[søksmål-på-vs-lh : nabokrangel = "Nabokrangel"]
```

Det er ikke plass til å presentere LTM i større omfang. For dypere introduksjon til og definisjon av emnekartformatet LTM viser jeg til Lars Marius Garshols rapport.[11]

5.2.3 Forekomster

Emnekart er en modell som beskriver emner og relasjoner på samme måte som et kart beskriver et geografisk område. Forekomster er skyggene som emnene kaster ned på dokumentsamlingen.

Forekomster knytter emnekart og dataressursene sammen, forekomstene til et emne inneholder relevant informasjon for emnet. Forekomstene kan være *interne* eller *eksterne*. Interne forekomster brukes typisk for å knytte relativt små mengder informasjon til et emne, som er så små at de like gjerne kan ligge i emnekartet, mens eksterne identifiseres ved en henvisning, f.eks. sidetall eller en URI²⁰.

¹⁹*Reifikasjon* betyr at et konsept eller en abstraksjon betraktes som noe virkelig og reelt. Man kan også bruke ord som "objektivering" eller "tingliggjøring". I Emnekart-perspektiv kan et nytt ord innføres: "emnifisering".

²⁰URI står for *Uniform Resource Identifier*. Det er en rekke tegn som identifiserer en

5.3 Muligheter

Fight ontological fascism!

— Steven R. Newcomb

Emnekart gir interessante muligheter. Modellen er enkel, men samtidig meget fleksibel. Den påtvinger verken forfattere eller lesere et regime av kataloger og hierarkier, men gir en frihet ved at det kan etableres vilkårlige relasjoner mellom vilkårlige tekster. Forfattere kobler sine arbeider til relevante emner ved hjelp av forekomster og til beslektede emner ved hjelp av assosiasjoner; lesere finner relevante tekster om et emne ved hjelp av forekomster og kan gå videre fra et emne til beslektede emner vha. assosiasjoner.

Hadde Vannevar Bush levd i dag, ville han sannsynligvis vurdert emnekart for *Memex*²¹.

En fascinerende egenskap er at emnekart lar seg flette sammen og på denne måten dannes det et nytt emnekart som inneholder emnene, assosiasjonene og forekomstene til de emnekartene som ble flettet sammen. Dersom samme emne var tilstede i begge emnekartene, vil emnet i sammenflettet emnekart få egenskapene fra samtlige opprinnelige emnekart som omtalte emnet.

Dersom Lovdata hadde et emnekart for alt i sine databaser, ville det ikke være behov for andre å ha samme innhold i sine baser. Norsk Lovkommentar kunne bygge videre på Lovdatas lov- og avgjørelsessamling og publisere eget emnekart som bare inkluderte kommentarene til lovene. Brukere som hadde abonnement på Norsk Lovkommentar kunne flette sammen emnekartene fra Lovdata som hadde lovene og Gyldendal

abstrakt eller fysisk ressurs. En type URI er en URL (*Uniform Resource Locator*), som brukes for å angi adressen til en ressurs. `http://xyz.example.com/dir/file.ext` er en URL.

²¹“Memex” er navnet på en maskin som skulle kunne lagre alt et menneske måtte ha av bøker, brev, dokumenter m.m. og deretter gjøre eieren av en Memex i stand til å finne frem i denne dokumentmengden på en enkel og intuitiv måte. Memex er beskrevet i Vannevar Bushs artikkel *As We May Think*, som er tilgjengelig på WWW på `http://ccat.sas.upenn.edu/~jod/texts/vannevar.bush.html`.

Rettsdata som hadde kommentarene, og til slutt sitte med en kommentert lovsamling. Gyldendal Rettsdata har ikke monopol på kommentering. Brukere kan ha sine egne emnekart der de føyer til nye assosiasjoner mellom eksisterende emner, legger til nye emner osv.

Mulighetene ender ikke der. I emnekart blir menneskelig forståelse av emner og relasjoner fanget opp. Det betyr at dersom kunnskap om innholdet i dokumenter foreligger som et emnekart, kan man utnytte denne kunnskapen ved søk etter relevante dokumenter om et gitt tema. Dette er mulig fordi emnekart beskriver *hvordan* et menneske forstår innholdet i et gitt dokument. Den formelle strukturen av denne beskrivelsen egner seg godt for formelle og presise spørrespråk. Dette skulle tilsi at man kan forbedre effektiviteten til søkesystemer ved at både presisjon og gjenfinningsgrad øker²².

²²Dette er bare en antakelse, jeg har ingen belegg for å forsvare mitt standpunkt om forbedret søkeeffektivitet.

Kapittel 6

Ontologisk analyse av rettskildefaktorene

Alle eventuelle forbedringer som en innføring av et emnekart-drevet system kan gi et eksisterende informasjonssystem, forutsetter at det blir laget et emnekart som beskriver emner og assosiasjoner som finnes i dokumentene i systemet. I mitt tilfelle ville det bety at det måtte lages et emnekart som beskriver alle emner og assosiasjoner som forekommer i alle rettskildefaktorer. En slik oppgave ville kreve mye arbeidsinnsats¹ og ville følgelig bli meget kostbar. Ambisjonsnivået *må* settes lavere og mer automatiserte metoder for etablering av emnekart må vurderes.

Emnekart kan lages på to måter som er grunnleggende forskjellige. I avsnitt 2.1 på side 12 nevnte jeg at indeksering kan gjøres manuelt eller automatisk. Det samme gjelder for emnekart.

En forsiktig start i å integrere emnekart i et eksisterende informasjonssystem kan være at et emnekart lages automatisk. Semantisk kvalitet blir selvfølgelig langt fra like god som ved manuell etablering av emner og assosiasjoner basert på forståelse av teksten.

¹Over 100 årsverk?

6.1 Tilgjengelige data

Henrik Ibsens skrifter hadde 90.000 sider med faksimiler som ble konvertert til maskinlesbar form ved at teksten var tastet inn for hånd og kodet i XML², slik at man i løpende tekst automatisk kunne se datoinformasjon, geografiske referanser, personreferanser og verksreferanser. Dette arbeidet tok sin tid, men resultatet ble meget godt.

Lovdatas dokumenter finnes allerede i digital form. Dokumentformatet Bokus er, vel, ganske sært, men det egner seg godt for automatisk analyse og konvertering. Utgangspunktet for etablering av et emnekart kan derfor tas i Lovdatas dokumentsamling. Lovdata representerer rettskildedefaktorer som dokumenter med både strukturerte og ustrukturerte data. Strukturerte data ligger i dokumentets hode.

Tilgjengelig dokumentsamling må analyseres med tanke på å avdekke emner og assosiasjoner som et datamaskinprogram deretter kan plukke opp og sette inn i emnekartet. Det er mye informasjon man kan hente ut av hvordan Lovdatas system er organisert i dag.

Det er forskjellige databaser som inneholder forskjellige typer rettskildedefaktorer. I hvilken base dokumentet ligger forteller noe om dokumentet, dets type.

Som jeg nevnte tidligere, kan emnekart flettes sammen. Det er derfor uten betydning om man velger å ha ett stort emnekart som omfatter alle basene, eller et emnekart for hver av basene.

Lovdata har dessuten flere typer registre: kronologisk, alfabetisk, systematisk, næringsregister, hjemmelsregister, departementsregister, kommuneregister og register over verneområder. Noen av disse registrene er utarbeidet manuelt og vi kan nyte godt av denne innsatsen når assosiasjonene mellom emnene skal etableres.

Noen typer dokumenter har et felt for stikkord. Innholdet i dette feltet er et resultat av en intellektuell prosess og kan brukes for å lage

²Man brukte et format som heter TEI-XML, jf. <http://www.tei-c.org/>.

assosiasjoner mellom emner. Lovdatas synonymliste kan også være nyttig.

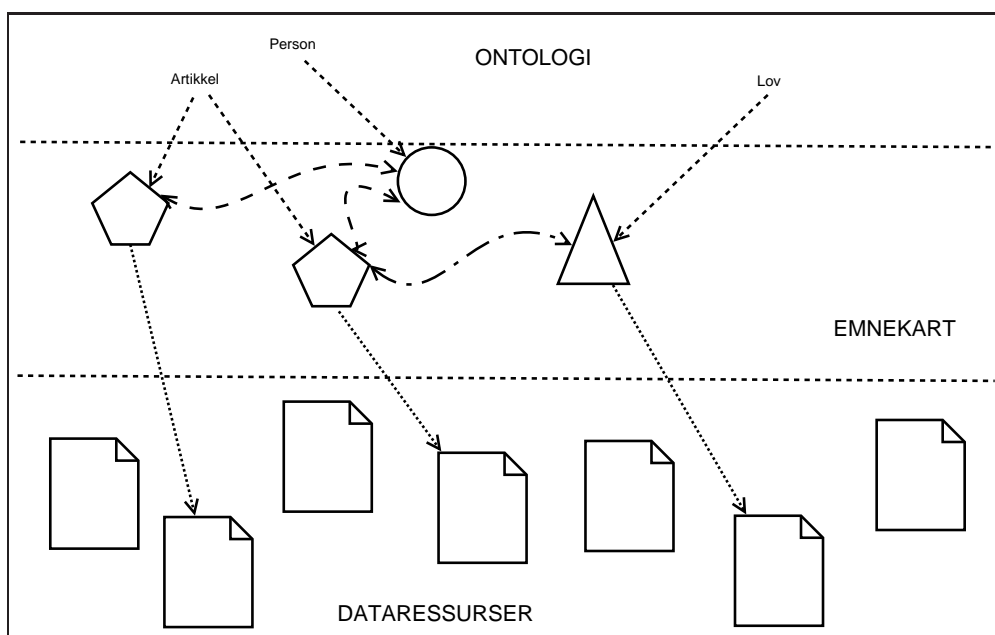
Skyggebasen inneholder oversikt over krysshenvisningene mellom dokumenter. Disse henvisningene kan benyttes for å etablere assosiasjoner mellom emner.

6.2 Ontologi-begrepet

I filosofi er *ontologi* en fundamental del av metafysikken, ontologi er læren om tingenes eksistens, vesen og egenskaper. “Hva eksisterer?” — er ontologiens grunnleggende spørsmål. Informatikere legger en annen mening i ordet “ontologi”. I informatikk er ontologi en datamodell som representerer en del av virkeligheten og som kan brukes til å arbeide med objekter og relasjoner i denne delen av virkeligheten innenfor rammene av datamodellen.

I forbindelse med Emnekart snakker man om ontologi når der er tale om en oversikt over hvilke typer av emner, assosiasjoner og forekomster et emnekart har; en ontologi kan sies å være et slags meta-emnekart. For en illustrasjon av forholdet mellom dokumentsamlingen, emnekart og ontologi se figur 6.1 på neste side.

Jeg skal videre forsøke å definere en ontologi for et emnekart som beskriver innholdet i et lite utvalg av Lovdatas dokumentbaser: *NL* (gjeldende lover) og *HRSIV* (sivile avgjørelser avsagt av Høyesteretts avdelinger). Denne ontologien er ikke ment å være uttømmende, men snarere skal illustrere hvordan prosessen foregår, dog slik at det skal være mulig å bruke ontologien til å lage et meget enkelt emnekart på bakgrunn av de nevnte dokumentbasene.



Figur 6.1: Ontologi, emnekart og dataressurser.

6.3 Analyse av dokumentbasen NL

Basen *NL* inneholder gjeldende regelverk i konsolidert form i henhold til Norsk Lovtidend avd. I og inneholder per 4. april 2006 777 lover.

6.3.1 Emner

Loven selv står så sentralt at det er naturlig å begynne med emnetypen *Lov*.

En lov er delt opp i mindre enheter. Kapittel, paragraf, ledd, litra, punktum, straffealternativ m.m. er identifikatorer som brukes for å angi nøyaktig hvilken del av loven det siktes til. Det beste hadde vært å la hver av de nevnte elementene være emner.

Lovdatas system kan identifisere en lovtekst med en paragrafs nøyaktighet, det gjøres ikke forskjell på ledd og punktum i samme paragraf. Det

er imidlertid ofte behov for å snakke om de enkelte leddene i en paragraf, f.eks. hvis det skrives en artikkel om avgrensning mot kontraktansvar i foreldelseslovens § 9 tredje ledd³ kan det være kjekt å kunne knytte den direkte til tredje ledd og ikke til § 9. Derfor vil jeg definere emnetypene *Ledd*, *Litra* og *Punktum*, selv om disse ikke kommer til å bli tatt i bruk med det aller første. I tillegg til disse tre kommer selvsagt *Kapittel* og *Paragraf*.

En lov identifiseres ved vedtakelsesdato og nummer. Vedtakelsesdato og nummer vil sammen gi en unik identifikator for loven, denne kombinasjonen egner seg godt til å bli brukt i temaidentitet.

Lover har alltid et navn og ofte et kortnavn. Lovens fullstendige navn kan brukes som navnet på emnet; dersom et kortnavn eksisterer, kan det plasseres i perspektivet "kortnavn".

Kapitler og paragrafer har også et nummer og ofte en tittel. Loven lar seg identifisere entydig ved dato og nummer. Kapitler og paragrafer i en lov kan dermed også identifiseres entydig bare ved å føye til kapittel- eller paragrafnummer til lovens unike identifikator. Kapittel- og paragraftitler kan brukes i emnenavn.

Tilsynelatende er det nok å oppgi loven og paragrafnummer for å entydig identifisere en paragraf. Så enkelt er det imidlertid ikke. En lov lever sitt liv frem til den blir opphevet. Den fungerer som en slags mappe der innholdet kan endre seg over tid. Og ofte gjør innholdet nettopp dette. Lovgiver har ingen betenkeligheter med å tilføye nye paragrafer, endre eksisterende og til og med flytte paragrafer for å oppnå en "bedre" struktur i loven (f.eks. har åndsverkloven av 1961 vært gjenstand for store endringer og omstruktureringer gjennom årene). Samme skjebne kan ramme kapitler. Kapittel 2 i dagens åndsverkslov ble i sin helhet gitt på ny ved lov av 2. juni 1995 nr. 27, tidligere inneholdt kapitlet §§ 11–24, nå har det §§ 11–38b. Paragrafer inneholder ledd på samme måte som kapitler inneholder paragrafer.

³Eksempelet er hentet fra Emnebank for privatrettslige emner på <http://www.jus.uio.no/ifp/emnebank.html>.

Jeg har i skrivende stund ikke noe godt gjennomtenkt forslag til hvordan endringsproblematikken best skal løses. Problemet kan skyves under teppet ved å si at jeg velger å se på rettskildefaktorene på et gitt tidspunkt. *Tid* er en kompliserende faktor.

6.3.2 Forekomster

Relevant informasjon om et emne befinner seg i forekomster. Forekomster er informasjonsressurser som forteller noe om emnet.

Lovene forberedes av departementer, har vedtakelsesdato og ikrafttredelsesdato. Resultatet av lovgivningsprosessen er en lovtekst. Lover og vedtak om lovers ikrafttredelse, oppheving og virksomhetsområde kunn­gjøres i Norsk Lovtidend.

Et emne av typen *Lov* kan dermed gis forekomster av følgende typer: *Dato*, *Ikrafttredelsesdato*, *Tekst*.

Man kan lagre informasjon om ansvarlig departement i en forekomst, men meget fort vil man finne ut at samme departement kan være ansvarlig for mange andre lover og det kan dessuten være mye annet interessant å si om et departement. Departement egner seg derfor til å bli gjort til et emne. Lov og departement kan knyttes sammen vha. en assosiasjon.

Når det gjelder publisering i Lovtidend, har man også to muligheter. Som med departementet, kan man lagre informasjon om publisering i Lovtidend i en intern forekomst. Man kan imidlertid også knytte loven og det aktuelle nummeret av Lovtidend (representert som et emne av typen *Lovtidend*) sammen vha. en assosiasjon. For meg fremstår det siste alternativet som mer logisk og naturlig.

Datoer (*Dato* og *Ikrafttredelsesdato*) er korte tekststrenger og egner seg godt til å bli lagret som interne forekomster. Selve lovteksten kan plasseres i forekomsten av typen *Tekst*. Denne er ekstern og kan representeres som en URL (adresse) hos Lovdata hvor loven kan leses⁴.

⁴URL-en til bustadsoppføringslova (buofl.) er <http://websir.lovdata.no/lex/lov/>

Kapitlene har ikke noen egenskaper ut over nummer og navn, og disse vil allerede fremkomme av emnenavnet. Et kapittel inneholder et sett med paragrafer. Forekomsten *Tekst* kan brukes til å romme innholdet i paragrafene som kapitler omfatter. Denne forekomsten er ekstern og kan representeres som en URL hos Lovdata hvor kapitlet kan leses⁵.

Med paragrafene er det som med kapitler: nummer og navn vil være i emnets navn og innholdet i paragrafen kan plasseres i en ekstern forekomst av typen *Tekst* med URL hos Lovdata som identifikator⁶.

Vi har brukt samme type forekomst (*Tekst*) til å referere til innholdet i både loven, kapitler og paragrafer.

URL-adresser for lov, kapittel og paragraf spiller to roller. For det første kan man få tilgang til teksten ved å hente dokumentet fra URL-adressen. For det annet fungerer URL-en som en entydig identifikator på loven, kapitlet eller paragrafen og egner seg til å bli brukt som temaidentitet.

URL-er hos Lovdata som temaidentitet lar seg bruke bare innenfor rammene av min oppgave der jeg ser på rettskildefaktorene på et gitt tidspunkt. Det hender imidlertid at lover blir opphevet, og opphevede lover flyttes fra basen *NL* til basen *ROL* med tilsvarende endring i URL-adressen⁷. Det er en grunnleggende forutsetning for temaidentitet at den for et tema ikke forandrer seg over tid.

6.3.3 Assosiasjoner

Assosiasjoner knytter emner sammen. Paragrafene vil fremstå som selvstendige elementer dersom de ikke knyttes til loven som de egentlig hører

nl/hl-19970613-043.html.

⁵URL-en til kapittel 4 i buofl. er <http://websir.lovdata.no/lex/lov/nl/hl-19970613-043.html#map004>.

⁶URL-en til § 31 i buofl. er <http://websir.lovdata.no/lex/lov/nl/hl-19970613-043.html#31>.

⁷Dersom buofl. oppheves, vil den opphevede loven være tilgjengelig på <http://websir.lovdata.no/lex/rol/nl/hl-19970613-043.html>. Merk at "nl" ble byttet ut med "rol".

til. Det må altså etableres en assosiasjon mellom paragrafer og loven. Paragraf er en del av loven, typen på assosiasjonen kan være *er-en-del-av*.

Dersom loven inneholder kapitler, må også disse knyttes til loven og paragrafene må også knyttes til sine kapitler. Assosiasjonstypen som brukes for å koble sammen loven og paragrafer kan benyttes her, dvs. *er-en-del-av*.

I assosiasjonen *er-en-del-av*, kan emner spille to typer roller: *Helhet* og *Del*. *Helhet* skal representere helheten som *Del* er en del av.

Dersom vi nøyer oss med den ene assosiasjonen, får vi til bare hierarkisk navigasjon, dvs. at man kan gå fra lov til kapittel og paragraf, men ikke fra en paragraf til neste paragraf. Sekvensiell navigasjon mellom kapitler og paragrafer er ønskelig, ellers ville papirbaserte lovsamling ha flere funksjoner enn vår emnekart-drevet system. En ny type assosiasjon må defineres for formålet, la oss kalle den *frem-tilbake*. I denne kan emner spille to typer roller: *Forrige* og *Neste*.

Lover, kapitler og paragrafer kan endres og oppheves av nyere lover. Disse endringene danner koblinger som vi aller helst bør ha med i vårt emnekart. Dette kan oppnås ved å introdusere assosiasjonene *endret* og *opphevet* med emnerollene *Endret av*, *Endrer*, *Opphevet av* og *Opphever*.

I avsnitt 3.5 på side 26 nevnte jeg at henvisningene er en av teknikkene som brukes for å gi en hyperstruktur en lineær utforming. I lovtekstene i lovsamlingen kan vi spore opp to typer henvisninger: autoritative henvisninger til andre bestemmelser som gis av lovgiver, og redaksjonelle henvisninger som tilføyes av Lovdatas redaksjon og står i noter. Rettskildemessig er det stor forskjell på disse to typene, derfor bør vi ha to typer assosiasjoner: *henv* for autoritative henvisninger og *red-henv* for redaksjonelle henvisninger. Ofte utgjør redaksjonelle henvisninger mellom to bestemmelser et svar på en autoritativ henvisning som knytter de samme bestemmelsene, men som går i motsatt retning. Den redaksjonelle henvisningen etableres da for å omgå den autoritative henvisningens enveisnatur. Assosiasjonene i emnekart lar seg følge begge veier, og når emnekartet

Emner	Forekomster	Assosiasjoner
Lov	Dato (intern)	forberedt-av
Kapittel	Ikrafttredelsesdato (intern)	publisert
Paragraf	Tekst (ekstern)	endret
Ledd		opphevet
Litra		henv
Punktum		red-henv
Departement		
Lovtidend		

Tabell 6.1: Emne-, forekomst- og assosiasjonstyper i dokumentbasen *NL*.

skal bygges bør man passe på at det ikke etableres en assosiasjon *red-henv* mellom to emner hvis det allerede eksisterer en assosiasjon *henv* mellom de samme emnene.

En oversikt over forskjellige typer emner, forekomster og assosiasjoner som kan identifiseres i dokumentbasen *NL* er gitt i tabell 6.1.

6.4 Analyse av dokumentbasen *HRSIV*

Basen *HRSIV* inneholder sivile avgjørelser avsagt av Høyesteretts avdelinger.

6.4.1 Emner

En god start i å finne emner kan være å finne noe man kan ønske å beskrive nærmere og å si mer om.

Avgjørelsen selv må uten tvil bli et emne, la oss kalle emnetypen *Avgjørelse*. Det er to typer avgjørelser, og vår ontologi bør ha egne emnetyper for de forskjellige typene av avgjørelser: *Dom* og *Kjennelse*. Avgjørelsen avsies av Høyesterett, som er en domstol og som dessuten er den høyeste rettsinstans. Både domstol og rettsinstans kan gjøres til

et egne emner, slik at disse kan knyttes sammen med avgjørelsene som domstolen avsier.

Det er to parter som får sin tvist løst ved domstolen. Partene har ofte prosessfullmektiger. Skal parter og deres advokater få anledning til å bli med i emnekartet? Advokater — kanskje. Parter — kanskje ikke. Hensynet til personvern er et argument for å la være å ha egne emner for parter i en tvist.

Derimot kan det være nyttig å ha advokater og dommere med i emnekartet. Advokater vil bli representert som emner av typen *Prosessfullmektig* og dommere som emner av typen *Dommer*. På denne måten kan man finne svar på spørsmål av typen: *Hvilke saker har advokat NN vunnet?* Dommer Gunnar Aasland argumenterer godt⁸. Avgjørelsene som Aasland er med på å avsi kan knyttes til emnet *Gunnar Aasland* av typen *Dommer*.

6.4.2 Forekomster

Jeg minner om at forekomster inneholder relevant informasjon om et emne.

Dommer avsies på et gitt tidspunkt. Dette kan være en intern forekomst av typen *Dato*.

Publisering av høyesterettsavgjørelser skjer i *Rettstidende*. Årgangen av *Rettstidende* og sidetall var eneste måte å identifisere en dom på tidligere. Etter hver har en nyere måte å vise til avgjørelsene på vært tatt i bruk: år og saksnummer. F.eks. kan dommen i *Rt-2003-1821* også nås ved å bruke referansen *HR-2003-00576*.

Jeg nevnte i forrige avsnitt at av hensyn til partenes personvern, kan opplysningene om parter plasseres i forekomster, istedenfor at det opprettes egne emner for parter.

Dommen ustyres med et sammendrag. Når Lovdata mottar avgjørelser

⁸Det er min personlige oppfatning, men jeg er ikke alene om å mene det.

fra Høyesterett, kommer de ofte med et sammendrag, men ikke alltid. Rettskildemessig har sammendragene status som teori uansett om de skrives av redaksjonen i Høyesterett eller Lovdata, og det er derfor ikke noe poeng å behandle sammendraget forskjellig avhengig av hvem som skriver det. Normalt er sammendraget relativt kort, derfor kan det plasseres i en intern forekomst *Sammendrag*.

For selve teksten i avgjørelsen kan vi benytte en ekstern forekomst av typen *Tekst*.

6.4.3 Assosiasjoner

“Høiesteret dømmer i sidste Instans.”⁹ Det betyr at før saken kommer opp til Høyesterett, har det vært behandlet i andre lavere instanser. Avgjørelsene slik de vises hos Lovdata har et felt *Saksgang*. Innholdet i dette feltet egner seg godt til å bli gjort om til en assosiasjon. Avgjørelsene i saken i forskjellige instanser kan knyttes sammen vha. assosiasjonen *saksgang* som går til den foregående avgjørelse (som også godt kan ha en *saksgang*-assosiasjon til sin forgjenger osv.).

Prosessfullmektiger kan knyttes til dommen vha. assosiasjonen *prosedert-av*, og dommere vha. assosiasjonen *avsagt-av*.

Det er en mangfoldighet av henvisninger til lover, avgjørelser, teori og andre rettskildefaktorene som avgjørelsen bygger på. Hver av disse bør gjøres til en assosiasjon. Assosiasjonens type bør avhenge av hvilken relasjon henvisningen gir uttrykk for. Her bør man vurdere om det bør innføres et nokså begrenset sett av typer assosiasjoner, f.eks. kan man ha de samme typene som brukes i *Australian Case Cimator*, jf. ovenfor i avsnitt 3.6 på side 27. Identifisering av forskjellige typer henvisninger i eksisterende tekster vil kreve mye menneskelig arbeid. I denne omgang nøyer jeg meg med å definere en type assosiasjon for alle henvisninger: *henv.*

⁹Jf. GrL. § 88

Emner	Forekomster	Assosiasjoner
Avgjørelse	Dato (intern)	saksgang
Dom	Ankende part (intern)	prosedert-av
Kjennelse	Ankemotpart (intern)	avsagt-av
Instans	Sammendrag (intern)	
Domstol	Tekst (ekstern)	
Prosessfullmektig		
Dommer		

Tabell 6.2: Emne-, forekomst- og assosiasjonstyper i dokumentbasen *HRSIV*.

En oversikt over forskjellige typer emner, forekomster og assosiasjoner som kan identifiseres i dokumentbasen *HRSIV* er gitt i tabell 6.2.

6.4.4 Perspektiv

Jeg ønsker også å si noen ord om et egenskap ved Emnekart med et stort potensial, men som ikke til nå ikke har vært tatt i bruk i noe stor grad – perspektiv.

Perspektiv tillater forskjellige virkelighetsoppfatninger til å komme til uttrykk. Sakens parter *har* forskjellige oppfatninger av sakens faktum, ellers ville det ikke ha vært noen rettssak.

Partenes motstridende syn kunne knyttes til perspektivene *Ankende parts syn* og *Ankemotpartens syn*. Partenes syn på hva? Nå ser vi at vi må ha flere emnetyper for avgjørelser, nemlig *Faktum* og *Jus*, som kan belyses forskjellig avhengig av om vi velger perspektivet *Ankende parts syn* eller *Ankemotpartens syn*.

Og så kunne vi ha egne perspektiv for dommernes syn på faktum og jus.

Kapittel 7

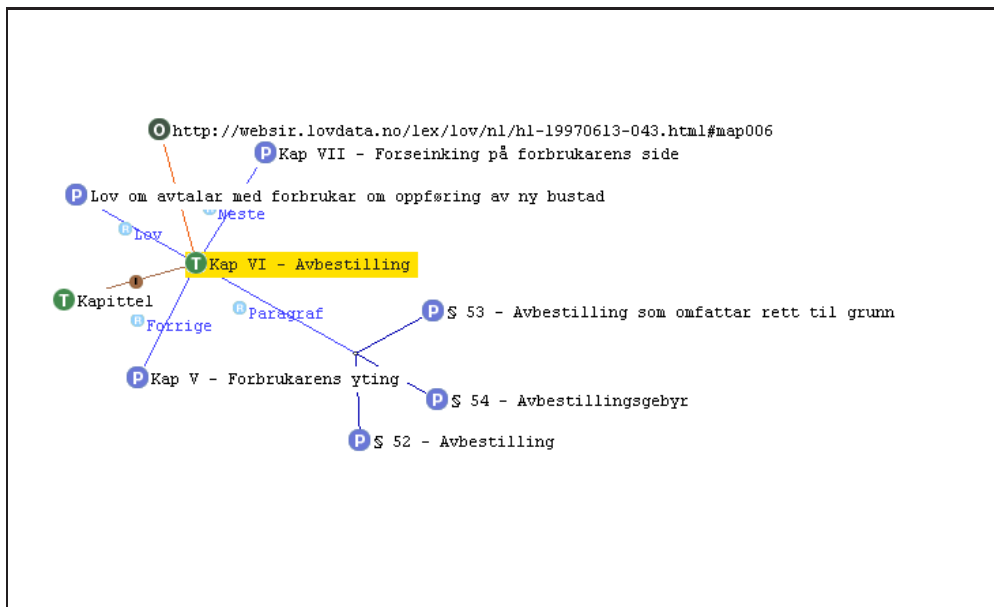
Fra lov til emnekart. Et forsøk.

Et bilde er tusen ord verdt, og når man prøver å lære noe nytt, lærer man mest ved å anvende det nye. Det går sjeldent bra første gang, men øvelse gjør mester og hvert nytt forsøk går bedre enn det forrige.

Jeg har snakket nok om Emnekart og mulig anvendelse av disse for å strukturere innholdet i juridiske tekster. Jeg ønsker å prøve å bygge et emnekart basert på en av juridiske tekster, en lov. Det er ikke det samme som å definere en ontologi. En ontologi forteller bare hvilke typer elementer et emnekart skal inneholde.

Jeg stiftet bekjentskap med Emnekart da jeg vurderte visualiseringsmetoder for rettslige hyperstrukturer og det følger derfor av seg selv at innenfor rammene av denne oppgaven måtte jeg prøve å overføre noe av de juridiske tekstene til et emnekart og mate dette inn i et program, som kunne tolke emnekart-formatet og vise emner, forekomster og assosiasjoner i grafisk representasjon. Den praktiske nytten av et slik eksperiment ville være trolig begrenset, men man ville få et annet syn på loven.

Jeg spente beltet om livet og gikk rett på sak. Veien fikk bli til mens jeg gikk. *Lov om avtaler med forbruker om oppføring av ny bustad m.m.* fikk rollen som prøvekanin. Jeg ville representere loven, kapitler og paragrafer som emner, samt knytte kapitler til loven og paragrafer til kapitler og loven. Dessuten skulle kapitler og paragrafer kobles slik at sekvensiell



Figur 7.2: Grafisk representasjon av buofl. med kap. 6 i sentrum.

navigasjon mellom kapitler og paragrafer kunne realiseres.

Med LTM[11] kan emnekart¹ lages i en vanlig tekstredaktor. LTM er også mer kompakt og leselig enn f.eks. XTM. Jeg valgte å bygge emnekartet i LTM.

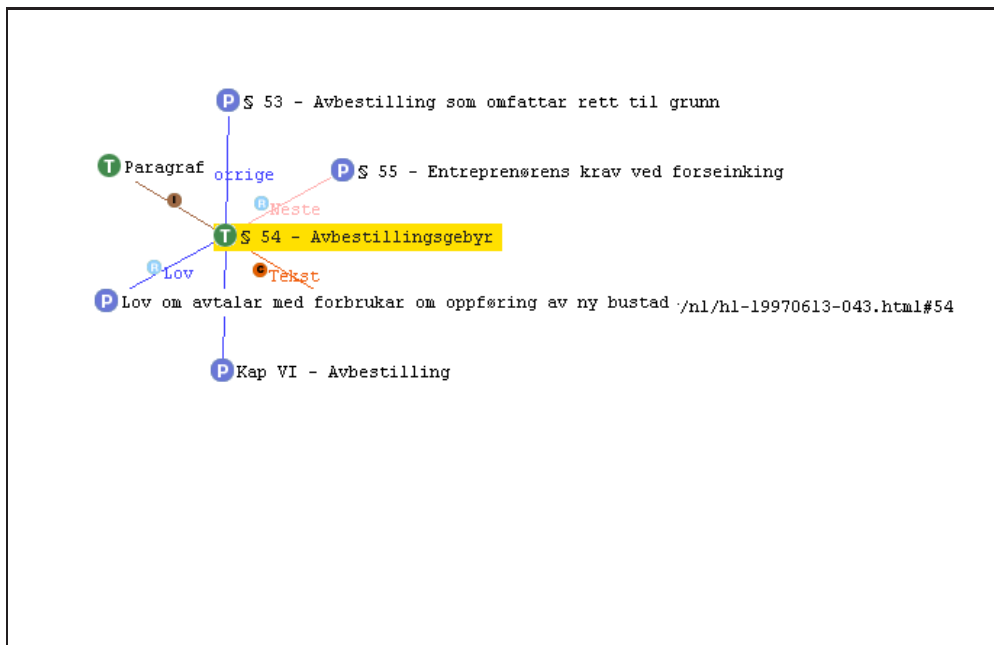
Jeg hentet loven i HTML-format² og laget deretter et lite program som analyserte HTML-filen og identifiserte kapitler og paragrafer. For hvert kapittel og hver paragraf var det generert et sett av emner, assosiasjoner og forekomster.

Resultatet av den automatiske prosessen var et enkelt emnekart i LTM format.

Til slutt føyde jeg til noen avgjørelser som viser til bestemmelsene i buofl. Disse avgjørelsene er representert som emner uten noen forekomster.

¹Merk liten "e" i "emnekart".

²URL-en til loven i fulltekst er <http://websir.lovdato.no/lex/lov/nl/nl-19970613-043.html>.



Figur 7.3: Grafisk representasjon av buofl. med § 54 i sentrum.

Et utdrag LTM-filen som ble produsert er inntatt i vedlegg C på side 79 og skjermbilder av resultatet av visualiseringen i *TMNav*³ er illustrert i figurene 7.1 på side 59, 7.2 på forrige side og 7.3.

³*TMNav* er en del av prosjektet *TM4J – Topic Maps for Java*. For detaljer om *TM4j* se <http://www.tm4j.org/>.

Kapittel 8

Representasjon og navigasjon

Så langt har jeg brukt relativt mye plass på å fortelle om Emnekart. Dette har vært nødvendig, ettersom emnekart kan danne grunnlaget for en bedre utnyttelse av hyperstrukturer i rettslige informasjonssystemer.

Strukturene som blir fanget opp i et emnekart kan anvendes for å gi brukere av rettslige informasjonssystemer alternative måter for å finne frem til relevante rettskildefaktorer og oppdage nye sammenhenger. Den grafiske representasjonen kan gi en topografi over dokumentstrukturen (eller til og med innholdet i dokumentene¹) og dermed hjelpe brukere med å finne frem til det interessante og nyttige ved at strukturen lar seg utforske.[13, s. 51]

Et emnekart med rettskildefaktorer kan potensielt inneholde mange millioner emner og assosiasjoner. Emner kan ha mange egenskaper. Utfordringen når man skal lage et grensesnitt for utforskingen av innholdet i emnekart blir å vise det som kan være nyttig for brukeren i søkeprosessen og skjule det som vil virke forstyrrende. Det sier f.eks. seg selv at med millioner av emner i emnekartet, vil bare et lite utvalg av disse la seg presentere om gangen. I dette avsnittet vil noen teknikker for representasjon av og navigasjon i hyperstrukturer bli presentert².

¹Dette forutsetter riktignok at det ytes en betydelig redaksjonell innstats.

²Det er ikke anledning til å foreta analyse av fordeler og ulemper ved de forskjellige

I avsnitt 1.4 på side 7 nevnte jeg om forskjellige funksjoner til et rettslig informasjonssystem: søkefunksjon, relevansvurdering og kildefunksjon. Emnekart eksisterer som et lag utenfor og uavhengig av dokument-samlingen og kan dermed ikke i seg selv oppfylle kildefunksjonen godt, men ved bruk av emnekart vil kildefunksjonen ivaretas ved at emner kan ha forekomster, som kan brukes i juridisk argumentasjon. Grensesnittet må utformes på en måte som vil gi brukeren en enkel tilgang på forekomstene som brukeren måtte være interessert i.

Det er to aspekter ved visualisering av hyperstrukturer som kan identifiseres og dermed bør belyses: *representasjon* og *navigasjon*.

8.1 Representasjon

Representasjon av hyperstrukturer og rettskildedefaktorer vil ha direkte innvirkning på relevansvurdering av innholdet i et informasjonssystem.

Man kan skille mellom relevansvurdering i vid og snever forstand. I avsnitt 1.4 på side 7 snakket jeg om relevansvurdering av et enkelt dokument, det var relevansvurdering i snever forstand. Ved visualisering av rettslige hyperstrukturer får relevansvurderingen en ny dimensjon — brukeren å gis en oversikt over strukturen, slik at han kan gjøre seg opp en mening om hva slags struktur han har foran seg og *hvor* i strukturen kan interessant materiale skjule seg. Relevansvurdering i vid forstand skjer altså fra et fugleperspektiv, man får “a view from the top of the mountain”. Etter hvert som brukeren blir mer og mer fortrolig med strukturen, kan han velge å se mer av detaljer, han begynner å se trær og ikke bare skog. Jo lenger ned man kommer, desto mer kan man lære om elementene i strukturen, men samtidig vil det globale bildet forsvinne.

Hvis hyperstrukturene er representert som emnekart, må brukeren ved visualisering av innholdet i emnekartet raskt introduseres med ontologien, med forskjellige typer emner og assosiasjoner. Det er ikke

teknikkene innenfor rammene av denne oppgaven.

nødvendig at brukere konfronteres med begrepene emne og assosiasjon, men grensesnittet må være utformet på en slik måte at emner og assosiasjoner av samme type lett lar seg identifisere og skiller seg fra emner og assosiasjoner av andre typer.

Brukeren må gis mulighet til å se på de forskjellige attributtene — dvs. forekomster — som emnene måtte ha. Det er også viktig å gi brukeren oversikt over hvilke andre emner emnet er knyttet til og på hvilken måte — grensesnittet må vise assosiasjoner som emnet deltar i, de andre deltaker-emnene i assosiasjonene og rollene de forskjellige emnene spiller i assosiasjonene.

8.2 Navigasjon

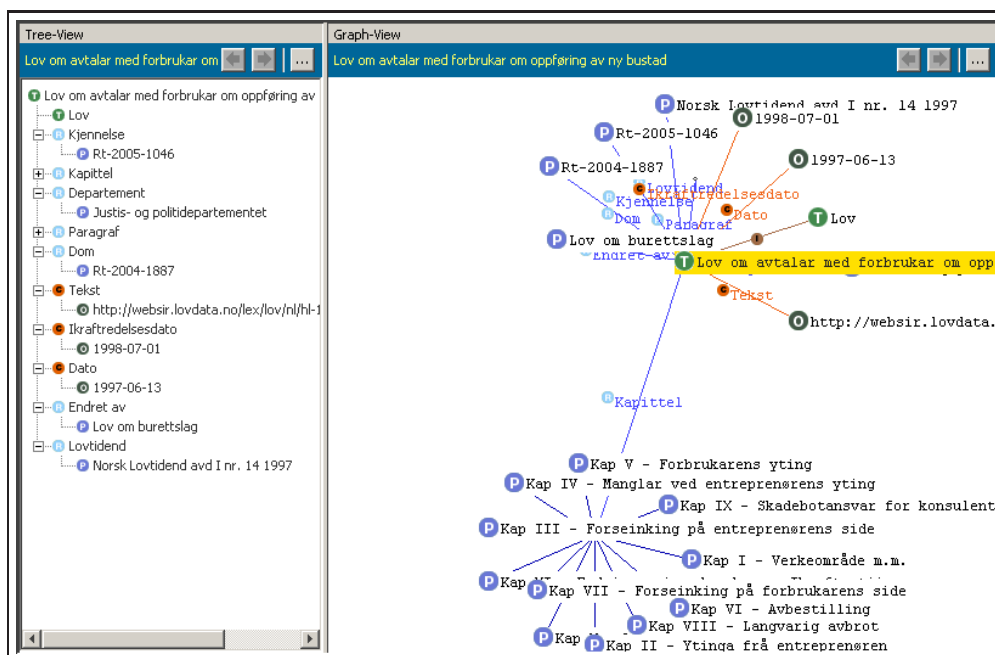
Navigasjon dreier seg om orientering og bevegelse i hyperstrukturene. Brukeren må føle seg trygg i hyperstrukturen, men må også gis muligheter til å utforske strukturens mørke hjørner uten å gå seg vill.

I Lovdatas system blir WWW-hyperlenkene brukt til navigasjon mellom forskjellige baser, dokumenter og i hyperstrukturen³. Systemet kan tvilsomt sies å gi brukeren noen som helst form for oversikt eller støtte på ferden.

8.3 Visualiseringsmetoder

I dette avsnittet vil jeg ta for meg visualisering av hyperstrukturene i vid forstand, ikke de enkelte elementene i disse, og presentere noen mulige visualiseringsmetoder.

³Se om forskjellige typer henvisningsknapper i Lovdata i kapittel 4.



Figur 8.1: Visualisering av hyperstrukturer: kombinasjon av tre og graf.

8.3.1 Grafer og trær

En hyperstruktur er en nettverkstruktur, og det er naturlig å representere emnekart som en graf, der emner er noder og assosiasjoner — kanter⁴.

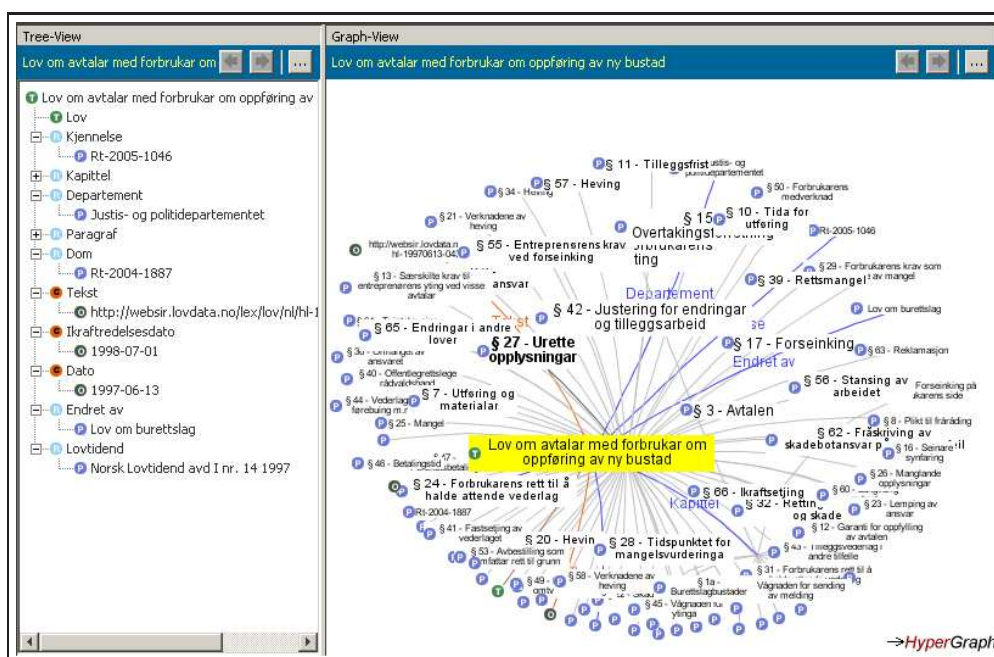
Det finnes grenser for hvor mange noder man kan representere i en graf på en gang, det blir fort kamp om plass på skjermen, når det er snakk om millioner av noder som skal vises. Overgang til hyperbolsk geometri⁵ i visualiseringen gir muligheter til å vise et større antall noder på en gang.

I figur 8.1 er det tatt med en visualisering av emnekartet som ble resultatet av forsøket i kapittel 7 på side 58, på figur 8.2 på neste side er det resultatet av visualiseringen av samme emnekartet ved hjelp av en hyperbolsk graf⁶. Det bør legges merke til at begge figurene egentlig viser

⁴Kanter er forbindelser som holder nodene i en graf sammen.

⁵Til forskjell fra "vanlig" s.k. euklidsk geometri, kan det gjennom et punkt på et hyperbolsk plan trekkes ubegrenset mange parallelle linjer til en gitt linje på planet.

⁶Selve visualiseringen var i dette tilfelle gjort av *HyperGraph*. I første eksempel ble

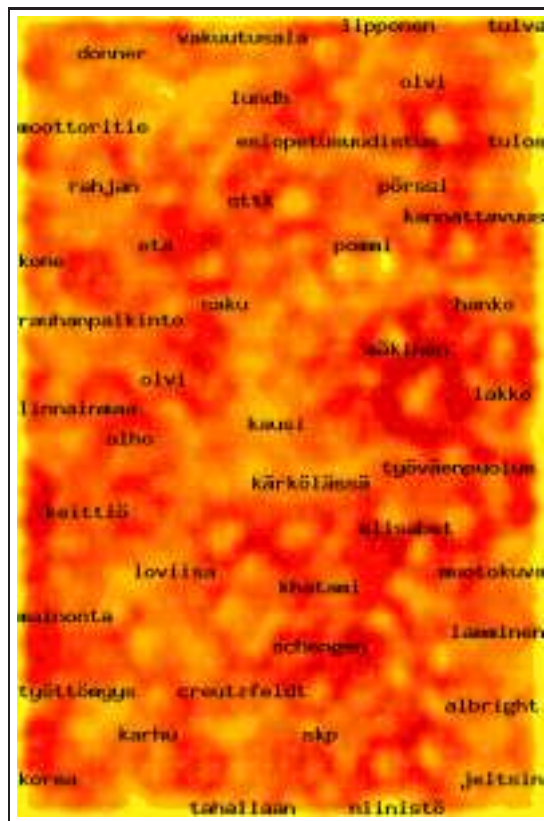


Figur 8.2: Visualisering av hyperstrukturer: kombinasjon av tre og hyperbolsk graf.

en kombinasjon av et tre og en graf: på venstre side er det en visualisering av den valgte noden med assosiasjoner, relaterte emner og forekomster, mens på høyre side er det en visualisering av strukturen som noden er en del av. Trær er hierarkiske datastrukturer. Emnekart er ikke et hierarkisk struktur og kan ikke *per se* representeres som et tre, men små deler av emnekart — f.eks. et emne med dets attributter — kan gi en mulighet for slik visualisering. Hierarkiske, ordnede trær er lettere for mennesker å tolke enn kompliserte, anarkiske grafer.[13, s. 53] En kombinasjon av et tre og en graf som de nevnte illustrasjonene viste, gjør at brukeren kan lære mer om en node man har valgt eller har tenkt å gå videre til uten å måtte forlate sitt ståsted.

Når informasjon ligger lagret i emnekart, kan man identifisere relaterte emner — de vil være koblet til hverandre vha. assosiasjoner, og for å gi brukeren en bedre oversikt over hyperstrukturen kan relaterte emner

TouchGraph benyttet.



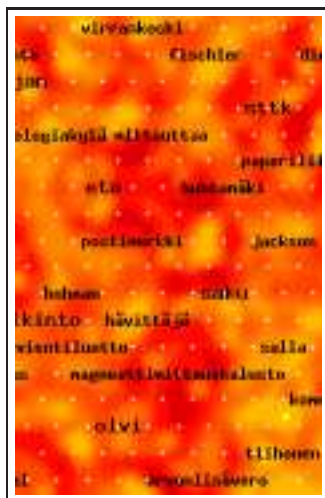
Figur 8.3: Self Organizing Map.

grupperes sammen i klaser, og emner som er koblet sammen ved en assosiasjon kan vises i nærheten av hverandre. Emner kan også grupperes etter emnetyper eller forekomster.

Det er også mulig å representere grafen i tredimensjonell rom, dermed blir det plass til å vise flere noder på samme tid.[13, s. 53–54]

8.3.2 Kart

Et emnekart er et kart, så hvorfor ikke vise det som et kart også? Det lar seg gjøre. Kartet kan være to- eller tredimensjonell og relaterte emner kan vises i nærheten av hverandre. Forskjellige farger kan benyttes for å markere forskjell i emnetettheten i de forskjellige regionene på kartet.



Figur 8.4: Self Organizing Map.

Dersom kartet er tredimensjonell, kan man ha fjell der mange emner samler seg og daler der det er få eller ingen emner.

Jeg har ikke hatt tilgang på verktøy for å lage et kart av mitt emnekart, men vil likevel illustrere hvordan visualisering av hyperstrukturer kan gjøres vha. kart. Jeg tar med to skjermbilder⁷ av såkalte *Self Organizing Maps* (SOM), som i dette tilfelle er bygget på 20.000 finske nyhetsartikler i perioden 1. januar–10. oktober 1997. Figur 8.3 på forrige side viser det opprinnelige kartet og figur 8.4 viser kartet slik det ble da jeg trykket på området ved ordet “eta”⁸. Målestokken på det siste kartet er større, og det er kommet frem noen navn som ikke var med på det opprinnelige.

Kart kan egne seg for å representere f.eks. alle paragrafer i alle lover på en gang. Grunnet massiv bruk av krysshenvisninger mellom lovbestemmelsene, ville relaterte bestemmelser sannsynligvis bli plassert i nærheten av hverandre på et SOM-kart. For identifikasjon av de forskjellige områdene på kartet kunne man bruke navn på lovene når målestokken er liten, og etter hvert som målestokken ble større, kunne man gå over til numre og titler på paragrafer. Dette kunne være et

⁷Eksemplene er fra <http://websom.hut.fi/websom/stt/doc/eng/>.

⁸Jeg vet ikke hva dette ordet betyr...

interessant forsøk.

En ulempe med kart er at de er svært omtrentlige. De egner seg godt for å la brukeren velge et større eller mindre område, men når området er valgt, bør andre representasjons- og navigasjonsteknikker brukes. Kart-, graf- og tre-representasjoner kan brukes parallelt. Ubegrenset mange emner kan dyttes inn i et kart, derfor kan kart brukes for representasjon og navigasjon på de aller største mengdene med dokumenter. Etter hvert som antallet blir mindre, kan man gå over til en graf for representasjon og navigasjon, og til slutt til et tre.

Kapittel 9

Resultatene og veien videre

Våre liv styres i stor utstrekning av grenser og frister av forskjellige slag. Jeg begynner nå å nærme meg to grenser: det maksimale antallet ord som en avhandling kan inneholde og innleveringsfristen. Tiden har derfor kommet for å foreta en oppsummering av det som har vært sagt, gjort og tenkt, og se på hva fremtiden kan bringe oss.

Det er mitt håp at denne oppgaven kan være nyttig for de som jobber med utvikling av rettslige informasjonssystemer. Det som står i de første kapitlene har ikke vært nytt stoff i seg selv, men er tatt med for å kunne gi en rask start til de som aldri har vært i nærkontakt med informasjonssøking. I de øvrige kapitlene dokumenteres resultatene av mitt forsøk på å velge en datamodell for å representere hyperstrukturer, lære mer om den valgte modellen (Emnekart), definere ontologi for et utvalg av rettskildedefaktorer og gjennomføre en enkel test av teknologien.

Vi er kommet til et tidspunkt der den tilgjengelige datamengden er blitt så stor at konvensjonelle metoder ikke lenger gir et tilfredsstillende søkeresultat og nye løsninger må prøves. Søkemotorselskapene satser fortsatt på statistisk syntaktisk analyse av tekst for å gi bedre presisjon og gjenfinningsgrad. De har ikke noe valg, for disse selskapene lever av syntaktiske søk. Men samtidig forsker de på og prøver ut nye metoder og

teknologier¹.

Det var jurister som var pionerene i EDB-basert leting etter tekstmønstre i dokumenter. Nå ser det ut at filologene ligger langt foran når det gjelder et system for leting etter mening i dokumenter² — der er referanser til steder, personer og verk markert på en identifiserbart måte.

Jeg har ikke hatt som mål å lage en prototyp til et system med et nytt grensesnitt der rettslige hyperstrukturer kan utnyttes for å gi bedre søk enn tilfelle er i dag, men har i kapittel 7 demonstrert hvordan en lov alternativt kan presenteres og nevnte flere visualiseringsmetoder i kapittel 8.

Dersom et fremtidig system vil bygge på emnekart, vil man i tillegg til navigasjon i henvisningsstrukturene etter hvert kunne utvide systemet til å kunne gjøre semantiske søk i dokumentene. Jeg mener Emnekart er spennende og godt verdt et forsøk.

Er Emnekart løsningen? Det er ikke uvanlig i IT-bransjen at leverandøren selger (eller i det minste gjør et forsøk på å selge) en teknologi som han gir ut for å være en løsning. Kunder vil ha løsninger, og en teknologi er aldri en løsning i seg selv. Emnekart er en teknologi.

Løsninger bygger på teknologier. Et fremtidig grensesnitt til rettslige informasjonssystemer kan nyte godt av at systemet baserer seg på emnekart. WWW-baserte grensesnittet til *Henrik Ibsens skrifter* kommer til å bygge på emnekart, og grensesnittet til det nye norske rettslige informasjonssystemet kan også bygge på emnekart.

Fritekst-søk vil ikke gjøres overflødig ved at emnekart tas i bruk som hjelpemiddel. Fritekst-søk pløyer gjennom innholdet og finner ord og uttrykk som stemmer overens med brukerens søkeargument. Emnekart gir ikke denne muligheten, de gir oversikt over innholdet, men gir ikke tilgang til selve innholdet. Men fordi det i emnekart ligger en

¹Det skjer mye forskning, utvikling og testing av prototyper hos de store selskapene, f.eks. i *Yahoo! Research* og i *Google Labs*.

²Arbeidet på prosjektet *Henrik Ibsens skrifter* ble påbegynt allerede i april 1998.

beskrivelse av hvordan data i dokumenter skal forstås, kan fritekst-søk gjøres mer presis ved at kunnskapen som ligger i emnekartet utnyttes ved silingsprosessen. Det kan også tenkes at brukeren gis en mulighet til å inkludere emner og assosiasjoner i søkeargumentet³. For å utforme søkeargumentet med riktige emner og assosiasjoner vil det imidlertid være en forutsetning at brukeren kjenner at det ligger et emnekart bak søkesystemet og ikke minst kjenne emnekartets ontologi (dvs. hvilke typer emner, assosiasjoner og forekomster som er representert i emnekartet). Det er urealistisk å kreve at brukere kjenner ontologien og konstruere sine søkeargumenter deretter. Skal man utvide fritekst-søket med emner og assosiasjoner, må det lages et grensesnitt som skjermer brukere for ontologiske spørsmål og vurderinger.

Det er en relativt lav terskel for å ta i bruk emnekart. De kan utvikles og bygges opp over tid. Til å begynne med kan emnekartet bak *Lovdata 3000*⁴ bygges på dokumenter og henvisninger, det er mulig å komme igang relativt fort.

Hvordan systemet bør utformes vil være opp til teknologer, informatikere og grensesnitteksperter å avgjøre. Jeg tror nok at systemet må forbli WWW-basert. Tendensen er at stadig flere av applikasjoner flyttes vekk fra brukernes personlige datamaskiner og gjøres om til WWW-tjenester. Denne trenden vil nok bare fortsette.

Skal man kunne gå videre, må man velge en vei. Dersom veien man velger fører til en vegg og prosjekt ikke går bra, har man vunnet en erfaring, ikke gjort et feilsteg. Neste gang man velger veien, har man blitt klokere.

Takk for meg.

³Det jobbes for tiden med en standard for et formelt spørrespråk for emnekart *ISO 18048: Topic Maps Query Language (TMQL)*. TMQL vil være for emnekart som SQL er for relasjonsdatabaser.

⁴Et tenkt fremtidig rettslig informasjonssystem.

Bibliografi

- [1] ISO/IEC 13250 Topic Maps. http://y12web2.y12.doe.gov/sgml/sc34/document/0322_files/iso13250-2nd-e%d-v2.pdf.
- [2] Jack M. Balkin. The Footnote. *Northwestern University Law Review*, (83), 1988/1989.
- [3] Jon Bing. *Rettslige kommunikasjonsprosesser*. Universistetsforlaget, 1982.
- [4] Jon Bing. Offentlighetslovens dokumentbegrep. *Lov og Rett*, (10):606–619, 2004.
- [5] David C. Blair og M. E. Maron. An evaluation of retrieval effectiveness for a full-text document-retrieval system. *Commun. ACM*, 28(3):289–299, 1985. ISSN 0001-0782.
- [6] Sergey Brin og Lawrence Page. The Anatomy of a Large-Scale Hypertextual Web Search Engine. <http://www-db.stanford.edu/pub/papers/google.pdf>.
- [7] Enid Campbell. *Legal research: Materials and Methods*. North Ryde, NSW, fjerde utgave, 1996.
- [8] Morten Daae. Lovdata. historie, lov- og forskriftssystem. *CompLex*, (7), 1991.
- [9] Torstein Eckhoff. *Rettskildelære*. Universistetsforlaget, femte utgave, 2001.
- [10] Thor Falkanger, redaktør. LOV, DOM OG BOK: *Festskrift til Sjur Brækhus 19. juni 1988*, side 45–55. Universistetsforlaget, 1988.
- [11] Lars Marius Garshol. The Linear Topic Map Notation. <http://www.ontopia.net/download/ltn.html>.

- [12] Lars Marius Garshol. Living with topic maps and RDF. <http://www.ontopia.net/topicmaps/materials/tmrdf.html>.
- [13] Vladimir Geroimenko og Caomei Chen, redaktører. *Visualizing the Semantic Web*. Springer-Verlag LONDON Limited, 2003.
- [14] Luuk Matthijssen. *Interfacing between Lawyers and Computers*. Kluwer Law International, 1999.
- [15] Steve Pepper. The TAO of Topic Maps. <http://www.ontopia.net/topicmaps/materials/tao.html>.
- [16] Martin Støren. Hvorfor vokser Norsk Lovtidend? *CompLex*, (2), 2003.
- [17] Colin Tapper. An experiment in the use of citation vectors in the area of legal data. *CompLex*, (9), 1982.

Tillegg A

En avgjørelse i XML-format

```
<ns0:forsendelse art="manuelt"
  xmlns:bt="http://computas/lovisa/ws/common/xsd"
  xmlns:ns0="http://computas/lovisa/ws/lovdata">
<sak nr="05/180675/HALO" embeteNavn="Hålogaland lagmannsrett">
<avgjoerelse nr="6659" dato="2006-01-03+01:00" type="Kjennelse">
  <kategorisering>
    jordskifte, fremmingsvedtak, alltidvarende bruksrett
  </kategorisering>
<sammendrag>
  Kjæremål over jordskifterettens vedtak om fremming av
  grensegangssak for alltidvarende bruksrett. Påstått
  saksbehandlingsfeil fordi den omtvistede veirett ikke var avklart
  før grensegangssaken ble fremmet, jf. jordskifteloven § 17 tredje
  ledd. Videre påstand om at saken falt utenfor området for
  jordskifteloven § 88. Kjæremålet førte ikke fram.
</sammendrag>
<dokument>1</dokument>
<part>
<aktoer rolle="Kjæremålsmotpart">
  <person>
    <navn>
      <fornavn>Karin Elisabeth</fornavn>
      <etternavn>Andresen</etternavn>
    </navn>
  </person>
</aktoer>
<processfulmektige rolle="Advokatfullmektig">
  <person>
    <navn>
      <fornavn>Cathrine</fornavn>
      <etternavn>Leistad</etternavn>
    </navn>
    <tittel>Advokatfullmektig</tittel>
  </person>
</processfulmektige>
<processfulmektige rolle="Prosessfullmektig">
  <person>
    <navn>
      <fornavn>Odd</fornavn>
      <etternavn>Enevold</etternavn>
    </navn>
    <tittel>Advokat</tittel>
  </person>
</processfulmektige>
```

```

    </person>
  </processfulmektige>
</part>
<part>
  <aktoer rolle="Kjørende part">
    <person>
      <navn>
        <fornavn>Øystein Ingolf</fornavn>
        <etternavn>Henriksen</etternavn>
      </navn>
    </person>
  </aktoer>
  <processfulmektige rolle="Advokatfullmektig">
    <person>
      <navn>
        <fornavn>Rune</fornavn>
        <etternavn>Karlsen</etternavn>
      </navn>
      <tittel>Advokatfullmektig</tittel>
    </person>
  </processfulmektige>
  <processfulmektige rolle="Prosessfullmektig">
    <person>
      <navn>
        <fornavn>Hans Alfred</fornavn>
        <etternavn>Hoffmann</etternavn>
      </navn>
      <tittel>Advokat</tittel>
    </person>
  </processfulmektige>
</part>
</avgjoerelse>
<dommer rolle="1. håndsdommer">
  <person>
    <navn>
      <fornavn>Trude</fornavn>
      <etternavn>Haugli</etternavn>
    </navn>
    <tittel>Kst. lagdommer</tittel>
  </person>
</dommer>
<dommer rolle="2. håndsdommer">
  <person>
    <navn>
      <fornavn>Arild O.</fornavn>
      <etternavn>Eidesen</etternavn>
    </navn>
    <tittel>Førstelagmann</tittel>
  </person>
</dommer>
<dommer rolle="3. håndsdommer">
  <person>
    <navn>
      <fornavn>Kjell Martin</fornavn>
      <etternavn>Haug</etternavn>
    </navn>
    <tittel>Lagdommer</tittel>
  </person>
</dommer>
</sak>
</ns0:forsendelse>

```

Tillegg B

En avgjørelse i Lovdatas Bokus-format

<h0>
<h1> Hålogaland lagmannsrett - Kjennelse .
<h2> 2006-01-03
<h2a> 2006-01-23
<h3> LH-2005-180675
<h4> Jordskifte , fremmingsvedtak , alltidvarende bruksrett .
<h5> Kjæremål over jordskifterettens vedtak om fremming av grensegangssak for alltidvarende bruksrett . Påstått saksbehandlingsfeil fordi den omtvistede veirett ikke var avklart før grensegangssaken ble fremmet , jf. jordskifteloven # 17 tredje ledd . Videre påstand om at saken falt utenfor området for jordskifteloven # 88 . Kjæremålet førte ikke fram .
<h98> ^LOV 1979-12-21 77 # 17^
<h98> ^LOV 1979-12-21 77 # 88^
<h99> ^LOV 1976-12-17 100^
<h99> ^LOV 1979-12-21 77 # 61^
<h6> Finnmark jordskifterett - Hålogaland lagmannsrett
<ref=LH-2005-180675>LH-2005-180675</ref> .
<h7> Øystein Ingolf Henriksen (advokatfullmektig Rune Karlsen og advokat Hans Alfred Hoffmann) mot Karin Elisabeth Andresen (advokatfullmektig Cathrine Leistad og advokat Odd Enevold) .
<h8> Førstelagmann Arild O. Eidesen . Lagdommer Kjell Martin Haug . Kst . lagdommer Trude Haugli .

<a> Kjæremålet gjelder Finnmark jordskifteretts vedtak om fremming av grensegangssak for alltidvarende bruksrett i sak 3/2005 - Høyvik i Hasvik kommune .

[...]

<a> Kjæremålet må etter dette forkastes . For lagmannsretten må kjæremålsmotparten tilkjennes saksomkostninger , som settes til 3.000 kroner .
<a> Kjennelsen er enstemmig .
<tf>
<z> <kb>Slutning :
<tf>
<a> Kjæremålet forkastes .
<a> I saksomkostninger for lagmannsretten betaler Øystein Ingolf

Henriksen til Karin Elisabeth Andresen 3.000 - tretusen kroner innen 2
- to - uker fra forkynnelsen av denne kjennelse.<ks>
<slutt>

Tillegg C

Bustadsoppføringslova i LTM-format

```
/* LTM-emnekart for buofl. */

/* Emnetyper */
[lov = "Lov"]
[kapittel = "Kapittel"]
[paragraf = "Paragraf"]

[ledd = "Ledd"]      /* brukes ikke */
[litra = "Litra"]   /* brukes ikke */

[departement = "Departement"] /* kunne ha vært en forekomst */
[lovtidend = "Lovtidend"]     /* ditto */

/* Emnetyper for emner som viser til buofl. */
[avgjørelse = "Avgjørelse"]
[dom : avgjørelse = "Dom"]
[kjennelse : kjennelse = "Kjennelse"]

/* Typer av assosiasjoner og roller */
[er-en-del-av = "Er en del av"]
[frem-tilbake = "Forrige/neste"]

[del = "Er en del av"]
[helhet = "Inneholder"]

[forrige = "Forrige"]
[neste = "Neste"]

[henv = "Henvisning"]
[red-henv = "Redaksjonell henvisning"]
[sist-endret = "Sist endret"]
[forberedt-av = "Forberedt av"]
[publisert = "Publisert"]

[endret = "Endret"]
[endret-av = "Endret av"]
[endrer = "Endrer"]
```

```

/* Typer av forekomster */
[dato = "Dato"]
[ikraftttredelse = "Ikrafttredelsesdato"]
[tekst = "Tekst"]

/* Buofl */
[lov-19970613-43 : lov = "Lov om avtaler med forbrukar om oppføring av ny ←
  bustad"
  %"http://websir.lovdato.no/lex/lov/nl/hl-19970613-043.html"]
{lov-19970613-43, dato, [[1997-06-13]]}
{lov-19970613-43, ikraftttredelse, [[1998-07-01]]}
{lov-19970613-43, tekst, "http://websir.lovdato.no/lex/lov/nl/hl←
  -19970613-043.html"}

publisert-i(lov-19970613-43, lt1-1997-14)
[lt1-1997-14 : lovtidend = "Norsk Lovtidend avd I nr. 14 1997"
  %"http://www.lovdato.no/ltavd/i1997-1-14.html"] /* finnes ikke! */

forberedt-av(lov-19970613-43 : lov, justis : departement)
[justis : departement = "Justis- og politidepartementet"
  @"http://odin.dep.no/jd/"]

sist-endret(lov-19970613-43 : endrer, lov-20030606-39 : endret-av)
[lov-20030606-39 : lov = "Lov om burettslag"
  %"http://websir.lovdato.no/lex/lov/nl/hl-20030606-039.html"]
{lov-20030606-39, dato, [[2003-06-06]]}
{lov-20030606-39, ikraftttredelse, [[2004-01-01]]}
{lov-20030606-39, tekst, "http://websir.lovdato.no/lex/lov/nl/hl←
  -20030606-039.html"}

/* Kapittel I */
[lov-19970613-43-kapI : kapittel = "Kap I - Verkeområde m.m. "
  %"http://websir.lovdato.no/lex/lov/nl/hl-19970613-043.html#map001"]
er-en-del-av(lov-19970613-43-kapI, lov-19970613-43)
{lov-19970613-43-kapI, tekst, "http://websir.lovdato.no/lex/lov/nl/hl←
  -19970613-043.html#map001"}

/* §1 */
[lov-19970613-43-par1 : paragraf = "§ 1 - Verkeområde"
  %"http://websir.lovdato.no/lex/lov/nl/hl-19970613-043.html#1"]
er-en-del-av(lov-19970613-43-par1, lov-19970613-43)
er-en-del-av(lov-19970613-43-par1, lov-19970613-43-kapI)
{lov-19970613-43-par1, tekst, "http://websir.lovdato.no/lex/lov/nl/hl←
  -19970613-043.html#1"}

/* §1a */
[lov-19970613-43-par1a : paragraf = "§ 1a - Burettslagbustader"
  %"http://websir.lovdato.no/lex/lov/nl/hl-19970613-043.html#1a"]
er-en-del-av(lov-19970613-43-par1a, lov-19970613-43)
er-en-del-av(lov-19970613-43-par1a, lov-19970613-43-kapI)
{lov-19970613-43-par1a, tekst, "http://websir.lovdato.no/lex/lov/nl/hl←
  -19970613-043.html#1a"}

/* [...] */

/* Kapittel XI */
[lov-19970613-43-kapXI : kapittel = "Kap XI - Endringer i andre lover. ←
  Ikrafttsetjing "
  %"http://websir.lovdato.no/lex/lov/nl/hl-19970613-043.html#map011"]
er-en-del-av(lov-19970613-43-kapXI, lov-19970613-43)

```

```

{lov-19970613-43-kapXI, tekst, "http://websir.lovdata.no/lex/lov/nl/hl↵
-19970613-043.html#map011"}

/* §65 */
[lov-19970613-43-par65 : paragraf = "§ 65 - Endringar i andre lover"
%"http://websir.lovdata.no/lex/lov/nl/hl-19970613-043.html#65"]
er-en-del-av(lov-19970613-43-par65, lov-19970613-43)
er-en-del-av(lov-19970613-43-par65, lov-19970613-43-kapXI)
{lov-19970613-43-par65, tekst, "http://websir.lovdata.no/lex/lov/nl/hl↵
-19970613-043.html#65"}

/* §66 */
[lov-19970613-43-par66 : paragraf = "§ 66 - Ikraftsetjing"
%"http://websir.lovdata.no/lex/lov/nl/hl-19970613-043.html#66"]
er-en-del-av(lov-19970613-43-par66, lov-19970613-43)
er-en-del-av(lov-19970613-43-par66, lov-19970613-43-kapXI)
{lov-19970613-43-par66, tekst, "http://websir.lovdata.no/lex/lov/nl/hl↵
-19970613-043.html#66"}

/* Bla frem og tilbake i kapitlene */
frem-tilbake(lov-19970613-43-kapI : forrige,
lov-19970613-43-kapII : neste)
frem-tilbake(lov-19970613-43-kapII : forrige,
lov-19970613-43-kapIII : neste)
frem-tilbake(lov-19970613-43-kapIII : forrige,
lov-19970613-43-kapIV : neste)
frem-tilbake(lov-19970613-43-kapIV : forrige,
lov-19970613-43-kapV : neste)
frem-tilbake(lov-19970613-43-kapV : forrige,
lov-19970613-43-kapVI : neste)
frem-tilbake(lov-19970613-43-kapVI : forrige,
lov-19970613-43-kapVII : neste)
frem-tilbake(lov-19970613-43-kapVII : forrige,
lov-19970613-43-kapVIII : neste)
frem-tilbake(lov-19970613-43-kapVIII : forrige,
lov-19970613-43-kapIX : neste)
frem-tilbake(lov-19970613-43-kapIX : forrige,
lov-19970613-43-kapX : neste)
frem-tilbake(lov-19970613-43-kapX : forrige,
lov-19970613-43-kapXI : neste)

/* Bla frem og tilbake i paragrafer */
frem-tilbake(lov-19970613-43-par1 : forrige,
lov-19970613-43-par1a : neste)
frem-tilbake(lov-19970613-43-par1a : forrige,
lov-19970613-43-par2 : neste)
/* [...] */
frem-tilbake(lov-19970613-43-par65 : forrige,
lov-19970613-43-par66 : neste)

/* Henvisninger */

/* HRSIV */
[hr-2004-02014a : dom = "Rt-2004-1887"
%"http://websir.lovdata.no/lex/avg/hrsiv/hr-2004-02014-a.html"]
henv(hr-2004-02014a, lov-19970613-43-par36)
henv(hr-2004-02014a, lov-19970613-43-kapIV)
henv(hr-2004-02014a, lov-19970613-43)

[hr-2002-01114a : dom = "Rt-2003-1312"
%"http://websir.lovdata.no/lex/avg/hrsiv/hr-2002-01114a.html"]

```

```

henv(hr-2002-01114a, lov-19970613-43-par43)
henv(hr-2004-02014a, lov-19970613-43-kapV)
henv(hr-2004-02014a, lov-19970613-43)

[hr-2002-00359a : dom = "Rt-2003-612"
 %"http://websir.lovdata.no/lex/avg/hrsiv/hr-2002-00359a.html"]
henv(hr-2002-00359a, lov-19970613-43-par26)
henv(hr-2004-02014a, lov-19970613-43-kapIV)
henv(hr-2004-02014a, lov-19970613-43)

[hr-2002-01028a : dom = "Rt-2003-486"
 %"http://websir.lovdata.no/lex/avg/hrsiv/hr-2002-01028a.html"]
henv(hr-2002-01028a, lov-19970613-43-par36)
henv(hr-2004-02014a, lov-19970613-43-kapIV)
henv(hr-2004-02014a, lov-19970613-43)

[rt1999-0768-00042a : dom = "Rt-1999-768"
 %"http://websir.lovdata.no/lex/avg/hrsiv/rt1999-0768-00042a.html"]
henv(rt1999-0768-00042a, lov-19970613-43-par66)
henv(hr-2004-02014a, lov-19970613-43-kapXI)
henv(hr-2004-02014a, lov-19970613-43)

[rt1998-0656-00035b : dom = "Rt-1998-656"
 %"http://websir.lovdata.no/lex/avg/hrsiv/rt1998-0656-00035b.html"]
henv(rt1998-0656-00035b, lov-19970613-43-par37)
henv(hr-2004-02014a, lov-19970613-43-kapIV)
henv(hr-2004-02014a, lov-19970613-43)

/* HKSIV */
[hr-2005-01382-u : kjennelse = "Rt-2005-1046"
 %"http://websir.lovdata.no/lex/avg/hksiv/hr-2005-01382-u.html"]
henv(hr-2005-01382-u, lov-19970613-43)

/* HUSIV */
[hr-2002-01249 : kjennelse = "HR-2002-01249"
 %"http://websir.lovdata.no/lex/avg/husiv/hr-2002-01249.html"]
henv(hr-2002-01249, lov-19970613-43-par48)
henv(hr-2004-02014a, lov-19970613-43-kapV)
henv(hr-2004-02014a, lov-19970613-43)

```