

*Sømløs kunnskap – om bruk av emnekart gir en innføring i emnekartstandard og de viktigste elementene i emnekart. Sømløs kunnskap peker på problematiske aspekter ved sammenfletting og et praktisk forsøk på sammenfletting illustrerer til slutt enkelte navigasjonsmuligheter som kan realiseres med emnekart.*

# SØMLØS KUNNSKAP

- om bruk av emnekart

av Ingvild Kongsbakk

Originaltittel: Emnekart og tesauri – gjenfinning i nettverk og hierarkier.

Først utgitt: Diplomoppgave ved Høgskolen i Oslo, Avdeling for journalistikk, bibliotek- og informasjonsfag, Oslo 2003.

Forfatter: Ingvild Kongsbakk. Hun avsluttet sivilbibliotekarstudiet ved Høgskolen i Oslo i 2003 og er nå tilsatt som redaktør for Bibsys emneportal.

#### ABM-UTVIKLING

POSTBOKS 8145 DEP  
N-0033 OSLO

TELEFON: 23 11 75 00  
TELEFAKS: 23 11 75 01

post@abm-utvikling.no  
www.abm-utvikling.no

FORSIDE: Brainmap/Unn Bjørge  
TRYKK: PDC-Tangen AS

ISSN 1503-5972  
ISBN 82-8105-009-8

ABM-utvikling 2004

# INNHold

Forord . . .	2
STEVE PEPPER	
Sømløs kunnskap med emnekart . . . . .	4
<b>1 Innledning</b> . . . . .	8
1.1 BAKGRUNN . . . . .	8
1.2 PROBLEMSTILLING . . . . .	8
1.3 FRAMGANGSMÅTE OG OPPGAVENS STRUKTUR . . . . .	9
<b>2 Teoretisk bakgrunn.</b> . . . . .	10
2.1 KUNNSKAPSORGANISASJON . . . . .	10
2.2 WORLD WIDE WEB . . . . .	12
2.3 SEMANTIC WEB . . . . .	12
<b>3 Emnekart</b> . . . . .	14
3.1 FORMÅL OG DEFINISJON . . . . .	14
3.2 HISTORIKK . . . . .	14
3.3 TEMA OG EMNE . . . . .	15
3.4 ASSOSIASJONER . . . . .	17
3.5 FOREKOMSTER . . . . .	18
3.6 PERSPEKTIV . . . . .	18
3.7 LAGDELINGEN I EMNEKART . . . . .	18
3.8 BEGRENSNINGER . . . . .	18
3.9 EMNEKART OG SEMANTIC WEB . . . . .	19
3.10 EMNEKART OG KUNNSKAPSORGANISASJON . . . . .	19
<b>4 Tesauri . . . . .</b>	22
4.1 TESAUROSRELASJONER . . . . .	22
<b>5 Emnekart og tesauri</b> . . . . .	26
5.1 TEORETISK SAMMENLIGNING AV EMNEKART OG TESAUORI . . . . .	26
5.2 EMNEKART I PRAKSIS . . . . .	27
5.3 RESULTATER FRA ANALYSEN . . . . .	29
5.4 PROBLEMENE I FORHOLD TIL STANDARDENE . . . . .	30
<b>6 Sammenfletting av emnekart</b> . . . . .	32
6.1 SAMMENFLETNING . . . . .	32
6.2 METODER FOR SAMMENFLETNING . . . . .	34
6.3 KONSISTENTE EMNEKART . . . . .	34
6.4 TEORETISK DRØFTING AV SAMMENFLETNING . . . . .	35
6.5 SAMMENFLETNING I PRAKSIS . . . . .	36
6.6 RESULTATET AV SAMMENFLETNINGEN . . . . .	37
<b>7 Oppsummering.</b> . . . . .	40
<b>Litteraturliste</b> . . . . .	42

# FORORD

Emnekart – eller engelsk Topic map – er det heteste begrepet når det gjelder organisering av fulltekst kunnskap i dag.

Vi velger å bruke begrepet sømløs kunnskap. Sømløs-begrepet er hentet fra datateknologien som en betegnelse på integrerte dataprogrammer. I biblioteksektoren brukes begrepet det sømløse bibliotek som et mål. Det vil si at du som bruker kan benytte deg bedre av alle tjenestene uten å kjenne grensene mellom de ulike bibliotekslagene. På samme måte kan emnekartet gi deg en mulighet til å utnytte innholdet bedre enn det den tradisjonelle databaseorganiseringa hittil har gjort, uten at du må plages med sømmene som binder stoffet sammen.

Internasjonalt er Norge i bresjen når det gjelder emnekart. Steve Pepper i Ontopia er leder av standardiseringsarbeidet for emnekart. Verdens første emnekartkonferanse ble arrangert i Oslo i 2002. Emnekart er bl.a. tatt i bruk i organiseringen av nettstedene forbrukerportalen.no og forskning.no. I løpet av 2004 lanseres en ny versjon av kulturnett.no som også bygger på emnekart. ABM-utvikling, ved Norsk Digitalt Bibliotek, har gitt støtte til et prosjekt i regi av Bibsys for å se på bruk av emnekart for organisering av bibliografiske data.

Det er ingen tvil om at interessen for emnekart er økende. Ifølge Steve Pepper er Ingvild Kongsbakks diplomoppgave det beste som hittil er skrevet på norsk om emnekart. Det er derfor

naturlig for ABM-utvikling å publisere Kongsbakks diplomoppgave for å bidra til å spre kunnskap om emnekart, og å fremme diskusjon og oppmerksomhet om dette.

Vi synes det er spesielt spennende å kunne bidra til å spre gode studentarbeider fra abm-utdanningsfeltet, som tilfellet er med denne diplomoppgaven fra Sivilbibliotekarstudiet ved Høgskolen i Oslo. Vi takker Ingvild Kongsbakk for at vi får lov til å publisere oppgaven og Steve Pepper for en ekstra innledning om emnekart.

Oslo, april 2004

*Jon Birger Østby*  
DIREKTØR

*Lars Egeland*  
AVDELINGS-DIREKTØR

## Forfatterens forord

Stian Danenberger, Creuna, og Lars Marius Garshol, Ontopia, skal ha takk som tok seg tid til å svare på spørsmål, og viste interesse for oppgaven. En takk går også til Jan Bruusgaard, SSB, som villig diskuterte emnekart, og som lot meg få et innblikk i innføring av emnekart hos Statistisk sentralbyrå.

En stor takk til veileder Ragnar Nordlie som utviste stor tålmodighet da arbeidet med oppgaven holdt på å stagnere. Takk for oppmuntring og faglige innspill.

Mine medstudenter fortjener også å bli nevnt. Marit Kristine Ådland og Marte Brenne skal vite at jeg har satt pris på de mange hyggelige middagene vi har hatt på Fyrhuset. Takk til Torunn Helene Fredriksen og Mona Thomassen som har gitt meg støtte, oppmuntring og gode råd underveis i prosessen.

Familien fortjener en stor takk for å ha holdt ut med meg i denne perioden. Det samme gjelder også venner som har støttet meg etter beste evne.

Spesielt takk til bestevenninne Carina Stjernholm som har hjulpet meg med arbeidet, og kommet med gode råd og forslag. Takk for oppmuntring, støtte og lange telefonsamtaler.

Oslo, november 2003

*Ingvild Kongsbakk*

STEVE PEPPER

# SØMLØS KUNNSKAP MED EMNEKART

*Da Denis Diderot og Jean d’Alembert skapte Den store encyclopedien i Frankrike på 1700-tallet hadde de som mål å samle all sin tids kunnskap i dette verket. Deres erklærte hensikt var å bryte elitenes kunnskapsmonopol og bringe kunnskapen til folkeflertallet, slik at det kunne bane vei for et folkestyre.*

Slik begynner en kronikk skrevet av undertegnede og tidligere leksikonredaktør Pål Steigan i *Aftenposten* den 10. desember 2001.<sup>1</sup>

Bakgrunnen for kronikken var en anbudskonkurranse som Kulturdepartementet hadde utlyst våren 2001 om å skape en nasjonal kunnskapsbase. To grupperinger nådde finalen men begge anbudene ble forkastet og statsråd Kristin Clemet oppfordret til å “tenke fritt og uavhengig av dagens teknologi og kommersielle begrensninger.”

## Nasjonal kunnskapsbase

Kronikken i *Aftenposten* tok til orde for en “åpen, demokratisk nasjonal kunnskapsbase” basert på en ny internasjonal standard for strukturering av kunnskap, ISO 13250 Topic Maps, nå bedre kjent i Norge som *emnekart*.

Forfatterne utmalte en visjon som blant annet gikk ut på følgende:

- En kunnskapsbase som er like lett å finne frem i som en bok med et godt stikkordregister, og som kan brukes enten man snakker norsk, samisk, urdu eller et annet språk.
- En kunnskapsbase som tilpasser seg din profil, så som aldersgruppe, kultu-

rell bakgrunn, interesser, osv., og hjelper deg å luke ut den informasjonen du ikke er interessert i.

- En kunnskapsbase som du kan slå opp i via en WAP-telefon for å avgjøre en livlig pubdiskusjon om hvem som scoret målene i den berømte kampen mellom Norge og England, og hvilket år det skjedde. Eller når og hvor operaen Tosca først ble oppført; eller for den saks skyld, hvem det var som skrev librettoen, og hvilke andre libretti han sto bak.
- En kunnskapsbase som har plass til innhold fra både Kunnskapsforlaget, Cappelen, og fra mange andre aktører i tillegg; fra store offentlige og kommersielle organisasjoner som NRK, Schibsted, Forskningsrådet og Nasjonalbiblioteket til universiteter og skoler, interesseorganisasjoner, minoritetsgrupper – ja, til og med enkeltpersoner.
- En kunnskapsbase skrevet av både profesjonelle leksikografer og andre fagfolk, men også av den norske befolkning.
- En kunnskapsbase som inneholder mange ulike syn på én og samme sak, ikke bare én halvoffisiell, “objektiv” sannhet. En kunnskapsbase som gjør det mulig for utilsørt subjektive meninger om EU, gasskraft og krigen i Afghanistan å komme til uttrykk, men som samtidig tillater brukeren å skille mellom disse, vite hvem som står for hvilke uttalelser, og velge hvem hun vil stole på.
- En kunnskapsbase som gjør det mulig

for Thea og de andre i 5A på Marienlyst å lage sin egen lille “kunnskapspar-sell” – f.eks. om miljøeffekten av hamburgerindustriens rovdrift i Amazonas – og hekte den på Nasjonal Kunnskapsbase (med hilsen fra 5A), slik at venninnen Nina i Trondheim og fetteren Tom i England kan se den (på hhv. norsk og engelsk), gjenbruke den, og føye til sine egne tanker og koblinger.

- En kunnskapsbase som blir så verdifull at den kan lisensieres til de som har råd til det, også utenfor Norge, og samtidig så kostnadseffektiv å vedlikeholde at Norge har råd til å gi den bort som u-hjelp.

Fra et teknologisk synspunkt var det fullt mulig å realisere denne visjonen allerede den gangen, men selv om forslaget vakte begeistring i brede kretser var ikke tiden inne politisk. Teknologien som skulle underbygge visjonen var ennå fersk og lite utprøvd. Emnekart var blitt vedtatt som internasjonal standard bare ett år i forveien og man hadde ingen større prosjekter basert på denne teknologien å vise til. Forslaget ble dermed liggende på skrivebordet.

## Emnekartportaler i Norge

I dag har situasjonen forandret seg. I de siste tre årene har vi sett vellykkede lanseringer av et tosifret antall offentlige og halvoffentlige nettsteder basert på emnekart. Visjonen om en åpen kunnskapsbase er i ferd med å bli realisert uten departementets hjelp.

Denne utviklingen begynte med ITU (Forsknings- og Kompetanse-nettverk for IT i Utdanning) som sommeren 2000 satte i gang en prosess for å skape et nytt nettsted, *itu.no*. ITU var ute etter noe litt utenom det vanlige, et nettsted som var “innovativt, kreativt, sprekt, seriøst, nettverksskapende, levende, dynamisk ... og *annerledes*”. De engasjerte konsulentselskapet Creuna og prosjektleder Stian Danenbarger.

Under behovsanalysen kom det frem at ITU så på feltet sitt som et komplekst nettverk av mennesker, prosjekter, teknologier, metoder, teorier, verktøy, aktiviteter og temaer som var vevet tett sammen med hverandre gjennom relasjoner av typen “er en del av”, “består av”, “ligner på”, “tilhører samme kategori”, osv. Man ønsket et nettsted som gjenspeilet dette og meislet ut et par bærende ideer for et nettverks-konsept:

- Relasjoner mellom objekter og ulike grupperinger av objekter gir brukerne mange veier inn til samme innhold, og det stimulerer til å utforske på tvers av innholdet i tjenesten.
  - Visualisering av dette nettverket skal gi brukeren en konseptuell modell av nettverket og gi en følelse av å befinne seg i et “relasjons-rom”.
- Man ønsket med andre ord å la selve strukturen på nettstedet beskrive kunnskapsuniverset til ITU og samtidig bruke denne strukturen som et gjenfinnings- og navigasjonsverktøy for innholdet på nettstedet. Dette var

en besnærende tanke som Stian lot seg fange av, men hvordan skulle man få det til i praksis? Akkurat da ble XML *Topic Maps* (XTM) annonsert. Stian så notisen og det gikk et lys opp for ham.

Notisen beskrev et “standardformat” for “utveksling av informasjon” om “emner og relasjoner mellom emner” og et “flerdimensjonelt rom”, der relasjonene “beskriver stien fra ett emne til et annet.” Litt undersøkning på nettet førte Stian frem til Ontopia, et lite norsk firma som var blant pionerene på emnekart-basert teknologi. Ontopia og Creuna inngikk et samarbeid om å skape et publiseringsverktøy med støtte for emnekart basert på “åpen kildekode”-systemet Zope og resultatet ble ZTM, Zope Topic Maps. ZTM ga ITU akkurat det nettstedet de ønsket seg og prosjektet ble en stor suksess.

*itu.no* ble etterfulgt av *forskning.no*, et enda mer ambisiøst prosjekt der det blant annet var behov for to parallelle men tett sammenfildrede klassifikasjonssystemer som avspeilet hvordan forskningens verden oppfattes av henholdsvis målgruppen (unge voksne) og innholdsleverandørene (fagfolk). Man ønsket seg en løsning som “kombinerte akademiske faghierarkier, vanlige folks forestillinger og redaksjonell tenkning i ett konsept”.

Det ble igjen et samarbeid mellom Creuna og Ontopia der førstnevnte sto for den tekniske løsningen (basert på ZTM), konseptutvikling og design, og sistnevnte var ansvarlig for utvikling av ontologien. Også dette nettstedet

ble en suksess. Jon Solberg hos Norges Forskningsråd “ble nesten svimmel over hvor mye mer tilgjengelige artiklene var” og redaktør Steinar Q. Andersen kommenterte hvordan “det blir veldig lett fordi *det er sånn hjernen fungerer*. Man tenker assosiativt.”

I løpet av de siste par årene har emnekartportaler kommet som perler på en snor i Norge: Forbrukerportalen, Matportalen, Nysgjerrigper og hoyre.no er bare noen av dem. Sistnevnte er spesielt interessant, ikke bare fordi nettstedet tilhører partiet til statsråd Kristin Clemet, men fordi det består av 450 portaler (én pr. lokallag) som er drevet av ett felles emnekart (og dermed gir en forsmak på hvordan en nasjonal kunnskapsbase kunne fungere).

## Mot sømløs kunnskap

Foreløpig eksisterer disse portalene som atskilte øyer av kunnskap, men etter hvert som de blir flere oppstår det stadig mer overlapping. På *forskning.no* finnes det en side viet til temaet “genmodifisert mat” med siste nytt fra forskningen. På *forbrukerportalen.no* finner man også en side om genmodifisert mat, denne gangen med informasjon om den nyeste lovgiving fra norsk og europeisk side. Hos *matportalen.no* snakkes det derimot om “genmodifiserte matvarer” – det samme temaet, bare et annet navn.

Tenk om man kunne kople disse sidene sammen automatisk og gi brukeren mulighet til å navigere sømløs fra den ene kunnskapsbasen til den andre!

Eller enda mer spennende: På basis av en spørring opp mot en virtuell nasjonal kunnskapsbase (“Finn stoff om genmodifisert mat”) kunne man få satt sammen en side som kombinerte innhold fra alle disse nettstedene hel-automatisk!

Med utgangspunkt i emnekartstandarden er denne visjonen om *sømløs kunnskap* nå i ferd med å bli realisert. I det dette ABM-skriftet går i trykken blir siste hånd lagt om en protokoll som vil gjøre det mulig for ZTM- og andre emnekartbaserte nettsteder å kommunisere, utveksle innhold, og sys sammen til en distribuerte kunnskapsbase. I neste omgang blir det til og med mulig for nettsteder som ikke er basert på emnekart å ta del i de nye mulighetene; i hvert fall et stykke på vei.

Hva er det med emnekart som gjør alt dette mulig?

Svaret er en helt unik kombinasjon av egenskaper. Emnekart er basert på en meget fleksibel og kraftig data-modell som er i stand til å favne om alle eksisterende kunnskapsorganisasjonsmodeller (inkludert tesaurus, glossar, og stikkordregister), og viktige informasjonsstrukturer som relasjonelle data, hierarkiske XML-dokumenter, og meta-data i form av Dublin Core og RDF.<sup>2</sup> Dermed kan emnekart fungerer som den overgripende strukturen som samler ulike typer data og kunnskap innenfor én modell.

Det at emnekart er en internasjonal standard, ISO 13250:2003 *Topic Maps*, gir

teknologien en troverdighet som gjør den verdt å investere tid og penger i: Det er en teknologi som man kan stole på og som ikke eies og kontrolleres av kommersielle interesser.

#### Publiserte temaer

Men det viktigste av alt, i hvert fall når det gjelder å oppnå sømløs kunnskap, er at emnekart har en modell for *identitet* som er helt unik.

Begrepet “identitet” er meget sentralt for at man skal kunne vite når man snakker om samme sak. Uten dette blir alle forsøk på å oppnå sømløs kunnskap forgjeves. For å ta et enkelt eksempel: Hvordan skal vi kunne vite at “genmodifisert mat” dreier seg om det samme som “genmodifiserte matvarer” (eller “genetically modified food”, eller “GM food” for den saks skyld)?

Kontrollerte vokabularer er ikke svaret (i hvert fall ikke i vanlig forstand). Vi snakker om en distribuert kunnskapsbase hvis omfang minst er nasjonalt og potensielt globalt. I en slik setting er det ikke mulig å legge restriksjoner på hvordan folk navngir ting; synonymi og polysemi er noe man må leve med. For å realisere visjonen om sømløs kunnskap trenger vi derfor intet mindre enn en global tesaurus, basert på unike, globale identifikatorer, som tillater vilkårlige navn for det samme emnet, og som kan fungere som en slags “semantisk superhighway” for sammenkobling av informasjon og kunnskap.

En umulig drøm? Kanskje det, i

hvert fall i all sin bredde. Men i mer begrenset omfang er dette fullt mulig å realisere, spesielt fordi mekanismen for bygging av en slik tesaurus allerede er en del av emnekartstandarden.

Kjernen i denne mekanismen er “published subjects” (eller “publiserte temaer”). Et publisert tema er et tema med en *temaidentifikator* i form av en URL som er blitt offentliggjort med tanke på å bli brukt av flere. Ideelt sett bør denne URLen også fungere som adresse til et dokument (kalt en *temaindikator*). Kombinasjonen av identifikator og indikator gir en tosidig identifikasjonsmekanisme som kan brukes av både maskiner og mennesker: Maskiner kan sammenligne identifikatorer; mennesker kan inspisere indikatoren.

Et viktig aspekt ved denne mekanismen er at den virker *nedefra og opp* og ikke ovenfra og ned. Alle står fritt til å definere identifikatorer. Finner man ikke en identifikator man trenger hos andre som man stoler på, kan man lage sin egen. Dersom den publiseres vil den kunne brukes av andre. En evolusjonsaktig utvelgelsesprosess (“den troverdigstes overlevelse”) vil føre til at det etter hvert vokser frem autoritative identifikatorer som dekker stadig større deler av menneskehetens felles kunnskapsunivers.<sup>3</sup>

#### kulturnett.no

Et stort skritt fremover i så henseende blir tatt når det nye *kulturnett.no* lanseres om kort tid. Denne portalen represen-

terer den hittil mest ambisiøse satsing på emnekart i Norge og vil blant annet definere flere tusen unike identifikatorer av typen <http://psi.kulturnett.no/museum/munch-museet> og lignende.

*kulturnett.no* vil bli en viktig milepæl i utviklingen frem mot en nasjonal kunnskapsbase og mot sømløs kunnskap i sin alminnelighet (som selvsagt ikke er begrenset til offentlig sektor eller nasjonale grenser). Det samme kan sies om Ingvild Kongsbakks meget leseverdige avhandling om temaet *Emnekart kontra tesaurus*, som gir en utmerket introduksjon til emnekart til en yrkesgruppe, bibliotekarer, som nå bør kjenne sin besøkestid.

Convenor SC 34/WG 3 (Topic Maps)

Oslo, april 2004

Steve Pepper

<sup>1</sup> “En åpen, demokratisk nasjonal kunnskapsbase”. Aftenposten 10. desember 2001, <http://www.aftenposten.no/meninger/kronikker/article.jhtml?articleID=242930>

<sup>2</sup> For en utførlig behandling av dette se Lars Marius Garshols artikkel, Metadata? Thesauri? Taxonomies? Topic Maps!, <http://www.ontopia.net/topicmaps/materials/tm-vs-thesauri.html>

<sup>3</sup> For mer informasjon om publiserte temaer se Published Subjects: Introduction and Basic Requirements, <http://www.ontopia.net/topicmaps/materials/pubsubj-gentle-intro.htm>

# INNLEDNING

## 1.1 Bakgrunn

Emnekart er en teknologi som muliggjør beskrivelse og framvisning av et kunnskapsområde. Dette blir gjort ved at emnekart beskriver kunnskapsstrukturer og assosierer disse med ressurser.

I 2002 ble verdens første emnekartkonferanse arrangert i Oslo. Norske firmaer har vært med på å utvikle standarden for emnekart, og har også stått for markedsføringen av standarden her i landet. Norge er også det landet i verden hvor flest web-redaksjoner har tatt i bruk emnekart (Vogt 2002). Etter emnekartkonferansen har det vært arrangert ulike seminarer hvor temaet har vært emnekart, og det jobbes nå for å opprette en egen emnekartgruppe i Norge ([www.emnekart.no](http://www.emnekart.no)).

Biblioteksektoren har også vist interesse for emnekart. Våren 2003 arrangerte BRODD et emnekartseminar rettet mot bibliotekarer. Her ble emnekart tatt opp i forhold til bibliotekfag, og et par av foredragene tok for seg emnekart kontra tesauri (BRODD 2003). Dette har også vært et tema på e-postlista [biblioteknorge.no](http://biblioteknorge.no) hvor diskusjonen har dreid seg om emnekart i forhold til tradisjonelle verktøy innen kunnskapsorganisasjon.

En sentral person i arbeidet med emnekart, Steve Pepper, har sammen med Pål Steigan i kronikken *En åpen, demokratisk nasjonal kunnskapsbase* (Steigan 2001) foreslått emnekart som løsning for en nasjonal kunnskapsbase. Hovedargumentet for dette er mulighe-

ten for å flette sammen flere emnekart. Det betyr at emnekart som dekker avgrensede samlinger, områder eller fag, kan kobles sammen slik at de danner et samlet emnekart over alle områdene.

Arbeidsgruppa for Norsk digitalt bibliotek tar i sin innstilling opp problemstillingen med sammenkobling av ulike vokabularer (Norsk digitalt bibliotek 2003). I innstillingen heter det at det må undersøkes nærmere hvordan man kan få til samsøk i flere samlinger. Arbeidsgruppa tar opp flere utfordringer rundt dette og sier at “mangelen på felles systemer for emnebeskrivende metadata (klassifikasjons- og emneords-systemer) er en særlig utfordring.”

## 1.2 Problemstilling

Pepper og Steigan skriver videre i sin kronikk at emnekart består av en “kunnskapsvev av emner og assosiasjoner – et slags semantisk nettverk – som er meget lett å navigere i fordi den avspeiler den assosiative måten vi mennesker tenker på.” Emnekart er altså et nettverk bestående av relasjoner. Utsagn fra emnekarthold som sier at “kunnskap lar seg ikke organisere i et enkelt hierarki” (Vogt 2002) gjør at denne oppgaven vil gå nærmere inn på denne nettverksstrukturen. Det blir også hevdet at hierarkiske kategorier ikke holder mål, og det vil derfor være interessant å sammenligne emnekartets nettverksstruktur med den hierarkiske strukturen i tesauri.

På bakgrunn av dette vil det være interessant å gå nærmere inn i emne-

kartstandarden for å få en bedre forståelse av selve modellen og funksjonene som ligger i den. Steve Pepper hevder at emnekart kan løse problemer knyttet til gjenfinning og gjenbruk av informasjon, samt problemer knyttet til sammenkobling av informasjon (Vogt 2002). Sammenfletting av emnekart vil derfor bli spesielt grundig behandlet. Med utgangspunkt i det som har vært sagt og skrevet om emnekart er det interessant å utforske muligheten for å bruke sammenfletting av emnekart til sammenkobling av vokabularer.

Opgaven har på bakgrunn av dette følgende problemstillinger:

- Hvordan skiller emnekart seg fra tesauri?
  - Hva er de vesentligste forskjellene?
  - Hvordan behandles relasjoner i henholdsvis emnekart og tesauri?
  - Hva er forholdet mellom nettverk og hierarkier?
- Hvordan fungerer emnekart som navigasjonshjelpemiddel?
- Kan emnekart brukes til å koble sammen ulike vokabularer?
  - Hvilke problemer finnes det i sammenfletting av emnekart?
  - Hva skyldes disse problemene? Hva kan være løsningen?
  - Tar emnekartstandard høyde for å løse disse problemene, eller er det ikke intensjonen at den skal gjøre det?

## 1.3 Framgangsmåte og oppgavens struktur

Tilnærmingen til emnekartmodellen er en forholdsvis grundig gjennomgang av emnekartstandard. Oppgaven går derfor nøye gjennom de viktigste elementene og funksjonene i emnekart. Gjennomgangen og beskrivelsen av emnekartstandard danner grunnlaget for resten av oppgaven, samt grunnlaget for problematiseringen av enkelte forhold ved emnekart. Dette er forhold knyttet til relasjoner, nettverksstruktur og sammenfletting av emnekart. Sammenfletting blir i tillegg behandlet i et eget kapittel.

For å få en bedre forståelse av hvordan emnekart fungerer i praksis, inneholder oppgaven en praktisk del. Denne delen tar utgangspunkt i to nettsteder som har tatt i bruk emnekart. Utvalgte deler av nettstedene blir analysert og er med på å danne et bilde av hvordan emner står i relasjon til hverandre i et emnekart. Analysen peker på problematiske forhold ved implementering av emnekart.

De to nettstedene blir også brukt i den praktiske delen som fokuserer på sammenfletting av emnekart. Utvalgte områder fra nettstedene danner grunnlaget for to emnekart som blir flettet sammen. Formålet med dette er å se hvordan sammenfletting fungerer i praksis. Det blir lagt vekt på å undersøke om forhold som ble beskrevet tidligere i oppgaven nå forsterkes eller forblir som de var, eller om nye momen-

ter dukker opp. Deretter blir det sett på hvordan emnekartstandard legger opp til å håndtere disse momentene.

Emnekart og tesauri blir sammenlignet i oppgaven. Derfor inneholder oppgaven en forholdsvis kort gjennomgang av tesauri. På den måten blir det mulig å se emnekart i lys av dette verktøyet. Den teoretiske gjennomgangen av henholdsvis emnekart og tesauri er grunnlaget for sammenligningen av de to modellene.

For å kunne sammenligne emnekart og tesauri er det nødvendig å sette dem inn i en sammenheng. Dette innebærer at formålet for modellene og tradisjonene de har sitt utspring i blir gjennomgått. For å kunne si noe om forskjellene mellom emnekart og tesauri er det også nødvendig å se nærmere på henholdsvis nettverksstrukturer og hierarkiske strukturer.

Funnene fra analysen av de emnekartbaserte nettstedene blir holdt opp mot prinsippene for tesauruskonstruksjon. Dette blir gjort med den hensikt å se hvordan de ville blitt i en hierarkisk ordning. Målet med dette er å tydeliggjøre forskjellen mellom nettverksstrukturer og hierarkiske strukturer.

På bakgrunn av dette blir oppgavens struktur som følger:

- Kapittel 2 inneholder en avklaring rundt emnebegrepet, og tar for seg kontrollerte vokabularer og gjenfinning, World Wide Web og navigering, samt at det gir en gjennomgang av hierarkier og nettverk. Semantic web blir også behandlet.

- Kapittel 3 tar for seg det teoretiske grunnlaget for emnekart, og setter modellen inn i en sammenheng. Deretter følger en grundig presentasjon av elementene i emnekart og emnekartstandard. Det blir gitt et praktisk eksempel på bruk av emnekart i to nettsteder.
- Kapittel 4 inneholder en gjennomgang av tesauri og tesaurusrelasjoner.
- Kapittel 5 er en teoretisk og praktisk sammenligning av emnekart og tesauri, og bygger på de foregående kapitlene.
- Kapittel 6 behandler sammenfletting av emnekart.
- Kapittel 7 oppsummerer oppgaven.

# TEORETISK BAKGRUNN

## 2.1 Kunnskapsorganisasjon

Kunnskapsorganisasjon handler om systematisering og organisering av kunnskap for senere gjenfinning. For at det skal være mulig å gjenfinne dokumenter er det utviklet flere metoder for å beskrive dem, både fysisk og innholdsmessig. Deskriptiv katalogisering beskriver dokumentenes ytre egenskaper slik at det skal være mulig å gjenfinne dem på tittel, forfatter, forlag etc. For å beskrive hva dokumentene handler om, dokumentenes indre egenskap, har man utviklet metoder som klassifikasjon og indeksering. Det er dokumenters indre egenskap denne oppgaven dreier seg om.

Dette kapittelet tar for seg hvordan emner, relasjoner og strukturer blir behandlet i kunnskapsorganisasjon. Kontrollerte vokabularer vil også bli gjennomgått. I tillegg blir World Wide Web, semantic web og navigering behandlet for å vise tradisjonen som emnekart har bakgrunn i.

### 2.1.1 Emne

Det et dokument handler om kalles mer presist for dokumentets emne. Emner står sentralt i kunnskapsorganisasjon. Siden dette også er et viktig element i emnekart er det derfor naturlig å gå nærmere inn på dette. Dette avsnittet vil ta utgangspunkt i hvordan emner defineres i litteratur som omhandler kunnskapsorganisasjon, og dette vil danne grunnlaget for en definisjon for videre bruk i denne oppgaven.

Det er skrevet mye om emner, men

likevel er det vanskelig å finne én allmenngyldig definisjon av begrepet. Ulike forfattere og ulike miljøer definerer emner forskjellig, men det ser samtidig ut til at de legger noe av det samme i begrepet. Birger Hjørland behandlet i sin doktoravhandling emnebegrepet i lys av kunnskapsorganisasjon og gjenfinning. Hjørland hevder at det i kunnskapsorganisasjon har vært tradisjon for å se emne som “en iboende egenskap ved dokumentet på den ene side, og på den anden side at opfatte et dokument emne som mere eller mindre bestemt af brugerens behov, forudsætninger etc.” (Hjørland 1993). Det er ikke nødvendigvis et motsetningsforhold mellom disse synspunktene. Emne kan sies å være en iboende egenskap ved dokumentet, men det gjelder å være klar over at et dokument kan ha flere emner. Brukernes behov er retningsgivende når det skal velges mellom ulike emner.

Hjørland hevder at emner i gjenfinningssammenheng er det som kommer ut av å spørre: hvilke potensielle spørsmål kan disse dokumentene være svaret på? Dette betyr at et dokument emne ikke er en statisk egenskap ved dokumentet, men at emnebegrepet (og emneanalysen) forholder seg til dokumentet på brukerens vegne (Hjørland 1993). På bakgrunn av dette kan hva som blir ansett som et dokument emne endre seg over tid. Det vil ikke dermed si at dersom et dokument etter noen år får andre emner enn det som ble gitt opprinnelig, vil bety at dette nå er feil, men at brukernes behov har endret seg.

Prosessen med å bestemme et dokument emne kalles emneanalyse. Det er utarbeidet en ISO-standard, ISO 5963 (1985), for dette, som beskriver hvordan dokumenter skal analyseres med tanke på å bestemme emnet. Både ISO-standard og Hjørland framhever at dokumenter har mange potensielle emner, og en emneanalyse innebærer at man tar stilling til hvilke emner som skal synliggjøres. Ulike fagmiljøer og fagområder har forskjellige interesseområder og derfor er det naturlig at emner prioriteres i henhold til disse interesseområdene. På den måten vil man oppnå at emneangivelsen fungerer optimalt i forhold til brukerne (Hjørland 1993). ISO-standard framhever, i likhet med Hjørland, at indeksereren bør ha i tankene hvilke spørsmål som kan bli stilt til systemet, og på bakgrunn av dette velge det eller de emnene, av de potensielle emnene, som best passer til brukergruppen.

Denne oppgaven støtter seg til Hjørlands definisjon som sier at et emne er bestemt av brukerens behov. Dette er også i overensstemmelse med ISO-standard. Emner er det et dokument handler om, og et dokument kan ha flere potensielle emner. Det er opp til indeksereren å bestemme hvilke emner som skal realiseres ut ifra kjennskap til fagområdet og brukergruppen.

Som tidligere nevnt finnes det ingen universell definisjon av emnebegrepet. I tillegg brukes ofte de engelske termene subject (emne) og concept (begrep) om hverandre i litteraturen.

Begrep er en term som brukes i andre fagområder som for eksempel psykologi og språkteori. I kunnskapsorganisasjon er det ikke et klart skille mellom disse termene, og de brukes ofte samtidig og refererer ofte til det samme. I ISO-standard, for eksempel, blir ikke emne spesifikt definert, mens begrep derimot blir definert som “a unit of thought” (ISO 5963 1985). I følge standarden er det derfor begrepene som skal identifiseres i emneanalysen, og som er de viktige elementene i beskrivelsen av dokumentets emne.

Den tyske bibliotekforskeren Gernot Wersig laget i sin doktoravhandling en modell som viser hvordan begrepsdannelse skjer hos mennesker. Modellen viser hvordan en organisme kommuniserer med verden rundt seg. I følge Wersig (1971) er persepsjoner fra omverdenen grunnlaget for begrepsdannelse. Ved å generalisere persepsjonene utvikles klasser, og disse klassene danner igjen begreper. Denne typen begreper kan så danne grunnlag for en videre generalisering og på den måten danne abstrakte begreper som ikke er direkte basert på sanseintrykk. Deretter navngis begrepene. Det er viktig å skille mellom selve begrepet og navnet fordi navnet kan uttrykkes på forskjellige måter. Etter hvert får man et nettverk av begreper og koblinger mellom dem. Begrepene er knyttet til hverandre etter som persepsjonsobjektene som ligger til grunn for begrepene, også står i relasjon til hverandre i virkeligheten.

Forholdet mellom emner og begreper forblir derfor noe uklart, men visse skiller og visse likheter er det likevel mulig å identifisere. Akkurat som begreper kan uttrykkes ved termer, skal emner uttrykkes ved emneord eller indekseringstermer (ISO 5963 1985). Det er naturlig å tenke seg at emner er det samme som begreper, og at emneord da svarer til navn på begrepene. Emner kan sies å være standardiserte eller de kan være bestemte begrep valgt innenfor et bestemt kunnskapsområde. Både begreper og emner har feste i virkeligheten, men i tillegg tjener emner til et formål i gjenfinning, et såkalt pragmatisk formål. Emner er derfor en beskrivelse av virkeligheten innenfor et avgrenset område, og for en bestemt samling.

Som nevnt uttrykkes emner ved emneord. I motsetning til naturlig språk, refererer imidlertid ikke emneord til objekter i den virkelige verden eller begreper i en mental verden, men til sine respektive emner. Som navn på et emne refererer i stedet emneordet til settet av alle indekserte dokumenter om

dette emnet. Emneordet referer dermed ikke til det faktiske objektet i den virkelige verden (Svenonius 2000).

### 2.1.2 Kontrollerte vokabularer

Kontrollerte vokabularer er en form for semantisk struktur som skal kontrollere synonymer, skille mellom homonymer, og lenke sammen termer hvis betydning er relatert til hverandre. Termer er her det samme som emneord. Det finnes tre typer kontrollerte vokabularer; bibliografiske klassifikasjonsskjemaer, emneordslister og tesauri (Lancaster 1986). Tesauri vil bli brukt som utgangspunkt i denne oppgaven.

Et kontrollert vokabular har gått gjennom en normalisering som innebærer å behandle termer semantisk for å legge begrensninger i betydningen av dem, og for å vise eksplisitt relasjoner mellom termene (Svenonius 2000). Resultatet blir at hver term refererer kun til ett begrep, og hvert begrep er representert ved kun én term.

Et vokabular brukes til å sette navn på emner innenfor et fagområde. Termene i kontrollerte vokabularer er derfor ord eller kombinasjoner av ord. Kontrollerte vokabularer har først og fremst vært et hjelpemiddel for indekserere for på en konsekvent måte å uttrykke et dokument emne. Etter hvert har kontrollerte vokabularer også blitt tilgjengelige for brukere. Dette for å hjelpe dem i søkeprosessen med å formulere søkeuttrykk. Før brukeren kan få hjelp til å formulere et søkeuttrykk må hun imidlertid ha kommet så langt i søkeprosessen at et spørsmål lar seg formulere. Konsekvensen av at et kontrollert vokabular er tilgjengelig for brukeren blir at hun bruker de sammen termene til å formulere søkeuttrykket som indeksereren brukte til å beskrive dokumentets emne. Dette er derfor med på å bedre gjenfinningen.

### 2.1.3 Relasjoner og strukturer

Emnene skal organiseres på en logisk måte som gjenspeiler det kunnskapsuniverset de representerer. Relasjoner viser hvordan ulike emner henger sammen, og forholdet mellom dem. I kunnskapsorganisasjon har man utarbeidet regler for hvilke relasjoner som skal uttrykkes og hvordan de skal realiseres. I kunnskapsorganisasjon snakker man om flere typer relasjoner. Denne oppgaven vil dreie seg om relasjoner knyttet til emne, og vil derfor ikke komme inn på bibliografiske relasjoner som dekkes av deskriptiv katalogisering.

Det finnes tre hovedtyper av emnerelasjoner; hierarkiske, assosiative og

ekvivalensrelasjoner. Svenonius (2000) peker på at disse relasjonene er avledet fra relasjoner mellom definisjoner i tradisjonelle leksika. Nesten alle definisjoner i naturlig språk refererer til hierarkiske relasjoner av typen slektart. I følge Svenonius har denne typen relasjoner dannet ryggraden i definisjoner helt siden Aristoteles' tid. Nesten alle definisjoner bruker synonymer og relasjoner mellom termer med overlappende betydning når de skal vise likhet i betydning.

Dagobert Soergel (1974) understreker imidlertid at hierarkier ikke er en tvangstrøye som et kunnskapsområde er nødt til å passe inn i. Hierarkier fungerer i stedet som hjelpemidler i indeksering og i utforming av søkeuttrykk. Hva som betraktes som hierarkiske relasjoner kan også variere. Det vanligste er imidlertid å bruke de som det er generell enighet om. Hierarkiske relasjoner kan bygge på logiske, filosofiske eller praktiske prinsipper. Utgangspunktet for hierarkiske relasjoner er vanligvis et sett emner. Ut i fra disse emnene lages det så gjensidig utelukkende kategorier, som igjen blir delt inn i underkategorier.

Hierarkiske, assosiative og ekvivalensrelasjoner vil bli behandlet nærmere i kapittelet om tesauri.

### 2.1.4 Gjenfinning

Marchionini (1995) definerer informasjonssøking som “a process in which humans purposefully engage in order to change their state of knowledge.” Informasjonsgjenfinning er da den delen av informasjonssøkeprosessen som omfatter brukerens interaksjon med et gjenfinningssystem.

Hovedformålet med gjenfinning er å møte et informasjonsbehov hos brukerne. For at dette skal være mulig må man vite hva brukeren er på jakt etter, og en må kunne finne dokumenter som tilfredsstiller dette behovet.

De siste 10-20 årene har gjenfinning vært gjennom et paradigmeskifte som har ført til større fokus på brukerne. Tidligere var det selve gjenfinningssystemet og muligheten for å forbedre den rent tekniske gjenfinningen av dokumenter som stod i sentrum. Etter hvert økte imidlertid interessen for brukerens rolle i gjenfinningsprosessen. Det ble gjort brukerstudier for å kartlegge hvordan de opplever gjenfinningssituasjonen, hva som påvirker gjenfinningen, og hva brukeren mener om resultatet. Gjennom disse studiene kom det fram at det var vanskelig for brukere å formulere et presist søkeuttrykk fordi de var usikre på hva det var de egentlig lette

etter. Dette har fått konsekvenser for gjenfinning fordi behovet for å hjelpe brukeren i å finne ut hva hun leter etter nå ble en oppgave for systemet. Med bakgrunn i dette blir det fokusert på hvordan systemene kan hjelpe brukeren underveis i søkeprosessen.

Indekseringsverktøy som for eksempel tesauri ble gjort tilgjengelig for brukeren slik som beskrevet i avsnitt 2.1.2.

Paradigmeskiftet har også ført til større oppmerksomhet på hvordan ulike navigasjonshjelpemidler kan hjelpe brukeren i søkesituasjonen. I stedet for å fokusere kun på søking, blir det også fokusert på navigasjon. Navigering blir derfor behandlet i avsnitt 2.2.1.

## 2.2 World Wide Web

I følge World Wide Web Consortium, W3C, er World Wide Web et nettverk av informasjonsressurser. World Wide Web støtter seg til tre mekanismer; URler for å lokalisere ressurser, protokoller som gir adgang til bestemte ressurser over World Wide Web, og hypertext som muliggjør enkel navigering mellom ressurser (W3C 1994).

W3C ble opprettet i 1994 av Tim Bernes-Lee, og formålet med organisasjonen er å standardisere den teknologiske utviklingen av World Wide Web. Det var også den samme engelskmannen som på begynnelsen av 90-tallet presenterte World Wide Web. Arbeidet startet da han var ansatt i CERN på 80-tallet. Bakgrunnen var at Bernes-Lee trengte et system som kunne hjelpe ham å holde oversikt over hvilke ansatte i CERN som jobbet med hvilke prosjekter. Dette utviklet seg etter hvert til et ønske om å koble sammen all informasjon som lå på datamaskiner rundt om i verden. Dette var brakgrunnen for at Bernes-Lee utviklet en teknologi som ble utgangspunktet for World Wide Web.

Bernes-Lee sier i boka *Weaving the Web* at han ønsket å lage et nettverk som gjenspeilte måten mennesker tenker på. Dette innebærer at alt kan kobles med alt. “The ultimate goal of the web is to support and improve our weblike existence in the world.” (Bernes-Lee 2000).

Sentralt i World Wide Web står hypertext og lenker. Hypertext er ikke-lineær eller ikke-sekvensiell skri-ving. Hypertextualitet er et tekststruktureringsprinsipp, det vil si en måte å ordne tekstsegmenter på. Hypertekstuell organisering innebærer derfor å opprette et antall eksplisitte relasjoner mellom ulike tekstfragmenter. Disse forbindelsene kalles linker og bestand-

delene har fått navnet noder. Linkenes synlige forankringspunkter i teksten kalles igjen for pekere. Anker er betegnelsen for stedet i den noden man kommer til. (Schwebs 2001).

Selve hypertextbetegnelsen ble skapt av amerikaneren Ted Nelson som presenterte begrepet på en konferanse i New York 1965. Tanken hans var å lage et universelt hypertextsystem som forente det komplekse og det bevegelige i form av en global elektronisk database. Først på 1980-tallet kom derimot de teknologiske forutsetningene for et verdensomspennende hypertextsystem (Schwebs 2001).

World Wide Web har også ført til nye muligheter for gjenfinning. Dokumenter blir her lagt ut i fulltekst noe som muliggjør fulltekstsøk, og søk i bestemte felt er ikke lenger så nødvendig. Med sin nettverksstruktur og lenkemuligheter byr World Wide Web også på andre måter å søke på. Det er ikke lenger nødvendig for en bruker å formulere et eksakt søkeuttrykk for å finne dokumenter, og brukeren kan i stedet navigere seg fram til hun finner det hun leter etter.

### 2.2.1 Navigering

Nettverksstrukturen, samt lenking og hypertext som World Wide Web bygger på, muliggjør som nevnt andre måter enn tradisjonell gjenfinning å søke etter informasjon på. Dersom det skal være mulig for en bruker å finne fram til det hun leter etter, er det viktig med gode navigasjonsmuligheter. Det vil si at brukeren underveis i letingen får hjelp til å velge hvilken vei som skal velges for å oppnå det ønskede resultatet.

I litteraturen blir navigering ofte sidestilt med browsing, som Hjørland gjør i *Informationsvidenskabelige grundbegreper*. Gary Marchionini (1995) på sin side skiller mellom browsing og navigering, og hevder i boken *Information seeking in electronic environments* at navigering kun er en del av browsingen. I denne oppgaven vil det imidlertid ikke være noe skille mellom betydningen av de to begrepene, og termen navigering vil konsekvent bli brukt.

I følge Marchionini blir ofte fysisk navigering brukt som metafor for det å manøvrere på nettet ved hjelp av lenker og hypertext. Navigering blir da den aktiviteten som brukeren er engasjert i når hun følger lenker i en hypertext.

Også i tradisjonelle gjenfinningssystemer blir det lagt vekt på navigasjon. Brukeren skal få hjelp av systemet til å navigere rundt på et område av potensi-

ell interesse. Det er derfor av interesse å se på hvordan navigasjon er behandlet i gjenfinning. Navigasjon slik det blir sett på i gjenfinning kan igjen overføres til navigering på nettstedet.

Navigering skiller seg fra tradisjonell gjenfinning på den måten at brukeren ikke lenger må vite eksakt hva hun leter etter. Som tidligere nevnt må en i tradisjonell gjenfinning formulere hva en leter etter i form av et presist søkeuttrykk. I navigering er det imidlertid nok at brukeren vet hvilket område som er av potensiell interesse for henne. Systemet skal da kunne hjelpe henne til å navigere slik at hun finner det hun er på jakt etter. Underveis kan brukeren ønske og behov endre seg, noe systemet må kunne ta hensyn til.

Marchionini mener at navigasjon er et kompromiss mellom bruker og system. Dette fordi systemet foreslår lenker som brukeren kan følge, men brukeren står fritt til å velge mellom mange lenker, eller eventuelt å fortsette i en lineær form.

Navigasjon innebærer derfor at brukeren blir presentert for ulike navigasjonsmuligheter. Dette betyr at hun må vite noe om hvilke veier hun kan velge og hvor disse vil føre henne. Dette fører igjen til at systemet må kunne gi brukeren gode tilbakemeldinger underveis slik at hun lettere kan navigere rundt på nettstedet.

I denne oppgaven vil navigering være knyttet til navigering i emner slik disse er definert i dette kapittelet. Det forutsettes her at brukeren har en for- mening om hva hun leter etter, og navigeringen skal da kunne hjelpe henne i denne prosessen. Navigeringen skal derfor vise hvor brukeren er og hvilke muligheter hun har for å gå videre. Navigasjonsmulighetene bør være slik at hun lett kan gå tilbake dersom en vei viser seg å ikke være interessant. Brukeren bør i tillegg få vite hva som skiller de ulike veiene og hva konsekvensen blir av å velge den ene veien framfor den andre.

## 2.3 Semantic web

“The Semantic Web is an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. The mix of content on the Web has been shifting from exclusively human-oriented content to more and more data content. The Semantic Web brings to the Web the idea of having data defined and linked in a way that it can be used for more effective discovery, automation, integration, and

reuse across various applications. For the Web to reach its full potential, it must evolve into a Semantic Web, providing a universally accessible platform that allows data to be shared and processed by automated tools as well as by people.” (W3C 1994)

Mesteparten av innholdet på World Wide Web i dag er representert slik at mennesker skal kunne lese det og ikke for at dataprogrammer skal kunne bearbeide det på en meningsfylt måte. Maskiner kan behandle layout på nettsider, men har ingen pålitelige måter for å behandle semantikk (Berners-Lee 2001).

Problemet med dagens World Wide Web er derfor at maskiner ikke kan behandle semantikk. Med dette menes at maskiner ikke har vært i stand til å behandle meningsinnholdet i World Wide Web, men bare har kunnet behandle ord statistisk. Det jobbes derfor med å finne metoder som sørger for at semantikk kan komme til uttrykk på en slik måte at også maskiner kan forstå det. I følge Tim Berners-Lee ligger utfordringen i å utforme et felles språk som uttrykker både data og regler for hvordan det skal “tenkes” om data. Det må også være mulighet for å eksportere regler fra eksisterende systemer ut på nettet (Berners-Lee 2001). Derfor har han lansert idéen om semantic web.

Det er tre grunnleggende komponenter i semantic web:

- XML
- RDF
- ontologier

XML står for *eXtensible Markup Language*, og er et tekstformat som bygger på SGML. XML gir mulighet for å beskrive innholdsstrukturer til dokumenter med egendefinerte elementer. I XML har man mulighet til å bruke XML-navnerom. “Namespaces allow element names and relationships in XML documents to be qualified to avoid name collisions for elements that have the same name but are defined in different vocabularies.” (Connolly 2002) Dette innebærer at data fra ulike kilder kan blandes uten at det oppstår uoverensstemmelser i forhold til bruken av dem.

RDF (*Resource Description Fraework*) er en standard for konsekvent bruk og utveksling av metadata. Bakgrunnen for RDF er at det ikke finnes en standard for metadata som alle bruker. Metadataformater som blir brukt har mye til felles, og RDF er et forsøk på å identifisere disse likhetene. Modellen bruker XML-syntaks for å utveksle og behandle metadata. “RDF [...] provides the foundation

for metadata interoperability across different resource description communities” (Iannella 1998).

### 2.3.1 Ontologier

Ontologier står sentralt i ideen om semantic web, og er også relatert til emnekart. Derfor vil dette avsnittet gi en litt grundigere gjennomgang av begrepet. Forholdet mellom emnekart og ontologier blir tatt opp i avsnitt 3.9.

Ontologier er et begrep som de siste årene har blitt knyttet til informasjonsvitenskap og informasjonsteknologi. I følge bokmålsordboka er ontologi et begrep innen filosofi, og betyr læren om tingenes eksistens, vesen og egenskaper. Alan Gilchrist (2003) har i artikkelen *Thesauri, taxonomies and ontologies – an etymological note* funnet at det er to områder hvor ontologier er særlig fram-tredne. Disse områdene er kunnskapsforvaltning (*knowledge management*) og semantic web. I disse miljøene heter det at en ontologi er et dokument eller en fil som formelt definerer relasjoner mellom termer (Berners-Lee 2001).

Tom Gruber (2003) sier på sin side at en ontologi er “a specification of a conceptualization.” Dette innebærer at en ontologi er en beskrivelse, eller en formell spesifikasjon, av begreper og relasjoner som kan håndteres av maskiner.

I følge Obrst og Liu (XML topic maps 2003) inkluderer en ontologi begreper og de relasjoner som finnes mellom dem, i tillegg til egenskaper ved begrepene, funksjoner og prosesser som involverer begrepene. En ontologi inneholder til sist også begrensninger og regler for disse begrepene.

På bakgrunn av dette kan man si at ontologier gir oversikt over begreper og relasjoner innenfor et fagområde. Relasjoner er ofte avhengig av sammenhengen begreper inngår i, og derfor er det i tillegg nødvendig at ontologien viser definisjoner av begreper og i hvilke sammenhenger de skal brukes.

Ontologier skal kunne håndteres av maskiner. For maskiner gjelder det at det som eksisterer er det som lar seg representere. Ontologier er derfor med på å gjøre begreper i den virkelige verden håndterbare for maskiner (jf ontologier i filosofien).

Ontologier kan blant annet brukes til å muliggjøre utveksling av kunnskap og gjenbruk. Flere maskiner knyttes da opp mot felles ontologier og dette innebærer at det er enighet om å bruke et vokabular på en konsistent måte (Gruber 2003). På den måten kan maskinene kommunisere om et fagområde.

Det som skiller ontologier og kon-

trollerte vokabularer er at ontologier har som mål å gi mulighet for å uttrykke flere begreper og andre begrensninger i forhold til begrepene. Kontrollerte vokabularer organiserer termer, mens ontologier forsøker å ordne begreper. Førstnevnte er begrenset av regler for hvilke termer og relasjoner som skal tas med. I ontologier er det imidlertid opp til den som er ansvarlig for utforming av den å utarbeide begrensningene. Kontrollerte vokabularer er heller ikke utformet med tanke på at maskiner skal kunne forstå dem. På bakgrunn av dette kan for eksempel en tesaurus være grunnlaget for en ontologi, men ikke fungere som en ontologi ettersom den ikke er forståelig for maskiner.

# EMNEKART

Dette kapitlet inneholder en grundig beskrivelse av emnekart og de mest sentrale komponentene i modellen. Kapitlet starter med å gi en definisjon av emnekart, for så å se på bakgrunnen for utarbeidelse av en emnekartstandard. Deretter følger beskrivelsen av selve komponentene i standarden.

## 3.1 Formål og definisjon

I emnekartmiljøet blir det gitt flere forskjellige definisjoner av emnekart. Dette avsnittet vil ta for seg noen av disse, samt se på formålet og bruken av emnekart. Ut i fra dette er målet å komme fram til en definisjon av emnekart til bruk i denne oppgaven.

I ISO-standarden for emnekart (ISO 13250) defineres emnekart ut i fra et teknisk perspektiv. Standarden sier at et emnekart er et dokument som tilfredsstillende de syntaktiske kravene som standarden definerer. I tillegg til ISO-standarden er det utarbeidet en standard, XTM 1.0, for emnekart uttrykt i XML. Her blir emnekart definert som en samling av emner, assosiasjoner og perspektiv, som eksisterer som “a serialized interchange format (e.g. as a topic map document expressed in XTM or some other syntax)” (XTM 1.0 2001). I likhet med ISO-standarden vektlegger også XTM 1.0 syntaksen i sin definisjon av emnekart.

Som definisjonen i XTM 1.0 sier, er utvekslingsformatet viktig. “Topic maps provide a standard approach to creating and interchanging finding aids” (XML

Topic Maps 2003). Det skal derfor være mulig for maskiner å utveksle emnekart, og følgelig må de være i et format som muliggjør dette.

I 2003 ble boken *XML Topic Maps – creating and using topic maps for the web* gitt ut. Boken er skrevet av sentrale personer innen utvikling av emnekart. I boken blir emnekart definert som “a set of topics and associations between them”. Denne definisjonen ser bort i fra det tekniske aspektet og fokuserer i stedet på emner og assosiasjoner mellom dem

I følge Steve Pepper er emnekart “en ny måte å organisere, gjenfinne og navigere informasjonsressurser” (BRODD 2003). Emnekart kan fungere som et navigasjonshjelpemiddel for store samlinger av informasjonsressurser, og blir blant annet brukt på nettportaler og nettstedet for å støtte og forbedre navigasjonsmulighetene. Michael Biezunski sier at “A topic map is a representation of information used to describe and navigate information objects” (XML Topic maps 2003).

XTM 1.0 tar også opp formålet med emnekart og sier at “The purpose of a topic map is to convey knowledge about resources through a superimposed layer, a map, of the resources. A topic map captures the subjects of which resources speak, and the relationships between resources, in a way that is implementation-independent”. (XTM 1.0 2001)

Emnekart gir mulighet for å lage et egendefinert kart over et avgrenset

fagområde, og beskriver kunnskapsstrukturene på dette området. Kunnskapsstrukturene blir så knyttet opp mot informasjonsressurser. Beskrivelsen av kunnskapsstrukturene implementeres ikke i selve ressursene, men eksisterer som et eget lag.

På bakgrunn av dette blir emnekart definert som følger i denne oppgaven: “Emnekart er en teknologi som muliggjør beskrivelse og framvisning av et kunnskapsområde. I tillegg understøtter emnekart navigasjon ved utstrakt bruk av assosiasjoner. Emnekart er et publiseringsverktøy for World Wide Web, og skal uttrykkes i et format som muliggjør utveksling av informasjon.”

## 3.2 Historikk

Den opprinnelige interessen for emnekart er relatert til behovet for å kunne koble sammen indekser. “Topic maps originated in the problem of formalizing and interchanging finding aids like back-of-the-book indexes, glossaries, and thesauri” (XML topic maps 2003). Etter hvert ble dette utvidet til også å gjelde andre former for navigasjonshjelpemidler som har til felles at de forsøker å gi adgang til informasjon basert på en modell av kunnskapen de uttrykker (Pepper 1999).

Arbeidet startet i 1991 da Davenport Group ble opprettet av ulike systemutviklere, blant annet fra UNIX. Målet var å forbedre konsistensen i trykt dokumentasjon fra de ulike firmaene. Problemet var at ulike termer ble brukt for det

samme emnet, og i tillegg ble det brukt ulike termer i forskjellig systemdokumentasjon. Målet var å samle ulike dokumentasjon om de enkelte systemene. Utfordringen lå derfor i å opprette en hovedindeks for uavhengig administrert og konstant endret teknisk dokumentasjon samlet i systemmanualer hos programutviklere (XML Topic Maps 2003). Arbeidet resulterte i SOFABED<sup>1</sup>.

I 1993 ble det dannet en ny gruppe som fortsatte arbeidet. Denne gruppen tilførte modellen hypertextfasiliteter fra HyTime-standard<sup>2</sup>. Etter å ha studert mulighetene som lå i utvidet hyperlenkenavigasjon, videreutviklet arbeidsgruppen SOFABED til emnekart. I 1995 ble arbeidet godkjent som basis for en ny internasjonal standard. Den endelige emnekartstandard ble så lansert som ISO 13250 i 2000.

Ettersom ISO 13250 bruker HyTime som bygger på SGML-syntaks, startet arbeidet med å utarbeide en spesifisering for emnekart bygd på XML-syntaks umiddelbart etter lanseringen av ISO-standard. Bakgrunnen var ønsket om å lage en utforming av emnekart til bruk i web-sammenheng. Det ble opprettet en uavhengig organisasjon, TopicMaps.org, som laget XML Topic Maps og stod for utgivelsen av XTM 1.0-spesifikasjonen. Dette er et standardformat for emnekart i XML, og spesifiseringen er nå innlemmet som et vedlegg til ISO 13250.

XTM 1.0 stiller med dette på lik linje med ISO-standard. “While some topic map applications support a

restricted subset of the ISO 13250 interchange syntax, almost all those available today support XTM 1.0” (Ahmed 2003). Denne oppgaven bygger derfor på XTM 1.0, og selv om det er enkelte elementer som skiller denne fra ISO-standard, er ikke forskjellene av stor betydning.

## 3.3 Tema og emne

Emnekartmodellen har et “subject-centric view” (XML Topic Maps 2003) på den måten at emner ligger til grunn både for assosiasjoner og forekomster. Dette innebærer at emnekart fokuserer på hva ressursene handler om, og ikke på selve ressursen (Ahmed 2003). Emner (på engelsk *topics*) står dermed sentralt i emnekart. Et emne kan være hva som helst; en person, hendelse, bok, et forfatterskap, en webside, en bygning etc. Alt kan derfor defineres som emne.

<topic>elementet brukes til å definere emner i emnekart. Alle komponenter i et emnekart må defineres med dette elementet for å kunne brukes i emnekartet. “Topics are what make subjects “real” for the system and come as close as a machine can to a representation of what is “real” for humans” (XTM 1.0 2001). For at et emne skal kunne håndteres av systemet er det altså nødvendig å definere det som et topic. “Topic [...] an electronically processible stand-in for a thing that doesn’t necessarily have to be processible itself” (Ahmed 2003).

Emner representerer ulike temaer (subjects) som informasjonsressursene omhandler. I XTM 1.0 heter det at et

tema er alt som kan omtales eller forstås av mennesker. “In the most generic sense, a subject is anything whatsoever, regardless of whether it exists or has any other specific characteristics” (XTM 1.0 2001). Et tema er noe abstrakt som konkretiseres ved et emne. Dette betyr at et tema blir definert som et emne ved å bruke <topic>-elementet. Et emne er altså et element i emnekartet, og representerer et tema. For at temaet helse skal kunne forstås av en datamaskin er det nødvendig å konkretisere det ved et emne:

```
<topic>helse</topic>
```

Pepper (2002a) sammenligner emner og tema med Platons hulelignelse; tema er det Platon kaller en idé, mens emnet er skyggen av idéen på huleveggen. Platons ideer kan sammenlignes med det Wersig kaller begreper. Forskjellen blir imidlertid at begrepene hos Wersig eksisterer i den interne omverdensmodellen hos en organisme, mens Platons ideer eksisterer i en parallell verden til vår egen, en verden som bare kan nås via tanker. Likevel vil det ikke være unaturlig å si at ideene og begrepene står for det samme. “Plato’s idea that there is a plane of existence that is accessible only by our minds is exploited by the topic maps paradigm in order to make data resources federable without endless layers of metadata upon metadata.” (XML topic maps 2003). Prosessen med å definere et tema som

emne beskrives i standardene som konkretisering av noe abstrakt. Det abstrakte temaet blir konkretisert ved et emne, og dette gjør det mulig å behandle temaet innenfor emnekartet.

Det skal være en én-til-én relasjon mellom emne og tema; et tema skal bare representeres ved ett emne, og et emne skal bare representere ett tema (XTM 1.0 2001). Én-til-én relasjonene sikrer at all kunnskap om et spesielt tema kan nås gjennom ett emne, jf. samlokaliseringsprinsippet.

For å illustrere dette og for å få en bedre forståelse av emnekart vil det nå følge et eksempel som er hentet fra nettstedet forskning.no. Dette nettstedet har tatt i bruk emnekart og en grundigere beskrivelse av dette vil følge i kapittel 5.

Utgangspunktet for eksemplene i det følgende er en anmeldelse av boka *The Blank Slate* av Steven Pinker. Anmeldelsen er skrevet av Erik Tunstad, faglig redaktør i forskning.no. I forskning.no er anmeldelsen blant annet relatert til kognitiv psykologi. Ut i fra dette kan man tenke seg at følgende emner er definert i emnekartet:

```
<topic id="Steven_Pinker">
</topic>
```

```
<topic id="The_Blank_Slate">
</topic>
```

```
<topic id="Erik_Tunstad">
</topic>
```

```
<topic id="Kognitiv_psykologi">
</topic>
```

id er her et attributt som kan tildeles et <topic>-element for at man senere kan identifisere emnet.

På bakgrunn av det som hittil er skrevet om emner kan man si at emnene i et emnekart peker til fenomener, dvs. temaer, i den virkelige verden. <topic>-elementet gjør det mulig for systemer å håndtere disse fenomenene.

Emner har tre typer karakteristiker; navn, forekomster og roller i assosiasjoner. Disse er bare gyldig innenfor et bestemt perspektiv (se avsnitt 3.6). Emnekarakteristikkene blir beskrevet i de påfølgende avsnittene.

3.3.1 *Emnenavn*

Et emne kan ha ingen eller flere navn, og hvert navn er kun gyldig innenfor et definert perspektiv. Navnet fungerer da som en etikett for emnet. Et emne har alltid et basisnavn, men kan i tillegg ha flere variantnavn. Et basisnavn er basisformen av emnenavnet og består alltid

av en streng med tegn. Et variantnavn er en alternativ form av et basisnavn, og med et spesielt formål, for eksempel sortering eller framvisning. Et variantnavn kan være en streng, et ikon eller en hvilken som helst ressurs.

I forlengelsen av eksemplet ovenfor kan emnet *Steven Pinker* for eksempel ha basisnavnet Steven Pinker og variantnavnene kan for eksempel være *Pinker*, *Steven*, eller et bilde av Steven Pinker:

```
<topic id="Steven_Pinker">
  <baseName>
    <baseNameString>StevenPinker<
      /baseNameString>
  </baseName>
  <variantName>
    <variantNameString>Pinker,Ste
      ven</variantNameString>
  </variantName>
</topic>
```

For emnet *The\_Blank\_Slate* kan basisnavnet være *The Blank Slate*. Et bilde av bokomslaget og den norske tittelen på boka kan for eksempel fungere som variantnavn.

Ved hjelp av et sett av emner kan man uttrykke i hvilke sammenhenger et variantnavn skal brukes. Dersom et emne ikke har noen variantnavn vil applikasjonen bruke basisnavnet.

For å sikre at et emnekart er konskvent og ikke inneholder dubletter av emner, er det i emnekartstandarden utarbeidet metoder for å sørge for dette. Emnekartstandarden opererer derfor med noe som kalles begrensninger av emnenavn (*topic naming constraints*). Denne begrensningen innebærer at to emner ikke kan ha eksakt samme basisnavn i det samme perspektivet. Perspektiv blir behandlet i avsnitt 3.6. Det betyr at dersom to emner har det samme basisnavnet i det samme perspektivet så representerer de det samme temaet. Etter navnebegrensningsregelen vil derfor de to emnene bli slått sammen. På den måten unngår man å operere med to emner som representerer det samme temaet i emnekartet. Begrensninger i basisnavn er viktig i forhold til prosessen med å flette sammen emnekart. Dette blir grundigere behandlet i kapittel 6.

Emnenavn har altså som funksjon å gi navn til emner i et emnekart. Et emne kan ha flere variantnavn, men bare ett basisnavn.

3.3.2 *Temaidentitet og temaindikatorer*

Det er viktig at et emnekart ikke inneholder flere emner som representerer det samme temaet. Dette vil ødelegge

for samlokaliseringsprinsippet som sier at det skal være mulig å finne alt om et emne på samme sted. I tillegg til begrensninger i emnenavn opererer emnekartstandarden med temaidentitet for å sikre konsistente emnekart. Temaidentitet skal kunne gi en unik identifisering av et tema og på den måten kunne si noe om to emner representerer det samme temaet.

XTM 1.0 definerer temaidentitet slik: “That which makes two subjects identical, or distinguishes one subject from another.” Denne definisjonen kan imidlertid være noe misvisende. Temaidentitet er med på å avgjøre om to emner representerer det samme temaet, men er ikke med på å gjøre to temaer identiske. I avsnittet om emne og tema ble det sagt at tema kan sammenlignes med Platons ideer. To ideer i idéverden kan ikke være identiske. Det eksisterer ikke flere ideer som er identiske, det finnes bare en idé for fenomener i den virkelige verden, og denne ideen er unik. For resten av oppgaven vil det derfor være slik at temaidentitet identifiserer temaer og kan derfor brukes til å bestemme om to emner representerer det samme temaet. Temaidentitet spiller en sentral rolle i sammenfletting av emnekart som behandles i kapittel 6.

Et tema kan være enten adresserbart eller ikke adresserbart. Dersom temaet er adresserbart kan identiteten utledes av adressen. Temaet er i slike tilfeller en adresserbar ressurs, for eksempel en nettside som er adresserbar gjennom en URI<sup>3</sup>. Nettversjonen av bokanmeldelsen *Hjernen som bøtte* av Erik Tunstad er et eksempel på en adresserbar ressurs. Temaidentiteten til anmeldelsen vil være adressen: *http://www.forskning.no/Artikler/2003/februar/1045572790.18*. For emnet Hjernen som bøtte vil denne adresserbare ressursen være temaet. XTM 1.0 understreker dette ved å si at en adresserbar informasjonsressurs blir betraktet som “a subject in and of itself, and not considered in terms of what an author meant by it”. Temaet er representert ved et emne i emnekartet.

Det er imidlertid slik at de færreste temaer er adresserbare. Personer, begreper innenfor et fagområde, eller trykte publikasjoner er eksempler på temaer som ikke er adresserbare. Forfatteren av bokanmeldelsen, Erik Tunstad, vil være et eksempel på et tema som ikke er adresserbart.

I emnekartstandarden brukes temaindikatorer som løsning for å identifisere temaer som ikke er adresserbare. I XTM 1.0 defineres temaindikatorer som “resource that is intended by the

topic map author to provide a positive, unambiguous indication of the identity of a subject”. En temaindikator er altså en ressurs som brukes til å identifisere temaer som ikke er adresserbare. Nettopp fordi temaidentifikatoren er en ressurs og har en adresse, kan den brukes som temaidentifikasjon. Temaet *Erik Tunstad* kan identifiseres ved å bruke for eksempel hjemmesiden hans som temaindikator. For den trykte utgaven av boka *The Blank Slate* kan man for eksempel bruke adressen til forlagets presentasjon av boka som temaidentitet. Emnet *kognitiv psykologi* er heller ikke adresserbart og kan for eksempel identifiseres ved å bruke definisjonen av begrepet i en online ordbok som temaindikator.

Når to emner bruker samme ressurs for å indikere temaet, representerer de per definisjon det samme temaet og skal derfor slås sammen (XTM 1.0 2001). Dette vil også bli tatt opp i kapittelet om sammenfletting av emnekart.

3.3.3 *Emnetyper*

Emnekartstandarden åpner for at emner kan kategoriseres etter type, emnetyper. Et emne trenger ikke å tilhøre noen bestemt emnetype, men kan også tilhøre flere emnetyper. Emnetype er en form for snarvei til relasjoner av typen klasseforekomst.

Dette innebærer at emner kan kategoriseres etter klassen de tilhører. For eksemplet med Erik Tunstad vil det være slik at emnet *Erik Tunstad* kan sies å være en forekomst av klassen redaktør, mens *Steven Pinker* kan sies å være en forekomst av klassen forfatter. Både redaktør og forfatter må defineres som emne før de kan brukes som emnetype. Dette er den eneste måten temaer blir forståelige for systemet.

```
<topic id="forfatter">
  <baseName>Forfatter</baseName>
</topic>
```

```
<topic id="Steven_Pinker">
  <instanceOf>
    <topicRef xlink:href="#forfatter"/>
  </instanceOf>
</topic>
```

Her blir emnetypen *forfatter* definert som et emne i emnekartet ved å bruke <topic>-elementet. Deretter defineres *Steven Pinker* som et emne på samme måte, i tillegg til at han defineres som et tilfelle av emnetypen *forfatter*. Dette gjøres ved å bruke <instanceOf>-elementet som inneholder et henvisningselement

hvor man angir hvilket emne det aktuelle emnet er en type av.

Ved å bruke emnetyper oppnår man at emner kan listes opp etter hvilken klasse de tilhører. Dette sier noe om emnene og gjør det også mulig for systemet å skjønne hvordan enkelte emner henger sammen. Emnetyper gjør det mulig å samle emner av samme type slik at det for eksempel går an å finne alle emner av typen forfatter ved å gå via dette emnet. Det er opp til den som lager emnekartet å bestemme emnetypane.

**3.4 Assosiasjoner**

Akkurat som temaer er relatert til hverandre i den virkelige verden, er også emner i et emnekart relatert til hverandre. Relasjoner mellom emner kalles i emnekart for assosiasjoner. For at det skal være mulig for maskiner å forstå at det er en relasjon mellom to emner, brukes <association>-elementet. Assosiasjoner gjør det altså mulig å beskrive relasjoner mellom to eller flere emner.

Det kan være greit å være oppmerksom på forskjellen mellom assosiasjoner og kryssreferanser. Begge er hyperlenker, men i kryssreferanser ligger endepunktene for hyperlenken i selve informasjonsressursen. For assosiasjonene derimot, er det snakk om lenker mellom emner som er uavhengig av hvilke informasjonsressurser som eksisterer eller som betraktes som en forekomst av de aktuelle emnene (Pepper 2002b).

Hvilke assosiasjoner som skal gjelde mellom emner er det opp til den som lager emnekartet å bestemme. Assosiasjoner kan brukes til å realisere tesaurusrelasjoner, eller databasestrukturer, eller til å vise hierarkiske strukturer (XML topic maps 2003). Emnene i det foregående eksemplet kan for eksempel være relatert til hverandre gjennom følgende assosiasjoner:

— *The Blank Slate* er skrevet av *Steven Pinker*
— *Hjernen som bøtte* er skrevet av *Erik Tunstad*
— *Hjernen som bøtte* handler om *The Blank Slate*.
— *Erik Tunstad* er ansatt i *forskning.no*.
*forskning.no* ligger i Oslo
— *Kognitiv psykologi* hører til *psykologi*

3.4.1 *Assosiasjonstyper*

Assosiasjoner kan grupperes etter type, og disse assosiasjonstypene defineres også som emner. Assosiasjonstyper gjør det mulig å gruppere emner med samme relasjoner til et annet emne. “It is the type of the association that defines the nature of the relationship” (Ahmed 2003). Emnetyper blir sett på som en

spesiell assosiasjonstype. Eksemplene ovenfor har følgende assosiasjonstyper:

```
er_skrevet_av
handler_om
er_ansatt_i
ligger_i
hører_til
```

Emner kan være relatert til hverandre gjennom flere assosiasjonstyper:

```
Emne1 – assosiasjonstype1 – emne2
Emne1 – assosiasjonstype2 – emne2
Emne1 – assosiasjonstype3 – emne2
```

I likhet med emnetyper må også assosiasjonstyper defineres som emner før de kan brukes i et emnekart. Assosiasjonstyper gjør det mulig å gruppere sett av emner med samme assosiasjon til et emne.

3.4.2 *Assosiasjonsroller*

“An association is a relationship between one or more topics, each of which plays a role as a member of that association.” (XTM 1.0 2001)

Emner som deltar i en assosiasjon er medlemmer i assosiasjonen. Hvert medlem har i tillegg en rolle i assosiasjonen; assosiasjonsrolle. Hvilken rolle et emne har i en assosiasjon bestemmes av perspektivet. For assosiasjonene ovenfor kan for eksempel disse assosiasjonsrollene være aktuelle: person, sted, dokument, organisasjon, emne og fag. Uttrykt ved assosiasjonsroller og assosiasjonstyper blir assosiasjonene da:

```
[dokument] er_skrevet_av [person]
[dokument] handler_om [dokument]
[person] er_ansatt_i [organisasjon]
[organisasjon] ligger_i [sted]
[emne] hører_til [fag]
```

Uttrykt med XML-syntaks blir assosiasjonen [dokument] er\_skrevet av [person] seende slik ut:

```
<topic id="person">
  <baseName>Person</baseName>
</topic>
```

```
<topic id="document">
  <baseName>Dokument</baseName>
</topic>
```

```
<association id="document_person_
  assosiasjon">
  <member>
    <topicRef xlink:href="#dokument"/>
  </member>
  <member>
```

```

<topicRef xlink:href="#perso
n"/>
</member>
</association>

```

Assosiasjoner går alltid begge veier (*are inherently multidirectional*): hvis A er relatert til B må også B være relatert til A, dersom *The Blank Slate* er skrevet av Steven Pinker, følger det automatisk at Steven Pinker skrev *The Blank Slate*. I stedet for å angi retning på assosiasjoner brukes assosiasjonsrollene for skille mellom ulike måter medlemmene er innblandet i assosiasjonene. Dersom man har en `ble_påvirket_av`-relasjon hvor Steven Pinker og Erik Tunstad deltar, trenger man å vite hvem som påvirket hvem.

I noen tilfeller vil en assosiasjon være symmetrisk, dvs. at den er den samme begge veier; dersom Steven Pinker hadde samarbeidet med Erik Tunstad, følger det at Tunstad også samarbeidet med Pinker. I dette tilfellet blir assosiasjonen `[person] samarbeidet_med [person]`. I andre tilfeller vil assosiasjonen være den samme, men rollene vil være annerledes. Dette gjelder i assosiasjonen `[ektemann] er_gift_med [hustru]`.

Andre assosiasjoner som uttrykker klasse/subklasse- og helhet/del-forhold er transitive. Dette betyr at dersom A er relatert til B, og B er relatert til C, så er også A relatert til C. Dersom psykologi er relatert til kognitiv psykologi, og kognitiv psykologi er relatert til personlighetsutvikling, så er også psykologi relatert til personlighetsutvikling. Hierarkisk arv er et eksempel på transitivitet; underordnede emner arver egenskaper til overordnede emner.

Relasjoner kan involvere en, to eller flere roller. Det er ingen begrensninger for antall typer relasjoner som kan eksistere mellom emner. I tillegg er det ingen standardisert måte for hvordan ulike relasjoner skal realiseres.

Assosiasjonsroller kan også deles inn i typer, og en assosiasjonsrolletype er også et emne.

### 3.5 Forekomster

Et emne kan være lenket til en eller flere informasjonsressurser som er relevante for emnet. Ressurser kalles i emnekart-standard for forekomster. Slike forekomster kan være en monografi, en artikkel, omtale, kommentar, deler av et dokument, en video etc. som behandler eller omtaler temaet som emnet representerer.

Forekomster kan være av ubegrenset antall forskjellige typer (forekomsttyper), og kan organiseres deretter ved å referere til et emne som definerer typen.

Forekomster kan for eksempel være av typen hjemmesider og da må hjemmesider være definert som et emne i emnekartet.

For eksemplet med *Steven Pinker* og *The Blank Slate*, vil anmeldelsen *Hjernen som bømte* være en forekomst som er relevant for emnet. Videre kan både anmeldelsen og boka være forekomster til emnet *psykologi*. Andre mulige forekomster for dette emnet kan være spesialister innen faget, institusjoner eller ulike artikler.

Forekomster er som regel eksterne og finnes ikke i selve emnekartet, men blir henviset til ved å bruke URIer. Dette innebærer at man ikke er avhengig av metadata i selve forekomsten.

### 3.6 Perspektiv

Perspektiv angir når de ulike emnekaraktistikkene; navn, forekomster og roller i assosiasjoner, er gyldige. Et perspektiv viser altså i hvilken sammenheng et navn eller en forekomst er gitt til et emne, og i hvilken sammenheng emner er relatert til hverandre gjennom assosiasjoner. Dersom det ikke angis et bestemt perspektiv, gjelder det ubegrensede perspektivet (*unconstrained scope*). Emnekaraktistikker i det ubegrensede perspektivet er alltid gyldige.

“The scope mechanism of topic maps enables any information provided about a topic to be qualified by defining a context within which the information is valid” (Ahmed 2003).

Eksempler på ulike perspektiv kan være språk, målgruppe eller tid. Dersom perspektivet er norsk vil man for eksempel bruke den norske tittelen til boka av Steven Pinker. Er perspektivet tid kan man bruke *Kristiania* for perioden 1877-1925, *Oslo* for tiden før og etter.

“Topic map authors must use scopes to distinguish between the different meanings of any name that is used for more than one subject” (ISO/IEC 13250). Perspektivet brukes altså til å løse homonymi og polynomi problemer. Perspektiv kan for eksempel brukes til å skille mellom de ulike betydningene av ordet *horn*. Dersom målgruppen er bakere skilles betydningen av horn som *bakverk* fra henholdsvis betydningen *instrumenter* og *utvekst på hodet hos mange hovdyr*. På grunn av begrensningen i emnenavn (avsnitt 3.3.1) kan ikke to emner ha samme basisnavn i det samme perspektivet.

Perspektiv kan lette navigeringen ved å forandre synsvinkelen inn mot emnekartet basert på brukerprofiler og etter måten emnekartet blir brukt på (Pepper 2002b).

Man kan indikere i hvilke sammenhenger de ulike navnene skal brukes, eller innenfor hvilket perspektiv. I perspektivet norsk kan den norske oversettelsen av boktittelen brukes.

### 3.7 Lagdelingen i emnekart

Emnekart er delt i to lag. I det ene laget befinner emner og assosiasjoner seg, mens forekomstene er i det andre laget. På den måten kan man si at selve emnekartet består av emner og assosiasjoner, mens forekomster eksisterer utenfor emnekartet og blir pekt til ved hjelp av URIer.

I kunnskapslaget finnes emner og assosiasjoner mellom emner. I informasjonslaget befinner ressursene seg.

Delingen mellom ressursnivået og selve emnekartet gjør det mulig å bruke det samme emnekartet for flere informasjonssamlinger. I og med at emnene og assosiasjonene ikke er avhengige av forekomstene er det mulig å bruke samme emnekart på ulike samlinger av forekomster. Dette kan gi flere synsvinkler på det området som emnekartet dekker og kan gi brukerne flere innfallsvinkler til det samme stoffet. Lagdelingen av emnekart gjør det også mulig å bruke ulike emnekart for den samme samlingen av forekomster og på den måten sørge for flere innfallsvinkler til innholdet.

Ettersom beskrivelsen av kunnskapsområdet er skilt fra selve ressursene ligger vekten i emnekart på form mer enn på innhold. Det vil si at ressursene kan være i et hvilket som helst format.

### 3.8 Begrensninger

Etter at emnekart ble tatt i bruk har behovet for å kunne legge begrensninger i det meldt seg. Dette fordi emnekart fort blir komplekse og det er lett å miste kontrollen over konsistensen. Det arbeides nå med å utforme et begrensningsspråk for emnekart, kalt Topic Map Constraint Language, TMCL. Arbeidet med TMCL er i fasen med å samle krav som et slikt språk bør dekke med tanke på en mulig standard for dette. Tanken med TMCL er imidlertid at det er den som lager emnekartet som, i samråd med brukerne av det, skal bestemme begrensningene.

Begrensningene vil gi den som lager emnekartet mulighet til å bestemme hvordan temaer og emner skal sees i sammenheng. Begrensninger vil være avhengig av hvilket domene emnekartet skal brukes innenfor. Begrensningene skal altså sikre at emnekartet er semantisk gyldig i henhold til forfatterens intensjoner.

Ifølge TMCL er en begrensning “any restriction placed on a topic map construct that is used to confine or narrow its semantics or to further define cardinality and data types where they had not been defined within the topic map model.” (ISO/IEC 19756 2003)

Alle mulige begrensninger vil ikke dekkes i TMCL. Bare de vanligste behandles, men det gis mulighet for å definere andre begrensninger i tillegg til disse. Hovedkravene for TMCL gjelder assosiasjonstyper, emnetyper, forekomsttyper, emnekaraktistikker, assosiasjonsrolletyper og perspektiv.

### 3.9 Emnekart og semantic web

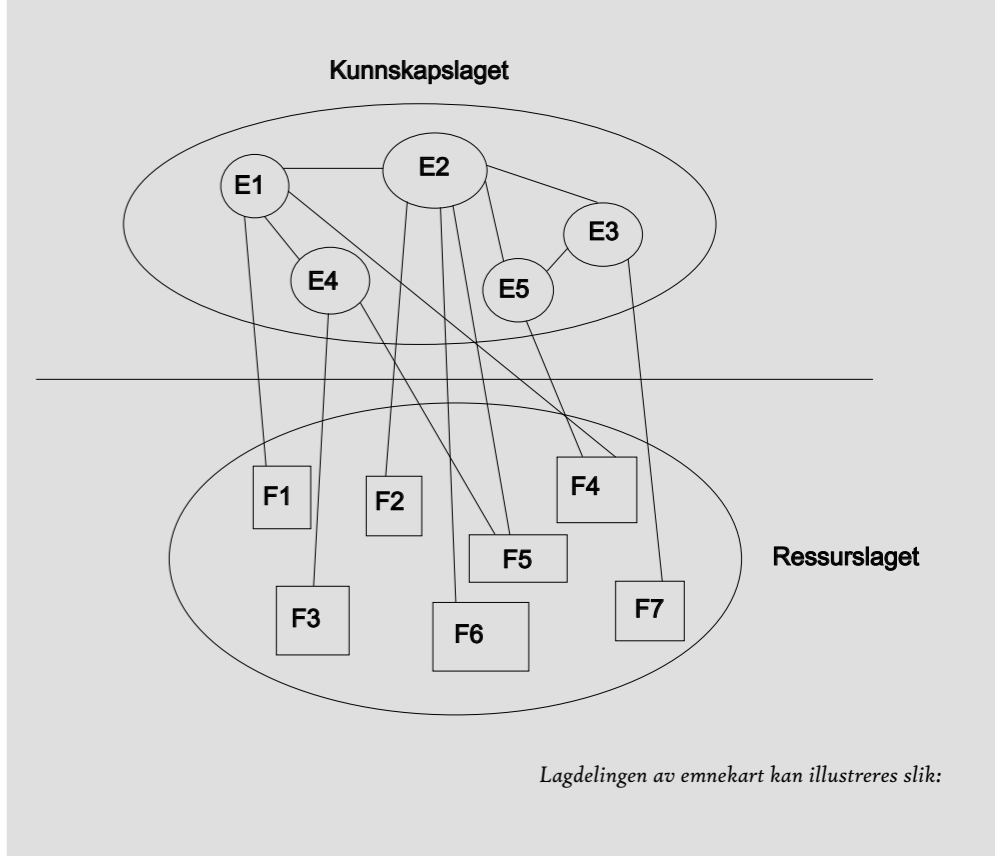
For å kunne realisere semantic web trengs det i følge Tim Berners-Lee tre grunnleggende komponenter, nemlig RDF, XML og ontologier. I følge personer som har vært med på å utvikle emnekart, kan også emnekart være med å realisere det semantic web.

Fra noen hold betraktes RDF og emnekart som konkurrenter, men flere tar nå til orde for at disse to sammen vil kunne være med på å realisere det semantic web. De to modellene skiller seg fra hverandre på flere måter. Den største forskjellen ligger i at RDF berører selve ressursen, mens emnekart beskriver ressurser i et eget lag, atskilt fra ressursene.

Som tidligere nevnt ble det med en gang ISO-standard for emnekart ble lansert, satt ned en arbeidsgruppe som skulle utarbeide en emnekartstandard hvor emnekart lot seg uttrykke i XML.

Et emnekart kan være relatert til en ontologi på to måter. Et emnekart kan inneholde deler av en ontologi. “The set of privileged topics and their characteristics, including associations between them, is what we call the topic map ontology” (Grønmo 2000). Dette innebærer at spesielt utvalgte emner, deres ressurser og assosiasjoner de inngår i, utgjør ontologien for emnekartet. Med bakgrunn i dette kan man si at emnetyper, assosiasjonstyper og forekomsttyper er eksempler på ontologiemer.

Emnekart kan også genereres fra ontologier. I den sammenheng er ontologien da en egen gjenstand utenfor det emnekartet den skaper (XML topic maps 2003). Ontologien vil da ikke være en del av selve emnekartet, men hele eller deler av den blir i stedet realisert i emnekartet. Det finnes foreløpig ikke verktøy som automatisk kan generere et emnekart ut i fra en ontologi (XML topic maps 2003 s. 129). Obrst og Liu hevder at det er flere fordeler med å generere emnekart ut i fra en atskilt



Lagdelingen av emnekart kan illustreres slik:

ontologi. Disse fordelene gjelder særlig ved oppdatering av emnekartet da innholdet blir vedlikeholdt ved å oppdatere ontologien. Anvendelse av en atskilt ontologi innebærer dessuten også mulighet for å bruke en som allerede eksisterer.

Forskjellen på ontologier og emnekart er at ontologier er konstruert på et språk som fokuserer på kunnskapsorganisasjon, mens emnekart er et produkt som er laget for spesielt å organisere internettressurser i et overliggende lag som ikke berører den opprinnelige organiseringen av ressursene. En enkelt ontologi kan brukes av flere emnekart (XML topic maps 2003 s. 130).

I følge Obrst og Liu sikrer emnekartstandard syntaktisk interoperabilitet, mens ontologier sørger for semantisk interoperabilitet. Dersom emnekart bygger på en god ontologi kan de derfor sørge for semantisk interoperabilitet ikke bare seg i mellom, men også blant applikasjoner som benytter seg av dem (XML topic maps 2003 s. 125).

Emnekart hører altså hjemme i tradisjonen med semantic web fordi de kan uttrykkes i XML-syntaks, erstatte eller eventuelt virke sammen med RDF, og realisere ontologier.

### 3.10 Emnekart og kunnskapsorganisasjon

Byggesteinene i et emnekart er emner som har navn og som peker til forekomster i eksterne ressurser. Assosiasjoner relaterer emnene til hverandre.

Emne står sentralt både i emnekart og kunnskapsorganisasjon. Derfor er det naturlig med en sammenligning av begrepet slik det blir brukt i de to miljøene.

Emnekartstandard opererer med både emne og tema. Termen emne har en bredere betydning i emnekartstandard enn emne i kunnskapsorganisasjon. Hver komponent som skal brukes i et emnekart må defineres som et emne ved hjelp av `<topic>`-elementet. Da blir emne derfor et element som har som formål å uttrykke semantikk slik at maskiner kan behandle det som er definert ved `<topic>`-elementer. Når det gjelder sammenhengen mellom tema og emne, er tema fenomener i den virkelige verden, mens emne gjør disse fenomenene håndterbare for systemet. I denne sammenheng blir det derfor naturlig å sammenligne tema med det man i kunnskapsorganisasjon legger i begrepet emne.

I kapittel 2 ble emne i kunnskapsorganisasjon definert som det et dokument handler om. I tillegg er emnet bestemt av brukerens behov. I det samme kapitlet ble emner og begrep sammenlignet. Ettersom tema i emnekart er fenomener i den virkelige verden, vil det være naturlig å sammenligne dette med begreper slik de ble behandlet i kap. 2.

Det er særlig en forskjell mellom tema og begrep som utpeker seg. I emnekartmiljøet blir tema sammenlignet med Platons ideer, mens Wersig snakker om begreper som noe som oppstår i et

individs interne omverdensmodell. Forskjellen blir da at tema eksisterer utenfor individet, i en egen verden, mens begreper i følge Wersig eksisterer i den interne omverdensmodellen.

Begreper eksisterer i den interne omverdensmodellen, mens emner er utvalgte begreper som har et spesielt formål i gjenfinning. Tema eksisterer i en egen idéverden. Begrep uttrykkes ved ord, emner ved emneord, og tema uttrykkes som emnenavn. Fordi det opereres med emnenavn i emnekart, kan man snakke om temaer. I tillegg gjøres tema håndgripelig for systemet ved å defineres som emne.

På den andre siden kan det diskuteres om dette er emnet i ontologien eller om det er det abstrakte emnet som blir et element i emnekartet ved å definere det med <topic>-elementet. Dette kommer ikke klart fram i standarden og er med på å problematisere hva et emne egentlig er.

Ut i fra dette kan man si at emner i et emnekart peker til fenomener (temaer) i den virkelige verden. Emnene er formaliserte som en allmenngyldig definisjon av begreper. Dokumenter handler om fenomener i den virkelige verden, begreper. På den måten vil emner også i denne sammenhengen peke til ulike begrep. Termer uttrykker begreper og gjør det mulig å snakke om dem.

Emnekartstandardens sier selv at det er vanskelig å skille mellom emner og tema, henholdsvis mellom *topic* og *subject*, og sier at begrepene brukes om hverandre (XTM 1.0 2001). Dette kommer blant annet til syne i diskusjonen rundt temaidentitet. I standarden heter det at temaidentitet brukes til å avgjøre om to temaer er identiske. Dette er som nevnt problematisk. Skal man følge tankegangen om at tema er det samme som ideer i Platons idéverden, blir dette umulig. To temaer kan ikke være identiske, men to emner i emnekart kan være identiske og på den måten representere det samme temaet. Dette er et eksempel på problemer knyttet til terminologi i emnekartstandardens.

Todelingen av emnekart tydeliggjør at det ikke er forekomstene som er fokusert i emnekart. Emner og assosiasjoner mellom dem, eksisterer uavhengig av forekomstene. Det er ikke et krav at alle emner skal ha forekomster. Dette kan sammenlignes med kunnskapsorganisasjon. Emner og relasjoner mellom dem eksisterer også her uavhengig av dokumenter, men fokuset ligger på å finne ut hva et dokument handler om, for så å tildele det emner.

---

<sup>1</sup> Standard Open Formal Architecture for Browsable Electronic Documents

<sup>2</sup> HyTime ble lansert i 1992 for å sørge for SGML med multimedia og hyperlenkingsmuligheter. ISO 10744

<sup>3</sup> Uniform Resource Identifiers (URIs, aka URLs) are short strings that identify resources in the web: documents, images, downloadable files, services, electronic mailboxes, and other resources. They make resources available under a variety of naming schemes and access methods such as HTTP, FTP, and Internet mail addressable in the same simple way. They reduce the tedium of "log in to this server, then issue this magic command ..." down to a single click. (hentet fra: [http://www.w3.org/Addressing/lesedata\\_23.05.03](http://www.w3.org/Addressing/lesedata_23.05.03))

# TESAURI

For å kunne si noe om forskjellene mellom emnekart og tesauri er det nødvendig å forstå bakgrunnen for henholdsvis nettverksstrukturer og hierarkiske strukturer. For å illustrere dette er tesauri valgt som sammenligningsgrunnlag på grunn av sin hierarkiske ordning.

En tesaurus er et kontrollert vokabular og brukes i gjenfinning til indeksering og søking. Tesaurusen hjelper “indexers and searchers consistently to choose the same terms for expressing a given concept or combination of concepts” (Clarke 2001). Formålet med kontrollerte vokabularer er som nevnt at den som søker etter dokumenter skal bruke samme kombinasjon av termer som den som indekserer dem.

Tesauri defineres som et indekseringsspråks kontrollerte vokabular, og består av termer som er organisert slik at de semantiske relasjonene mellom dem vises eksplisitt (ISO 2788). Et indekseringsspråk er et kontrollert sett av termer som er valgt fra naturlig språk. Termene i en tesaurus brukes til å beskrive emner. På denne måten blir en tesaurus en liste med termer som fungerer som identifikatorer for de underliggende begrepene (Clarke 2001).

Konstruksjon av tesauri baseres hovedsakelig på to ISO-standarder. ISO 2788 gir generelle retningslinjer for tesauruskonstruksjon, og gjelder både for énspråklige og flerspråklige tesauri. ISO 5964 kan sies å være en utvidelse av ISO 2788 da den tar opp elementer spesifikt knyttet til tesauri på mer enn

ett språk. Denne oppgaven tar utgangspunkt i ISO 2788 da fokuset i den videre framstillingen er énspråklige tesauri.

I ISO 2788 gis det retningslinjer for hvilke termer som skal velges, formen på dem, hvilke relasjoner som skal realiseres og hvordan de skal uttrykkes. Tesaurusrelasjoner behandles i neste avsnitt. Termene representerer flere begrepskategorier. Disse begrepskategoriene er konkrete enheter, materialer, handlinger og hendelser. I tillegg er abstrakte enheter, egenskaper ved ting, materiale og handlinger, fagområder, måleenheter, og individuelle enheter, begrepskategorier som dekkes av standarden. Videre gir ISO 2788 retningslinjer for termenes form. Generelt skal det brukes substantiver i ubestemt form. Verb representeres ved substantiver, og adjektiver bør ikke brukes alene. Dersom det er mulig å velge mellom flere termer, som for eksempel norske ord og fremmedord, gjelder det å ta hensyn til hva som er best for brukerne av tesaurusen.

Regelverket som gjelder for tesauri er med på å sikre vokabularkontroll. Ved vokabularkontroll får termer en fast form og relasjonene uttrykkes konsekvent. Reglene fører til konsekvent utforming og bruk, noe som vil lette gjenfinningen.

## 4.1 Tesaurusrelasjoner

En tesaurus viser semantiske relasjoner mellom termer. Denne typen relasjoner kalles analytiske eller *a priori* relasjoner.

Disse relasjonene viser hvordan termene er semantisk relatert til hverandre, og er dokumentuavhengige. Dette innebærer at relasjonene eksisterer uavhengig av om de blir behandlet i dokumenter. Analytiske relasjoner er allment akseptert og understøttes av tradisjonelle referanseverk. Motsetningen er syntetiske relasjoner som er dokumentavhengige. Det vil si at de bare eksisterer dersom de behandles i et dokument. En syntetisk relasjon er en formell relasjon som eksisterer som en potensiell mulighet i indekseringsspråket og som kan realiseres i indeksering. Syntetiske relasjoner er relasjoner som til sammen uttrykker emnet for et dokument. Disse relasjonene kan uttrykkes ved emneord i streng eller ved å kombinere termer ved boolske operatører i søking. Syntetiske relasjoner uttrykkes ikke i en tesaurus.

Hensikten med relasjoner er å hjelpe brukerne. Relasjonene viser i hvilken sammenheng en term er brukt, og kan lede brukeren til andre termer som supplerer eller erstatter den første. “If the terms are viewed as nodes in a semantic network, then the relationships serve as a complex pathway or navigation guide through the network.” (Clarke 2001). Tesaurusrelasjoner blir med dette også et navigasjonsverktøy i et semantisk nettverk innenfor et gitt fagområde, ettersom de viser sammenhengen en term er brukt i. Termen menneskekroppen kan for eksempel brukes både i sammenheng med helse og kunst, og her vil da perspektivet for tesaurusen avgjøre

sammenhengen. Er det en kunsttesaurus vil sammenhengen være kunst. Slik kan en tesaurus være med på å bedre forståelsen av et fagområde ved å gi semantiske kart som viser hvordan ulike begreper er relatert til hverandre. Perspektivet bestemmes av fagområdet og brukergruppen tesaurusen gjelder for.

Det er tre typer tesaurusrelasjoner; ekvivalensrelasjoner, hierarkiske relasjoner og assosiative relasjoner. For de ulike relasjonstypene er det i ISO 2788 utarbeidet en terminologi bestående av prefikser som skal brukes for de ulike relasjonene. Noen av disse prefiksene blir brukt i den videre framstillingen.

Relasjonene som er behandlet her gjelder semantiske lenker mellom individuelle termer. Dette kalles mikronivået. I tillegg finnes det relasjoner på makronivå. Disse ser på hvordan termene og lenkene mellom dem forholder seg til den overordnede strukturen for fagfeltet som tesaurusen dekker (Aitchison 2000). Da det ikke er denne oppgavens formål å studere tesauri innen et spesifikt fagområde, blir ikke relasjonene på makronivået behandlet.

### 4.1.1 Ekvivalensrelasjoner

Ofte er det slik at et enkelt begrep kan representeres ved flere forskjellige termer. Disse termene kalles for synonymymer og blir behandlet som ekvivalensrelasjoner. “This is the relationship between preferred and non-preferred term where two or more terms are regarded, for indexing purposes, as referring

to the same concept.” (ISO 2788). Ved synonymymer velges den ene termen som foretrukket foran den/de andre. Brukerne henvises fra den ikke foretrukne termen til den foretrukne.

Prefiksene BRUK og BF (brukt for) brukes for ekvivalensrelasjoner. Et eksempel på en slik relasjon er graviditet og svangerskap:

- graviditet – BRUK svangerskap
- svangerskap – BF graviditet

I praksis finnes det flere typer av synonymymer. ISO 2788 lister opp noen av de vanligste synonymtypene:

- Termer med forskjellig opprinnelse
- Populærnavn og vitenskapelige navn
- Vitenskapelig/populærnavn og handelsnavn
- Eldre og nyere termer
- Ulik stavemåte og forskjellige navneformer
- Forkortelser og fullstendige termer

Kvasisynonymer, eller nærsynonymer, blir også behandlet som ekvivalensrelasjoner. Dette er termer som i utgangspunktet representerer forskjellige begrep, men som blir behandlet som synonymymer av hensyn til indekseringen. Dette gjøres blant annet for å redusere vokabularet. Hvilke termer som betraktes som kvasisynonymer kan imidlertid avhenge av fagområdet tesaurusen dekker. Eventyr og sagn kan i enkelte generelle tesauri behandles som kvasisynonymer, men ikke i tesauri som spesifikt behandler norsk folkediktning.

De fleste kvasisynonymer er enten termer med overlappende betydning eller antonymer. Et eksempel på termer med overlappende betydning er urbane strøk og byer. Antonymer, eller motsetninger, er termer som representerer ulike synspunkt på samme sak. Eksempler på dette kan være høyde/dybde og likhet/ulikhet. Dersom det er sannsynlig at et dokument som omtaler likheter også vil omtale ulikheter, kan en velge å behandle disse som kvasisynonymer. Det blir da, på samme måte som for synonymymer, laget henvisning fra den ikke foretrukne termen til den termen som er foretrukket.

Ekvivalensrelasjoner gjelder i hovedsak for synonymymer og kvasisynonymer. I tillegg kan også hierarkiske relasjoner behandles som ekvivalensrelasjoner i visse tilfeller. Dette gjelder når den ikke foretrukne termen har en snevrere betydning enn den foretrukne termen. Denne teknikken kalles *upward posting* og behandler spesifikke termer som synonymymer til bredere termer. *Upward posting* brukes for å redusere antall termer i indekseringsspråket, men ISO 2788 anbefaler å unngå bruk av denne løsningen.

Ekvivalensrelasjoner sørger for at termer som uttrykker det samme begrepet blir samlet. En term blir valgt som representant for begrepet, og de andre termene sørger for flere innganger i gjenfinningssystemet, og leder brukere og indekserere til den foretrukne termen. Med ekvivalensrelasjoner unngår

man derfor at dokumenter som handler om det samme blir spredt på forskjellige emneord (Soergel 1974).

#### 4.1.2 Hierarkiske relasjoner

ISO 2788 sier at hierarkiske relasjoner er basert på “degrees or levels of superordination and subordination, where the superordinate term represents a class or whole, and subordinate terms refer to its members or parts” (ISO 2788). Denne relasjonstypen eksisterer altså mellom termer hvor den ene representerer et snevrere begrep enn den andre. Termene skal tilhøre samme begrepskategori.

Hovedhensikten med disse relasjonene er å gjøre brukeren oppmerksom på alternative termer som er relatert til begrepet, men på et annet spesifisitetsnivå. I tillegg klargjør hierarkiske relasjoner i hvilken sammenheng en term blir brukt i. Prefiksene OT (overordnet term) og UT (underordnet term) brukes for hierarkiske relasjoner. Et eksempel på en hierarkisk relasjon vil være:

- psykologi – UT kognitiv psykologi
- kognitiv psykologi – OT psykologi

For å ordne termer i hierarkier brukes inndelingskriterier. For eksemplet over kan inndelingskriteriet være etter *fag*. I en systematisk tesaurus kan kriteriene for hvordan inndelingen er gjort også tas med. Inndelingskriteriet er en merkelapp og inngår ikke i relasjonene. Det er viktig å bruke bare ett inndelingskriterium om gangen. Enkelte begreper passer imidlertid inn i flere inndelingskriterier, og en får da en polyhierarkisk relasjon hvor en term får flere overordnede termer.

Det er tre typer hierarkiske relasjoner; generiske, partitive og attributive. *Generiske relasjoner* kan identifiseres relasjonen mellom en klasse og dens medlemmer (Aitchison 2000). Medlemmene i klassen arver egenskapene til denne. Generiske relasjoner brukes for typer av handlinger, egenskaper, utøver og entiteter. For å identifisere disse relasjonene kan en bruke termene alle og noen; alle mennesker er pattedyr, noen pattedyr er mennesker.

- pattedyr – UT mennesker
- mennesker – OT pattedyr

*Partitive relasjoner* er relasjoner som innebærer helhet-del. Det er tillatt med fire typer partitive relasjoner:

- systemer og organer i kroppen
  - øre – det ytre øret
  - øregangen

- geografiske steder
  - Norge – Oslo
  - Grünerløkka

- disipliner og fagområder
  - psykologi – kognitiv psykologi

- hierarkiske sosiale strukturer
  - Stortinget – Odelstinget
  - Lagtinget

Hvor spesifikk tesaurusen skal være, samt fagområdet og brukergruppen vil være med på å avgjøre antall nivåer i hierarkiet.

*Attributive relasjoner* er relasjoner mellom en generell kategori av ting eller hendelser, og en individuell forekomst av kategorien. I ISO 2788 heter det at denne relasjonstypen “identifies the link between a general category of things and events, expressed by a common noun, and an individual instance of that category, the instance then forming a class-of-noe which is represented by a proper name.” Et eksempel på dette er:

- Hav – Atlanterhavet
- Svartehavet
- Middelhavet

#### 4.1.3 Assosiative relasjoner

Noen begrep er sterkt assosiert med hverandre og lenken mellom dem er derfor så sterk at de bør uttrykkes i en tesaurus. Termene som representerer disse begrepene er ikke ekvivalente og kan ikke plasseres i hierarkier. Denne typen relasjoner kalles assosiative relasjoner og uttrykkes som sideordninger i en tesaurus.

Hensikten med assosiative relasjoner er å vise brukeren til relaterte termer. Brukeren blir gjort oppmerksom på termer som kan brukes i stedet for eller sammen med termen som ble brukt i utgangspunktet. Det finnes ingen klar definisjon av assosiative relasjoner, og det blir derfor ofte opp til den enkelte tesaurusredaktør å avgjøre hva som best vil hjelpe brukerne.

I ISO 2788 fremheves viktigheten av å ha kontroll av denne typen relasjoner. Assosiative relasjoner lenker sammen termer som både tilhører samme begrepskategori, og som ikke tilhører samme begrepskategori. I det første tilfellet er termene i slekt med hverandre og har overlappende betydninger, som for eksempel båter og skip. I det andre tilfellet gir ISO 2788 en liste over situasjoner hvor det er vanlig å realisere assosiative relasjoner. Denne listen er ikke fullstendig og gir bare eksempler på ulike situasjoner.

- a) et fagområde og dets studieobjekt
- b) operasjon/prosess og dens instrument
- c) operasjon og dens produkt
- d) handling og dets objekt
- e) begreper relatert til deres egenskaper

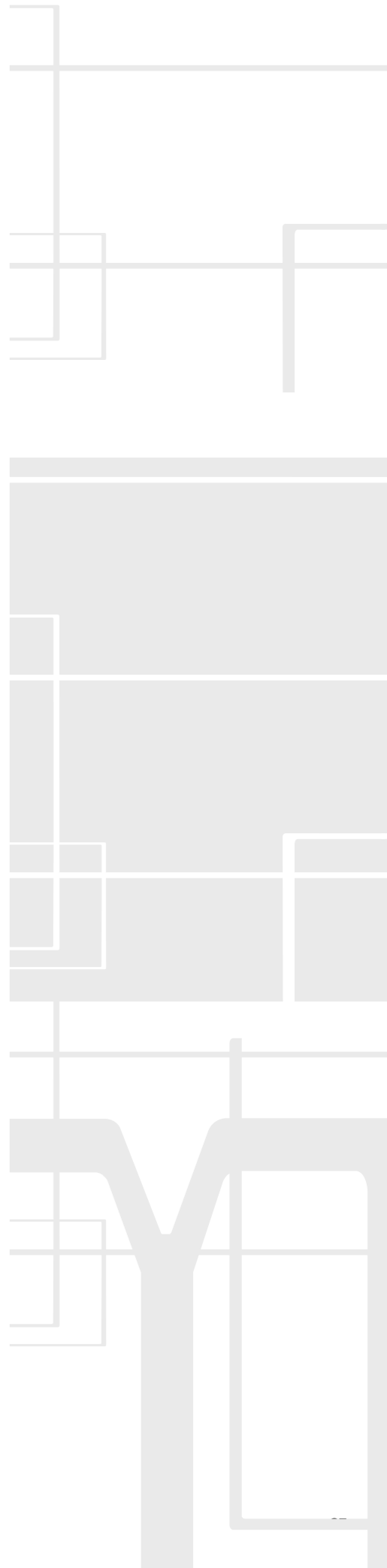
- f) begreper relatert til deres opprinnelse
- g) begreper lenket ved årsaksavhengighet
- h) en ting og dens motpart
- i) et begrep og dets målenhet
- j) helhet/del

(Hentet fra ISO 2788)

#### 4.1.4 Idéplan og verbalplan

I tesaurisammenheng har man relasjoner mellom idéplanet og verbalplanet. På idéplanet er begrepene, mens verbalplanet inneholder termene som skal representere disse begrepene (Opsahl 1987). Det oppstår problemer når det ikke er én-til-én relasjon mellom de to planene. Homonymi og synonymi fører til mange-til-én relasjoner og én-til-mange relasjoner mellom idéplanet og verbalplanet. Synonymer dekkes av ekvivalensrelasjoner, mens det er utviklet en egen teknikk for å håndtere homonymer. For å skille mellom termer som skrives likt, men som representerer forskjellige begrep, brukes kvalifikatorer. Dette innebærer at man gir en forklaring på sammenhengen termen brukes i. Forklaringen settes i parentes bak termen, for eksempel horn (bakverk) og horn (instrument).

Homonymi og synonymi er verbale fenomen og ekvivalensrelasjonene eksisterer derfor bare på verbalplanet. Generiske, partitive og attributive relasjoner eksisterer mellom begreper på idéplanet og uttrykkes som hierarkiske relasjoner på verbalplanet. Assosiative relasjoner er assosiative lenker mellom begreper på idéplanet og uttrykkes som sideordninger på verbalplanet.



# EMNEKART OG TESAUURI

## 5.1 Teoretisk sammenligning av emnekart og tesauri

I oppgaven er det så langt forsøkt å gi en innføring i emnekart og emnekartstandarden, samt en innføring i teorien rundt tesauruskonstruksjon. Dette kapittelet tar for seg forskjellene mellom emnekart og tesauri. Kapittelet er konsentrert om en teoretisk sammenligning av emnekart og tesauri. Dette for å se hvilken betydning bakgrunn og formål har hatt for utforming og anvendelse av de to modellene. Det blir også tatt opp hvordan modellene løser ulike problemer. Utgangspunktet for sammenligningen er standardene modellene bygger på, og forskjellene mellom disse.

Den teoretiske sammenligningen kan oppsummeres som vist i tabellen til høyre.

### 5.1.1 Bakgrunn

Emnekart og tesauri har forskjellig bakgrunn og er utviklet i to forskjellige tradisjoner. Tesauri er et eldre verkøy enn emnekart, og har lange tradisjoner i kunnskapsorganisasjon. Tesauri har vært i bruk lenge før internett og World Wide Web, og ble i utgangspunktet publisert i trykte utgaver. Etter lanseringen av World Wide Web er mange tesauri gjort tilgjengelige via internett. Den største forandringen i dette er navigasjonsmulighetene som hyperlenkene sørger for. I tillegg er tesaurusene i større utstrekning tilgjengelige for brukerne.

Emnekart ble utviklet i datamiljøet av systemutviklere. I utgangspunktet

var målet å skape et verktøy som gjorde det enkelt å slå sammen indekser. Dette ble utviklet, og emnekart ble et verktøy til å organisere internettressurser gjennom et emnebeskrivende lag uten å berøre de underliggende ressursene. Emnekart er laget som et publiseringsverktøy for intra- og internett, og er et gjenfinningssystem som særlig støtter navigasjon.

### 5.1.2 Formål og funksjon

Formålet med tesauri er å lede indekserere og brukere til å velge de samme termene i henholdsvis indeksering og søking. På den måten bedres gjenfinningen ved at de som beskriver dokumentene konsekvent bruker de samme

termene til å beskrive innholdet i dokumentene, og ved at de som søker etter dokumenter bruker de samme termene i søkeuttrykket som indeksererne. Det er altså viktig hvordan dokumenter blir beskrevet for å sikre at det blir gjenfunnet når brukeren kan ha bruk for det. En tesaurus kan med bakgrunn i dette sees på som en del av et gjenfinningssystem, eller som et hjelpemiddel i gjenfinningssystemet.

Emnekart sies å være et gjenfinningssystem i seg selv. Emnekart inneholder søkemuligheter, navigasjonsmuligheter, og mulighet for å knytte forekomster opp mot emner, noe som kan sees på som innholdsbeskrivelse av forekomstene.

	Emnekart	Tesauri
<b>Bakgrunn</b>	Datamiljø, semantic web	Kunnskapsorganisasjon
<b>Formål</b>	Hjelpe brukere til å navigere på et nettsted ved hjelp av emner og assosiasjoner	Hjelpe indekserere og brukere av et gjenfinningssystem
<b>Funksjon</b>	Navigering	Indeksering og søking
<b>Struktur</b>	Nettverk	Hierarki
<b>Emne</b>	Alt må defineres som emne for å kunne brukes i et emnekart	Det et dokument handler om
<b>Relasjoner</b>	Alle typer relasjoner er tillatt	Hierarkiske, ekvivalens- og assosiative relasjoner
<b>Språklige fenomen</b>	Perspektiv, temaidentitet	Ekvivalensrelasjoner, kvalifikatorer
<b>Standarder</b>	ISO 13250; er en standard for implementasjon av emnekart, inneholder ikke regler for relasjoner og termer	ISO 2788; inneholder regler for relasjoner, realisering av disse, hvilke termer som skal tas med og hvilken form de skal ha

### 5.1.3 Struktur

De to modellene har forskjellig struktur som kommer av at de bygger på to ulike standarder. Tesaurusstandarden tar utgangspunkt i en hierarkisk struktur, mens emnekartstandarden tar sikte på en nettverksstruktur. Tesauri kan med sin hierarkiske struktur brukes til å “provide a hierarchical browsing capability” (Aitchinson 2000). Tesauri gir mulighet for navigering i hierarkier, det vil si til bredere og snevrere termer, men tillater også navigering mellom hierarkier ved en begrenset bruk av assosiative relasjoner. Emnekart på sin side har en utstrakt bruk av assosiative relasjoner og oppnår med dette en nettverksstruktur som ikke binder navigasjonen til hierarkier. Utbredt bruk av assosiative relasjoner gir mulighet for navigering på tvers av hierarkier.

### 5.1.4 Emner

Emner står sentralt i begge modellene. Det er imidlertid forskjellig hva som anses som emner i henholdsvis emnekart og tesauri. I emnekart kan alt defineres som emne, og i praksis må alle elementer defineres som emner før de kan brukes. Det betyr at alle assosiasjoner, assosiasjonstyper, forekomsttyper osv. er definert som emner. I tesauri er et emne det som dokumentene handler om, og uttrykkes ved termer som representerer begreper. Dette kan også være emner i emnekart. Eksemplet med *Steven Pinker* og boka *The blank slate* illustrerer dette. Her vil *kognitiv psykologi* og

eventuelle undertemaer til dette være det som regnes som emner i en tesaurus, mens både *Steven Pinker* og bokanmeldelsen i tillegg kan være emner i et emnekart.

### 5.1.5 Relasjoner

Relasjoner er et annet sentralt element i modellene. I emnekart er det ingen begrensninger når det gjelder relasjoner. Alle relasjoner kan tas med og det er ingen regler for hvordan relasjonene skal uttrykkes. I tesauri er det kun tillatt med tre relasjonstyper, og tesaurusstandarden inneholder i tillegg regler for hvordan disse skal uttrykkes. Emnekart dekker også relasjoner som det ikke er nødvendig å ta med i en tesaurus da disse dekkes av andre verktøy i kunnskapsorganisasjon, som katalogisering. Relasjonen at *Steven Pinker* har skrevet boka *The Blank Slate* vil ikke være med i en tesaurus, men kan tas med i et emnekart. Relasjoner i tesauri gjelder innholdsbeskrivelser av dokumenter, og viser formelle relasjoner mellom termer. Emnekart dekker både innholdsbeskrivelse og dokumentbeskrivelse. Begge modellene beskriver et kunnskapsunivers.

### 5.1.6 Språklige fenomen

Språklige fenomen som homonymi og polysemi håndteres i emnekart ved å ta i bruk perspektiv. Perspektiv angir i hvilken sammenheng et emnenavn eller en forekomst er gitt til et emne (se kapittel 3). I tesauri brukes kvalifikatorer. Syno-

nymer håndteres i emnekart av temaidentitet og temaidentifikasjon, mens tesauri bruker ekvivalensrelasjoner. Perspektiv brukes også i tesaurisammenheng, men da om fagområdet tesaurusen dekker. Tesauri har ikke mulighet for å definere flere perspektiv slik som emnekart har. Perspektivet for en tesaurus vil være den sammenhengen som tesaurusen er laget i, det vil si det fagområdet den dekker. Perspektivet vil således være avhengig av både kunnskapsområdet som dekkes og av brukerne.

### 5.1.7 Standardene

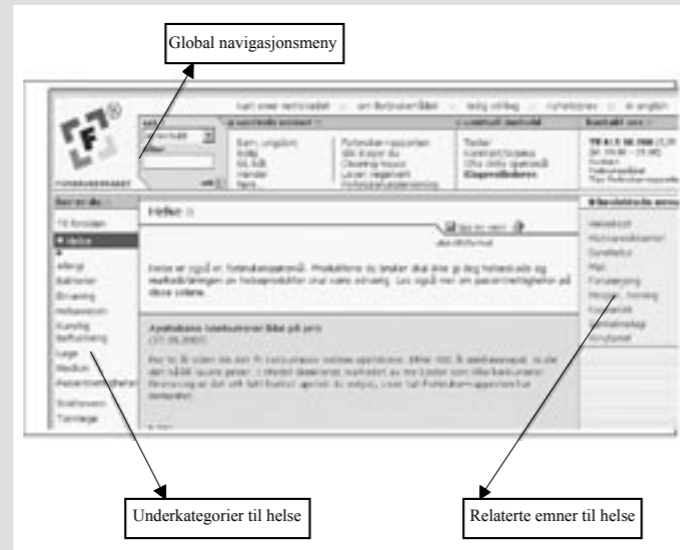
Tesaurusstandarden inneholder klare regler for hvilke termer som skal være med i en tesaurus, og hvordan termene skal utformes. Hovedregelen er at substantiver som kan telles uttrykkes i ubestemt form, flertall, mens substantiver det bare finnes ett av skal skrives i bestemt form, entall. I emnekart brukes verb og preposisjoner for å uttrykke relasjoner. Emnekartstandarden inneholder ikke regler for utforming av emner og assosiasjoner.

Reglene i tesaurusstandarden sikrer en konsekvent utforming av verktøyet. Målet er at alt om et emne skal havne på samme sted, slik at brukerne ikke blir forvirret. Likevel dekker ikke reglene alle mulige situasjoner, og det gis derfor rom for skjønn ut i fra kunnskap om hva som er best for den aktuelle brukergruppen.

Emnekartstandarden er i tillegg en standard for hvordan emnekart skal



Skjermbildet for helsekategorien, forskning.no



Forbrukerportalen.no

implementeres. Emnekart skal uttrykkes i et format som sikrer utveksling av informasjon blant maskiner.

## 5.2 Emnekart i praksis

For å få en bedre forståelse av hvordan emnekart fungerer i praksis tar resten av kapittelet for seg en mer praktisk sammenligning av emnekart og tesauri. Utgangspunktet for sammenligningen er to nettstedene som har implementert emnekart. Nettstedene er analysert og resultatene av analysen blir lagt fram. Funnene fra analysen blir også sammenlignet med tesaurusstandard på den måten at funnene blir holdt opp mot løsninger i standarden. Dette kapittelet inneholder i tillegg en kort presentasjon av nettstedene.

Formålet med analysen var å se hvordan emnekart fungerer i praksis. Dette innebærer at nettstedene er analysert med utgangspunkt i emnekartstandard. Analysen er ikke en analyse av selve nettstedene i den forstand at løsninger hos det enkelte nettsted blir vektlagt, eller at organisering og layout hos nettstedet blir gjennomgått. Analysen konsentreres om funn som kan si noe om hvordan emnekart fungerer når de er implementert i et nettsted, det vil si å se om det finnes eventuelle problemer knyttet til bruken av emnekart. I navigasjonen som ligger til grunn for analysen, var det brukerens situasjon som stod i fokus.

Utgangspunktet for analysen har vært samlokaliseringssprinsippet. I følge

Steve Pepper er samlokaliseringssprinsippet en metode for samlokalisering av all viten om et gitt tema (BRODD 2003). Samlokaliseringssprinsippet står sentralt i emnekartstandard, og går ut på at alt om et emne skal kunne finnes på ett og samme sted selv om brukeren benytter ulike veier eller innfallsvinkler for å komme til emnet. Ved navigering gjennom de to nettstedene var det derfor viktig å se etter løsninger eller problemer som ville ødelegge for dette.

I tillegg ble det også sett etter momenter som kunne gjøre navigasjonen problematisk for en bruker. For at navigasjonen skal fungere er det viktig at brukeren underveis får hjelp til å vite hvor hun skal starte og hvordan hun bør gå videre. Dette innebar at relasjoner, emner og hvordan ulike emner var gruppert sammen ble vektlagt. I tillegg bør det også være en viss forutsigbarhet i den forstand at emner med lik spesifisitet befinner seg på samme nivå i nettverket.

Analysen dreier seg om emnene på nettstedene. Som tidligere vist har emne forskjellig betydning i henholdsvis kunnskapsorganisasjon og emnekart. Når emne blir brukt i resten av dette kapittelet, vil det være i betydningen av hvordan man oppfatter emne innen kunnskapsorganisasjon. Det vil med andre ord si at kapittelet fokuserer på det som kalles tema i emnekartsammenheng.

Funnene fra analysen blir diskutert og sett i lys av emnekartstandard.

Diskusjonen tar sikte på å finne ut om problemene som er funnet i analysen skyldes emnekartstandard eller om de er redaksjonelle. For å kunne avgjøre om problemene skyldes emnekartstandard vil problemene holdes opp mot tesaurusstandard. Dette vil vise om de samme problemene hadde oppstått dersom en tesaurus hadde vært grunnlaget for nettstedene. Dette oppsettet av analysen vil da forhåpentligvis avdekke om problemene skyldes emnekart eller om det er problemer av mer universell karakter.

Utgangspunktet for analysen er nettstedene forskning.no og Forbrukerportalen.no. Begge har tatt i bruk emnekart, og begge er anerkjente offentlige nettsteder. Nettstedene har forholdsvis mange emne kategorier. Begge nettstedene har en emne kategori som kalles helse, og denne kategorien egnet seg derfor best, ettersom den ville være med på å illustrere hvordan forskjellige emnekart kan brukes på samme emneområde. Helsekategoriene fra Forbrukerportalen.no og forskning.no har dessuten en sentral rolle i kapittel 6 hvor sammenfletting av emnekart blir behandlet.

### 5.2.1 forskning.no

Forskning.no ble initiert av Norges forskningsråd og eies av ulike forsknings- og utdanningsinstitusjoner i Norge. Nettstedet ble lansert 10. april 2002 og formidler norsk og internasjonal forskning. Målgruppen er hele befolkningen med vekt på ungdom fra 15 år og oppover (forskning.no 2003).

Forskning.no har en distribuert redaksjon, og kan endre emnekartet etter behov. Nettstedet består av en sentral redaksjon i Oslo, et utstrakt frilansnettverk og andre bidragsytere. Eierinstitusjonene til forskning.no har direkte tilgang til publiseringssystemet og sender sine saker inn til publisering hos den sentrale redaksjonen.

Nettstedet har åtte hovedkategorier for emner. Disse er kultur, samfunn, helse, miljø, teknologi, hav og fiske, jord og skog, samt naturvitenskap. Hovedkategoriene utgjør en navigasjonsmeny til venstre i skjermbildet, og ligger fast på alle sidene hos nettstedet. Den fungerer dermed som en global navigasjonsbar.

Helsekategorien har 23 underkategorier som igjen har egne underkategorier. Dette fører til at det blir mange nivåer av kategorier i nettverket. Underkategoriene til helse er: psykologi, framtidens medisin, vaksiner, bakterier, bioteknologi, medisinske metoder, sex, medisiner, sykdommer, menneskekroppen, forebyggende helse, funksjonshemming, mat og helse, medisin, vår helse i 2020, kjønn og seksualitet, hel-sepolitikk, alkohol og narkotika, graviditet, mat, røyking, virus samt alternativ behandling. Disse underkategoriene vises i en navigasjonsmeny til høyre i skjermbildet. Denne menyen endrer seg etter hvilket emne man er inne på.

I forskning.no brukes termene undertema, beslektede tema, emne og beslektede fag for å definere ulike emnetyper. Tema slik det blir brukt på nettstedet må ikke forveksles med tema slik termen blir brukt i emnekartstandard. Emne vil, som presisert i innledningen til dette kapittelet, brukes i analysen slik det er definert i kunnskapsorganisasjon.

### 5.2.2 Forbrukerportalen.no

Forbrukerportalen.no er nettstedet til Forbrukerrådet og Forbrukerrapporten og ble lansert 20. juni 2002. Nettstedet ble kåret til et av de ti beste offentlige nettstedene i 2003 (Forbrukerportalen.no 2003). Innholdet på nettstedet er naturlig nok forbrukerrettet. Eksempler på dette er tester utført på forskjellige produkter og faktainformasjon om ulike varer. Nyheter av forbrukerinteresse blir presentert på forsiden. I tillegg er det lagt ut materiale som det er mulig å laste ned, for eksempel standard kontrakter og Forbrukerrapporten.

Nettstedet har søkemuligheter og har en global navigasjonsmeny som ligger fast på alle sidene. Det er menyer både på høyre og venstre side av skjermbildet. Menyene til venstre viser hvordan emnene i de enkelte kategoriene hører sammen, mens menyene til høyre gir en

oversikt over hvilke emner et emne er assosiert med. Menyene vil endre seg etter hvilken artikkel brukeren er inne på, og viser hvilke assosiasjoner artikkelen har.

Nettstedet har 12 hovedkategorier or emner. Hovedkategoriene er Barn/ ungdom, bolig, bil/båt, handel, helse, it/tele/data, kraft/strøm/energi, markedsføring/reklame, mat, priser, tjenester, og økonomi. Hovedkategoriene er igjen delt inn i underkategorier. Helsekategorien er delt inn i undergruppene allergi, bakterier, ernæring, helsevesen, kunstig befruktning, lege, medisin, pasientrettigheter, smittevern og tannlege.

## 5.3 Resultater fra analysen

Under analysen ble det registrert flere problemer hos forskning.no enn hos Forbrukerportalen.no. Dette kan komme av at Forbrukerportalen.no er et mindre nettsted, samt at helsekategorien også er mindre enn den tilsvarende kategorien hos forskning.no. Dette kan også være noe av årsaken til at helsekategorien på Forbrukerportalen.no virker mer oversiktlig. Den var enklere å navigere gjennom, og hadde færre assosiasjoner mellom emnene.

Problemer som kom fram i analysen blir i det videre gjennomgått og illustrert med eksempler fra nettstedene.

### 5.3.1 Bruk av polyhierarkier

Polyhierarkier innebærer at et emne tilhører flere overordnede emner. Dette betyr at brukeren kan finne det samme emnet selv om hun går via forskjellige overordnede emner, og kan likeså henvises fra det aktuelle emnet til forskjellige overordnede emner.

Et eksempel på polyhierarkier hentet fra forskning.no er funksjonshemming som hører til under kategoriene helse og samfunn. Dette kan gjøre det vanskelig å vite hva som er fokus for emnet funksjonshemming. Brukeren kan ikke vite om det er ut i fra et helsemessig eller et samfunnsmessig aspekt at funksjonshemming blir behandlet. Det mest nærliggende for brukeren er å anta at begge aspektene blir i varetatt i denne kategorien. I dette tilfellet er det imidlertid slik at forekomstene som er knyttet opp mot funksjonshemming, som regel tilhører samfunnskategorien, i tillegg til flere andre kategorier, men ikke til helsekategorien.

Bioteknologi er et annet emne fra forskning.no som hører til flere kategorier. Disse er teknologi, naturvitenskap, biologi, og helse. Dette er et eksempel på at en bruker kan ledes til emnet fra ulike innganger.

Ernæring hos Forbrukerportalen.no er også et emne som hører til under to kategorier, nemlig helse og mat. Dette er et eksempel på et polyhierarki som fungerer i og med at dette emnet i virkeligheten er forbundet med både helse og mat.

Helsevesen hører i Forbrukerportalen.no til under tre ulike kategorier; helse, tjenester og offentlige tjenester. Både helse og offentlige tjenester kan sies å ha forbindelse til helsevesen. Emnet tjenester sier imidlertid lite om hva emnet dreier seg om, og hva som eventuelt skiller det fra offentlige tjenester. Brukeren kan ikke vite hvorfor dette emnet er relevant i forhold til helsevesen.

Det siste eksemplet er også hentet fra Forbrukerportalen.no. Her hører smittevern til under landbruk og helse. Dette illustrerer noe av det samme som eksemplet med funksjonshemming. Brukeren kan ikke vite hvordan emnet blir behandlet i forhold til de overordnede emnene, og må gå ut i fra at det dekker begge aspektene. I dette tilfellet er det faktisk slik at forekomstene handler om smittevern både i forhold til landbruk og helse.

### 5.3.2 Ulikt spesifisitetsnivå

Analysen viste også at emner med ulik spesifisitet befinner seg på samme nivå i hierarkiet. Dette kan illustreres ved å se på underkategoriene til helse i Forbrukerportalen.no. Allergi, bakterier, lege og tannlege er emner som har et høyere spesifisitetsnivå enn for eksempel helsevesen og medisin.

Hos forskning.no finnes det også emner av ulike spesifisitet under helsekategorien. Her er for eksempel medisiner, menneskekroppen, bakterier og virus emner som ligger på samme nivå og er direkte relatert til helse.

Kategorien medisiner hos forskning.no illustrerer det samme. Her er Viagra er en underkategori, på samme nivå som bivirkninger og smertelindring. Dette viser store forskjeller i hvor spesifikke emnene er. Det samme gjelder når AIDS, kreft, psykiske lidelser, diabetes, stress, dysleksi og allergi er underkategorier til sykdommer, på samme nivå som plantesykdommer, dyresykdommer og fiskesykdommer. De tre sistnevnte er emner med mindre spesifisitet enn de andre emnene.

### 5.3.3 Assosiasjoner

I tillegg til at plantesykdommer, dyresykdommer og fiskesykdommer er mindre spesifikke enn de andre emnene, er det også vanskelig å se hvorfor disse emnene er plassert under sykdommer. De tre kategoriene angår dyr, mens det ikke er una-

turlig å forvente at *sykdommer* er relatert til mennesker. Utgangspunktet for at disse er tatt med i kategorien *sykdommer* må skyldes de mulighetene som ligger i emnekartstandarden for å uttrykke assosiasjoner. Mens de andre underkategoriene kan sies å være generiske relasjoner til *sykdommer*, er *plantesykdommer*, *dyresykdommer* og *fiskesykdommer* eksempler på assosiative relasjoner.

Når det gjelder assosiative relasjoner kan det være nødvendig å ta hensyn til hvor allmenne disse er. Dersom relasjonene ikke er naturlig for andre enn den som lager dem, er de ikke til mye hjelp på et nettsted. Ettersom *sykdommer* i denne sammenheng er relatert til mennesker og underordnet *helse*, kan man spørre seg om det er naturlig for en bruker å gå via *helse* for å finne noe om sykdommer hos fisk. Dette er et eksempel på hvordan emnekart kan lede en bruker til nye emner hun ikke visste hadde sammenheng med det hun opprinnelig var interessert i.

**5.3.4 Manglende navn på assosiasjoner**  
Det ble identifisert eksempler hvor det ikke er satt merkelapper på assosiasjonene. Dette innebærer at det ikke kommer klart fram hvilken type relasjoner som eksisterer mellom emnene. Dette fører til at det ser ut som om ulike aspekter ved et emne blir behandlet som om de er like. Dette gjelder for eksempel i tilfellet hvor både spesifikke sykdommer og dyresykdommer er underkategorier til *sykdommer*. Det er ingenting som sier om det dreier seg om hierarkiske eller assosiative relasjoner. Relasjonene blir med dette sidestilt.

Under *menneskekroppen* hos forskning.no finnes det også et eksempel på dette. Her er emner som *svangerskap*, *graviditet* og *slanking* underkategorier til *menneskekroppen* på samme nivå som *hjerne*, *hjerne* og *øye*. Blant underkategoriene skilles det ikke mellom lemmer som for eksempel *hender* og *føtter*, og indre organer som *hertet*.

Et annet eksempel hentet fra forskning.no er *framtidens medisin* som er en underkategori til både *helse* og *næringsliv*. Dette er ikke nødvendigvis intuitivt for en bruker. Det må kunne sies å være en assosiativ relasjon mellom *framtidens medisin* og *næringsliv*. Det ville imidlertid være mer naturlig at *næringsliv* var et beslektet emne til *framtidens medisin*. Også her blir assosiasjonene sidestilt.

**5.3.5 Forskjellige begrepskategorier**  
I helsekategoriene for begge nettstedene finnes det blant underkategoriene emner fra forskjellige begrepskategorier. Hos

forskning.no kan *psykologi*, *framtidens medisin*, *bioteknologi*, *medisin* og *alternativ medisin* sies å være fagområder innen helse. De andre underkategoriene tilhører ikke denne begrepskategorien. *Vaksiner* kan for eksempel sies å være et forbyggende helsetiltak eller en behandlingsform. *Bakterier* og *virus* er mikroorganismer som forårsaker *sykdommer* som er et annet emne under *helse*. *Røyking* er en handling som kan forårsake helseskader, mens *medisiner* er behandlingsmetoder.

Hos Forbrukerportalen.no finnes lignende eksempler. Her er *lege* og *bakterier* plassert direkte under *helse*. *Bakterier* er mikroorganismer, mens *lege* sammen med *tannlege* er eksempler på yrkesgrupper. *Allergi* er en type sykdom, mens *pasientrettigheter* er et aspekt ved helsevesenet.

Årsaken til dette kan være at relasjonene ikke er hierarkiske, men heller assosiative. De presenteres imidlertid som hierarkiske og dette illustrerer igjen problemet med assosiasjoner som mangler navn.

**5.3.6 Emnetyper**  
Hos forskning.no er emner definert som forskjellige emnetyper. Emnetyper som blir brukt hos forskning.no er blant annet *fag* og *tema*. *Psykologi* er et emne som er definert både som *fag* og som *tema*. Som bruker er det vanskelig å vite forskjellen på psykologi brukt som fag og psykologi brukt som tema. Det er ingenting som hjelper brukeren til å velge mellom disse. Problemet blir altså at brukeren ikke settes i stand til å velge mellom disse to, og heller ikke får hjelp til å velge den riktige kategorien.

**5.3.7 Synonymer**  
Hos forskning.no dukket det opp problemer knyttet til synonymer. Både *svangerskap* og *graviditet* er emner på nettstedet. Begge er underkategorier til *menneskekroppen*, og i tillegg er *graviditet* en kategori direkte under *helse*, på samme måte som *menneskekroppen*. Dersom man klikker seg inn på henholdsvis *graviditet* og *svangerskap* kommer man til forskjellige sider som inneholder både artikler som finnes begge steder og artikler som bare finnes på et av stedene. Dette ødelegger for samlokaliseringsprinsippet. En bruker er ikke garantert å finne alle artikler om det å vente barn ved å klikke på et av emnene. Dersom brukeren klikker på *graviditet* fra helsesiden er det ikke en gang sikkert at hun får vite at det også finnes en kategori som heter *svangerskap*.

I emnekartstandarden heter det at

perspektiv løser problemer knyttet til synonymer. Det kan være mange årsaker til at *graviditet* og *svangerskap* ikke blir behandlet som synonymer hos forskning.no. For det første kan det hende at det ikke er definert noe perspektiv for nettstedet, og for det andre kan det hende at de to emnene har forskjellig temaidentitet og dermed ikke blir oppfattet som synonymer av emnekartapplikasjonen.

*Ultralyd* under *medisinske metoder* er beslektet med *svangerskap*, og omvendt er *svangerskap* beslektet med *ultralyd*. *Ultralyd* er imidlertid ikke beslektet med *graviditet* og omvendt. Denne inkonsekvente løsningen skyldes kanskje menneskelige feil, men det virker ikke som emnekartet slik det er implementert hos forskning.no har mekanismer for å håndtere slike situasjoner.

## 5.4 Problemene i forhold til standardene

**5.4.1 Emnekartstandarden**  
Dette avsnittet tar opp problemene som kom fram og diskuterer dem i forhold til henholdsvis emnekartstandarden og tesaurusstandarden. Dette blir blant annet gjort for å se om problemene i analysen er direkte knyttet til emnekartstandarden, eller om de er redaksjonelle problemer.

Emnekartstandarden inneholder ikke begrensninger i forhold til assosiasjoner. Dette gjør det mulig at et emne kan være relatert til flere, og overordnede emner. Polyhierarkier kan lett bli konsekvensen av dette. Det blir derfor opp til den som konstruerer emnekartet å bestemme hvorvidt polyhierarkier skal være en godkjent løsning.

Graden av emners spesifisitet, og plassering i hierarki, er ikke behandlet i emnekartstandarden. Den grunnleggende strukturen i emnekart er nettverksorientert og det er derfor ikke interessant hvorvidt emnene skiller seg fra hverandre i spesifisitet. Det viktigste er å koble sammen emner som er assosiert med hverandre uansett grad av spesifisitet.

Det ligger ingen begrensninger i emnekartstandarden for antall assosiasjoner mellom emner, eller hvilke type assosiasjoner som kan realiseres. Dette innebærer utstrakt bruk av assosiative relasjoner. Assosiative relasjoner kan imidlertid lett bli personlige. Det vil si at en assosiasjon kan være intuitiv for en person, mens assosiasjonen ikke er innlysende for andre. I slike tilfeller vil det kunne være vanskelig for brukeren å navigere rundt i emnekartet.

Emnekartstandarden krever ikke at

det skal settes navn på assosiasjonene. Konsekvensen blir at assosiasjoner ikke får navn, som igjen fører til at brukeren ikke får vite hvordan emnene er relatert til hverandre. Det er igjen opp til den som utformer emnekartet å bestemme dette.

Begrepskategorier blir ikke behandlet i emnekartstandarden. Standarden er åpen og gir flere muligheter for utforming av emnekart, og gir derfor stor frihet til den som lager emnekartet. I og med at det ikke finnes krav eller restriksjoner med hensyn på begrepskategorier faller ansvaret for dette på den som utformer emnekartet.

Emnetyper er et element i emnekart og standarden legger ingen begrensninger for hvor mange emnetyper et emne kan tilhøre. Derfor vil ikke en emnekartapplikasjon fange opp at et emne defineres som to emnetyper. Dette ble blant annet illustrert i forskning.no hvor psykologi er definert både som fag og tema. Standarden gir heller ikke en løsning for hvordan brukeren skal kunne skille mellom ulike emnetyper. Det gis med andre ord ingen hjelp til brukeren i forhold til navigering mellom de ulike emnetyperne.

Emnekartstandarden tar i bruk temaidentitet for å håndtere synonymer. Imidlertid legger ikke standarden opp til at temaidentitet og temaindikatorer skal være obligatoriske. Dette kan føre til situasjoner hvor synonymer ikke blir oppfattet som synonymer. Dette ble illustrert med eksemplet om graviditet og svangerskap som ikke ble oppfattet som synonymer.

Emnekartstandarden inneholder altså få begrensninger med hensyn til emner og assosiasjoner. Dette får konsekvenser for blant annet emnetyper, navn på assosiasjoner, samt håndtering av synonymer. Friheten som standarden åpner for innebærer at mye av ansvaret for at navigeringen skal fungere for brukeren faller på den som utformer emnekartet.

**5.4.2 Tesaurusstandarden**  
Tesaurusstandarden åpner også for bruk av polyhierarkier. I tesauruskonstruksjon brukes inndelingskriterier for å ordne termer i hierarkier og ofte vil det være slik at et emne passer til flere inndelingskriterier, og dermed oppstår polyhierarkiske relasjoner. Standarden fraråder imidlertid utstrakt bruk av denne løsningen.

I en tesaurus er det hierarkiske relasjoner som viser relasjoner mellom termer med forskjellig spesifisitetsnivå. Overordnede termer har en mindre grad

av spesifisitet enn underordnede termer. Termer på samme nivå i hierarkiet vil imidlertid være av samme spesifitet. Dette gir en viss forutsigbarhet. Tesaurusstandarden åpner imidlertid for å kutte ut ledd for å unngå veldig dype hierarkier.

Tesaurusstandarden opererer med relasjoner i stedet for assosiasjoner. Standarden inneholder regler for hvilke relasjoner som skal realiseres og hvordan. I tesauri er det tillatt med tre typer tesaurusrelasjoner; ekvivalensrelasjoner, hierarkiske og assosiative relasjoner. Standarden foreslår at det kun er et bestemt utvalg av assosiative relasjoner som bør realiseres.

Tesaurusstandarden opererer også med navn på relasjonene. Med betegnelsene *overordnet term* og *underordnet term* angis det at termene er hierarkisk relatert til hverandre. Assosiative relasjoner angis med se også, mens ekvivalensrelasjoner angis med *bruk* og *brukt for*. Inndelingskriterier sier også noe om emnene og hvordan de står i forhold til hverandre.

I tesaurusstandarden heter det at termer som inngår i hierarkiske relasjoner skal tilhøre samme begrepskategori. Det ligger derfor i disse relasjonenes natur at termene må være fra samme begrepskategori.

I motsetning til emnekartstandarden operer ikke tesaurusstandarden med emnetyper. Inndelingskriteriene kan imidlertid fungere som emnetyper ved at de forteller hvordan emnene er sortert og hva de har felles.

Synonymer håndteres i tesaurusstandarden ved ekvivalensrelasjoner. Dersom *graviditet* og *svangerskap* hadde vært termer i en tesaurus ville den ene termen bli foretrukket og det ville ha vært en henvisning fra den ikke foretrukne til den foretrukne (se eksempel s. 29).

Tesaurusstandarden inneholder mange regler for utforming av tesauri. Dette fører til konsistente løsninger, men legger begrensninger på hva som kan uttrykkes i en tesaurus. Dette gjelder særlig assosiative relasjoner.

# SAMMENFLETNING AV EMNEKART

I emnekartstandarden er det mulighet for å flette sammen ulike emnekart. Denne sammenflettingen innebærer at forskjellige emnekart kobles sammen slik at de danner ett felles emnekart. Dette kapitlet går nærmere inn på dette, ved først å gi en teoretisk beskrivelse av metoder som da benyttes. Deretter følger en teoretisk drøfting rundt tekniske aspekter ved sammenfletting. For å få en bedre forståelse av hvilke konsekvenser en sammenfletting får for brukeren, er det også gjort et forsøk på å flette sammen to ulike emnekart. Til slutt vil resultatene av dette bli diskutert.

## 6.1 Sammenfletting

Bakgrunnen for sammenfletting av emnekart ligger i samlokaliseringssprinsippet, et prinsipp som står sentralt i emnekartstandarden. Ettersom flere emnekart kan inneholde emner som representerer det samme temaet er det ønskelig å kunne samle disse slik at det blir mulig for en bruker å finne alt om et emne på samme sted. Ved sammenfletting av emnekart får brukeren dermed tilgang til flere emner, assosiasjoner og forekomster.

Sammenflettingen skjer ved at emnekartene som skal flettes sammen blir lagt oppå hverandre. Duplikater av emner og assosiasjoner blir så fjernet. Like emner blir deretter flettet sammen slik at de danner ett emne som arver karakteristikkene fra hvert av de opprinnelige emnene.

Emner vil også i dette kapitlet bli brukt slik emner er definert i kunnskapsorganisasjon dersom ikke annet er presisert. Det betyr at emner i emnekart-sammenheng da vil representere et tema.

Som vist i kapittel 3 eksisterer emner og assosiasjoner i et eget lag atskilt fra forekomstene. Det er imidlertid ingen begrensninger for hvordan emner kan knyttes sammen ved hjelp av assosiasjoner og dette gir stor frihet til å lage nye assosiasjoner mellom emner. På den måten vil emnene inngå i flere assosiasjoner. Det er denne delingen mellom kunnskapslaget og ressurslaget som gjør det mulig å flette sammen ulike emnekart. Nettopp fordi emnene og assosiasjonene beskrives uavhengig av forekomstene kan emner være forbundet på tvers av ulike emnekart, og med mange ulike sett av informasjonsressurser. På samme måte lar ressursnivået seg beskrive med forskjellige emnekart. Ut i fra dette vil det kunne oppstå situasjoner hvor ulike emnekart inneholder emner som representerer det samme temaet (XTM 1.0 2001), og som det derfor vil være aktuelt å flette sammen.

Som tidligere nevnt har emner tre typer karakteristikk; navn, forekomster og roller i assosiasjoner. Det sammenflettede emnet vil da få summen av karakteristikkene, unionen av karakteristikkene, fra de to emnene. Det sammenslåtte emnet vil også erstatte de andre to emnene dersom noen av disse er definert som medlemmer i et eller flere perspektiv.

Sammenfletting av to emnekart kan illustreres ved eksemplet på neste side. Emnekartene A og B dekker hvert sitt område, men inneholder samtidig emner som er like. Som tegningene viser inneholder emnekart A fire emner og assosiasjoner mellom disse. Emnene er igjen koblet til ressurser. Det samme gjelder for emnekart B. Det er ikke definert noe perspektiv for emnekartene og begge befinner seg derfor i det ubegrensede perspektiv. Emnekartene har ett felles emne, nemlig E3. Når emnekartene flettes sammen til emnekart AB vil E3 og E7 bli flettet sammen. Resultatet blir da at emnet Eab3 vil inneholde de samme assosiasjonene som både E3 og E7 inneholdt, samt være knyttet til alle forekomstene for begge de opprinnelige emnene.

Eksempelet kan konkretiseres ved de to nettstedene forskning.no og Forbrukerportalen.no hvor helse er definert som emne i begge emnekartene. Dersom emnekartene for de to nettstedene flettes sammen vil resultatet bli en helsekategori som inneholder alle navn, forekomster og roller i assosiasjoner fra begge nettstedene. Det betyr at helse i det nye, sammenflettede emnekartet vil inneholde alle underkategoriene fra helse i forskning.no og fra helse i Forbrukerportalen.no. I tillegg vil det sammenflettede helseemnet være med i alle assosiasjoner som begge helseemnene i utgangspunktet var med i. Det sammenflettede emnet vil også være med i de ulike perspektivene som de opprin-

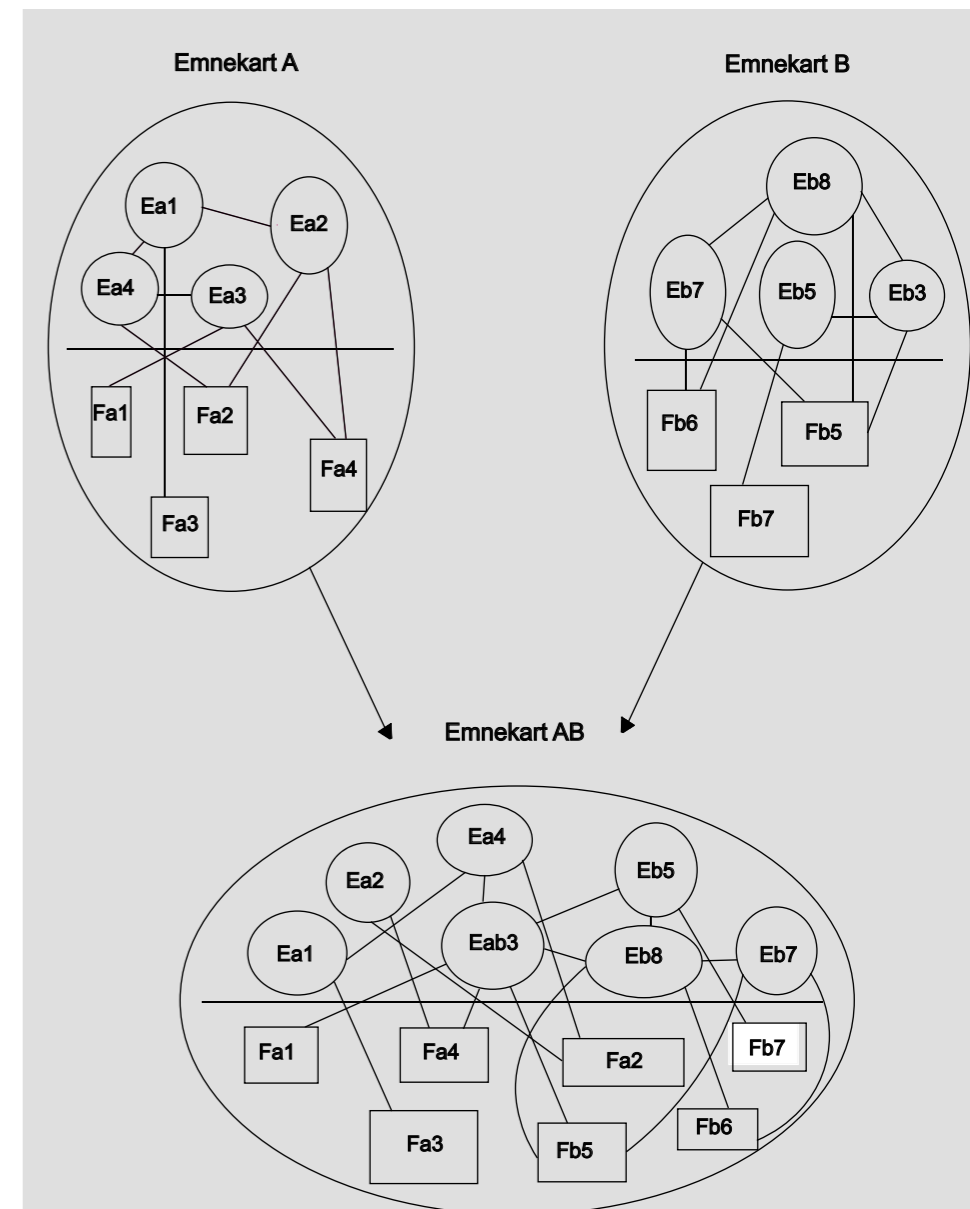
nelige helseemnene var med i. Det betyr at dersom helse fra Forbrukerportalen.no er definert til å være med i perspektivet forbruker, vil det sammenflettede helseemnet være med i det samme perspektivet. Forskjellen blir at helse nå vil inneholde alle karakteristikkene fra helse i forskning.no i tillegg til karakteristikkene fra helse i Forbrukerportalen.no.

Emnekartstandarden inneholder mulighet for å flette sammen hele emnekart eller kun utvalgte emner. Resultatet blir imidlertid det samme; alle emner som representerer det samme temaet blir slått sammen og duplikater blir fjernet (Ahmed 2003). Det sammenflettede emnet representerer det samme temaet som de to emnene som blir slått sammen. Forskjellen er at det sammenflettede emnet får flere og andre karakteristikk enn de opprinnelige emnene hadde.

Brukeren kan få anledning til selv å bestemme om emnekart skal flettes sammen, og eventuelt hvilke emnekart som skal flettes sammen. Det kan også være lagt opp til at det er det enkelte emnekartprogrammet som skal avgjøre når det er best å flette sammen emnekart. I tillegg sier XTM 1.0 (2001) at sammenflettingen "may be indicated by the topic map's author at the time of its creation."

### 6.1.1 Publiserte temaindikatorer

For å kunne identifisere temaet et emne representerer ble temaidentitet



introdusert i kapittel tre. Dersom et tema ikke er en adresserbar ressurs blir temaindikatorer tatt i bruk. Temaidentitet og temaindikatorer er med på å danne grunnlaget for sammenfletting av emnekart. Dersom to emner har samme temaidentitet eller samme temaindikator representerer de det samme temaet og kan derfor flettes sammen. På bakgrunn av dette er det en fordel at samme tema blir identifisert på samme måte i forskjellige emnekart. Emnekartstandarden introduserer derfor *published subject indicators*, PSI. Disse publiserte indikatorene er ment å sikre at de samme identifikatorene blir brukt for det samme temaet.

Formålet med publiserte temaindikatorer er at alle skal benytte seg av den samme indikatoren for et gitt tema. Dette vil bedre sammenflettingen og sørge for utveksling av semantikk i emnekart. XTM 1.0 gir følgende definisjon på publiserte temaindikatorer: “A subject indicator that is published and maintained at an advertised address for the purpose of facilitating topic map interchange and mergeability.” Temaidentitet vil derfor kunne gjøre det mulig for maskiner å utveksle semantikk nettopp fordi de ved hjelp av temaindikatorer kan avgjøre om emner representerer det samme temaet (XML topic maps 2003).

Et eksempel på publiserte temaindikatorer kan være standardiserte ontologier innenfor ett bestemt fagområde, eller en autorisert definisjon av temaet, gjort tilgjengelig på World Wide Web. Innenfor et bestemt fagområde kan publiserte temaindikatorer være en ontologi hvor begreper innenfor fagområdet blir definert. Denne ontologien må være gjort tilgjengelig på World Wide Web. Ontologien bør være av en slik karakter at flest mulig innen fagområdet vil slutte opp om og benytte den. Det vil gjøre det lettere å flette sammen emnekart fra det samme fagområdet.

I XTM 1.0 har topicmaps.org laget en ontologi for ulike elementer i emnekart. Ontologien definerer hvert enkelt element, og definisjonene har en URI som kan benyttes som temaindikator for elementene. Dette betyr at en assosiasjon av typen klasse-enhet kan identifiseres ved en PSI. Denne ontologien er et bidrag til publiserte temaindikatorer for elementer som er obligatoriske i emnekart. For <topic>-elementet gir XTM 1.0 følgende temaindikator:

topic

*The core concept of topic; the generic class to which all topics belong unless otherwise specified.*

(<http://www.topicmaps.org/xtm/1.0/core.xtm#topic>)

<topic>-elementet blir definert og har fått adressen som fungerer som temaindikator. På samme måte blir blant annet assosiasjoner, forekomster, klasse-enhet relasjoner, klasse, enheter, superklasse-subklasse relasjoner, superklasser, subklasser, og i tillegg er det laget indikatorer for sortering og visning av variantnavn.

Denne ontologien er et forslag fra topicmaps.org og er ment å løse problemet med publiserte temaindikatorer for elementer i emnekart. Det er imidlertid ikke påkrevd å benytte denne ontologien for disse elementene.

Publiserte temaindikatorer skal altså gi en utvetydig definisjon av temaer. Denne definisjonen skal, i tillegg til å være forståelig for mennesker, også kunne forstås av maskiner.

## 6.2 Metoder for sammenfletting

Det er mulig å flette sammen hele emnekart eller deler, angitte enkeltemner, av et emnekart. Emnekartstandarden inneholder metoder som gjør det mulig å bestemme om to eller flere emner representerer det samme temaet.

I XTM 1.0 angis det tre måter som bestemmer hvilke emner som representerer det samme temaet:

1. to emner representerer det samme temaet dersom de har samme ressurs som tema. Dette innebærer at temaet er en adresserbar ressurs.
2. to emner som bruker samme ressurs for å beskrive temaet, representerer det samme temaet. Det vil si at de to emnene bruker samme temaindikator.
3. dersom basisnavnet for to emner består av samme streng og basisnavnene er i samme perspektiv, representerer emnene det samme temaet

### 6.2.1 Temabasert sammenfletting

Punkt en og to inngår i metoden for temabasert sammenfletting som er basert på temaidentitet. Dette innebærer at to emner skal flettes sammen dersom de har samme temaidentifikator.

Punkt 1 kan eksemplifiseres med et emnekart hvor nettstedet forskning.no er definert som et emne. Siden forskning.no er en adresserbar ressurs blir temaidentiteten adressen til nettstedet, <http://www.forskning.no>. Det finnes også et emnekart over forskning i Norge og her er også forskning.no emne med temaidentitet <http://www.forskning.no>. Dersom disse to emnekartene blir flettet sammen vil de to emnene bli oppfattet å være like siden de har samme temaidentitet.

Et eksempel på punkt 2 kan være to emnekart over helse. I det ene emnekartet er graviditet definert som et emne. Dette er et eksempel på et tema som ikke er adresserbart, og derfor er adressen til et online medisinsleksikon og oppslag på graviditet, brukt som temaindikator. I et annet nettsted er graviditet også definert som et emne, og bruker samme ressurs som det første emnekartet som temaindikator. Etter punkt to ville emnene bli behandlet som om de er like ettersom de har samme temaindikator. Dersom emnene ikke hadde hatt samme temaindikator ville de ikke ha blitt flettet sammen. Dette tydeliggjør viktigheten av å bruke publiserte temaindikatorer.

### 6.2.2 Navnebasert sammenfletting

Punkt tre er metoden for navnebasert sammenfletting. Denne metoden går ut på å sammenligne basisnavn til emner. Etter navnebegrensningen for emner (se kapittel 3) kan ikke to emner ha det samme basisnavnet i det samme perspektivet. Dersom to emner har det samme basisnavnet i det samme perspektivet, antas de å representere det samme temaet og blir flettet sammen. For eksempelet ovenfor ville de to graviditet-emnene blitt slått sammen dersom begge hadde graviditet som basisnavn i samme perspektiv. Dersom det ene emnet hadde basisnavn graviditet, men var definert i perspektivet kjønn og livsfaser, og det andre emnet hadde samme basisnavn, men var definert i det ubegrensede perspektiv, ville emnene ikke ha blitt flettet sammen.

## 6.3 Konsistente emnekart

Konsistente emnekart er med på å ivareta samlokaliseringssprinsippet. Det at et emnekart er konsistent betyr at det ikke inneholder duplikater av emner og assosiasjoner. Det vil for eksempel være ødeleggende for samlokaliseringssprinsippet dersom et emnekart inneholder to emner som representerer det samme temaet, og har samme navn i samme perspektiv. Da vil brukeren kunne finne orekomster på flere forskjellige steder, og ikke gjennom et enkelt emne. Konsistente emnekart er derfor spesielt aktuelt når det gjelder sammenfletting av emnekart. Når emnekart flettes sammen vil det ofte være slik at de har flere emner felles. Det er da nødvendig at disse emnene flettes sammen og danner ett emne for å sikre at samlokaliseringssprinsippet ivaretas.

“Theoretically and ideally, after completing all topic map processing, every topic has exactly one subject, and

no subject is represented by more than one topic – there’s a one-to-one correspondence between topics and subjects.” (Biezunski 2001).

XTM 1.0 er det utarbeidet metoder som tar sikte på å håndtere situasjoner som ødelegger for samlokaliseringssprinsippet. Disse metodene skal fange opp duplikater ved å bestemme om emner i emnekartet er like. Metodene består av et sett med normative regler.

Reglene tar blant annet for seg muligheten for å bedømme likheten mellom ulike emnekartkonstruksjoner ved å sammenligne URIer, perspektiv, assosiasjoner, emnenavn og forekomster. Reglene beskriver også hvordan ekvivalente syntaktiske representasjoner av emnekarthelementer skal kunne kjønes igjen slik at de ikke vises i visualiseringen av emnekartet. Disse reglene forhindrer at et emnekart inneholder duplikater. Reglene for konsistente emnekart omfatter også hvordan variantnavn skal behandles.

Metodene og reglene for sammenfletting av emnekart er også sentrale i forhold til konsistente emnekart, og skal sammen med reglene ovenfor være med å sikre at samlokaliseringssprinsippet blir ivaretatt når emnekart flettes sammen.

## 6.4 Teoretisk drøfting av sammenfletting

Dette avsnittet tar for seg ulike teoretiske problemer knyttet til det tekniske aspektet ved sammenfletting av emnekart. Utgangspunktet for å drøfte dette er temabasert og navnebasert sammenfletting. Det sentrale spørsmålet som blir behandlet er om disse metodene virkelig sikrer at det kun er de emnene som representerer det samme temaet som blir flettet sammen. Likeledes er det også viktig å se om metodene sørger for at de emnene som representerer det samme temaet faktisk blir flettet sammen. Dette blir drøftet og illustrert ved ulike eksempler.

### 6.4.1 Rekkefølgen for temabasert og navnebasert sammenfletting

Sammenfletting av emnekart skjer automatisk når et <topicMap>-element inneholder et <mergeMap>-element som peker til et annet emnekart. Ut i fra dette kan det virke som om temabasert og navnebasert sammenfletting brukes samtidig. XTM 1.0 sier ingenting om at det kan angis en bestemt rekkefølge for metodene. Det er heller ikke noe i XTM 1.0 som indikerer at den ene metoden har rang foran den andre.

Det kan oppstå situasjoner hvor det vil være ønskelig å kun benytte en

av metodene. Dersom det for eksempel ikke er definert noe perspektiv for emnekartene som skal flettes sammen, kan det være problematisk at navnebasert sammenfletting ikke kan kuttes ut. Ved å bruke navnebasert sammenfletting i en slik situasjon, vil man risikere at emner som representerer forskjellige temaer blir flettet sammen. Dette er et eksempel på problemer knyttet til homonymi.

Det er likevel mulig at løsningen på dette ligger i selve programmeringen av emnekartprogrammer. I avsnitt 6.5 som omfatter sammenfletting av emnekart i praksis, ble emnekartleseren Omnigator brukt. I denne leseren er det mulighet for både temabasert og navnebasert sammenfletting, samtidig som navnebasert sammenfletting kan kuttes ut dersom det skulle være ønskelig. Dette kan tyde på at det er mulig å bare benytte seg av en av de to sammenflettingsmetodene, men dette er imidlertid ikke angitt i standarden.

### 6.4.2 Problemer knyttet til navnebasert sammenfletting

I navnebasert sammenfletting spiller basisnavn og perspektiv en viktig rolle, men det kan være problemer knyttet til dette. Dersom det ikke er definert noe perspektiv for et emne eller et emnekart, gjelder det ubegrensede perspektivet. Denne situasjonen kan oppstå for flere emnekart. Da emnekartene ble konstruert ble det ikke tatt hensyn til at de kunne brukes i andre sammenhenger enn den de ble opprettet i, eller at de kunne flettes sammen. Det å definere perspektiv vil innebære mer arbeid for den som lager emnekartet og dersom gevinsten ved å gjøre dette ikke blir klar umiddelbart, kan det være lettere å utelate dette. Etter reglene vil emner som har samme basisnavn i det samme perspektivet bli slått sammen. Dermed kan det oppstå situasjoner hvor emner som har det samme navnet, men som burde være definert i hvert sitt perspektiv, likevel blir flettet sammen. Dette gjelder for eksempel for homonymer.

### 6.4.3 Problemer knyttet til temabasert sammenfletting

Hovedfundamentet for temabasert sammenfletting er temaidentitet og temaindikatorer. Denne metoden forutsetter at to emner har samme URI som identifikator for at de skal kunne flettes sammen og her spiller publiserte temaindikatorer en sentral rolle. Det virker imidlertid urimelig å forvente at alle emnekart vil bruke den samme temaindikatoren for det samme emnet.

Ulike fagmiljøer og grupper vil sannsynligvis benytte seg av forskjellige indikatorer. I slike situasjoner vil ikke temabasert sammenfletting sikre at emner som representerer det samme temaet blir flettet sammen.

På en annen side kan det imidlertid være nærliggende å tro at temabasert sammenfletting vil fungere godt i et lukket miljø, det vil si innenfor et bestemt fagområde hvor det er aktuelt at flere emnekart deler samme ontologi (XML topic maps 2003).

### 6.4.4 Problemer knyttet til publiserte temaindikatorer

I følge emnekartstandarden representerer to emner det samme temaet dersom de bruker samme ressurs for å identifisere temaet. Det kan likevel oppstå situasjoner hvor to adresserbare ressurser peker til det samme temaet selv om adressene ikke er identiske. Adressen <http://www.topicmaps.org> viser til den samme ressursen som <http://www.TOPICMAPS.org>, men emnekartapplikasjonen vil ikke anse disse to for å vise til samme ressurs fordi det i emnekartstandarden skilles mellom store og små bokstaver.

Problemet med at emner som representerer det samme temaet ikke bruker samme temaindikator håndteres i XTM 1.0 ved et tredje emne som kan bygge bro mellom emnene. Dette kalles brobyggingsmetoden. Identiteten til det tredje emnet blir etablert gjennom indikatorene til de to andre emnene. Dette er mulig fordi et emne kan ha flere temaindikatorer. Det tredje emnet vil derfor bli slått sammen med hvert av de to andre emnene, og resultatet blir ett emne med karakteristikk fra alle de tre emnene.

Emnekartstandarden inneholder imidlertid ikke metoder som kan fange opp tilfeller hvor brobyggingsmetoden bør brukes. Ut i fra dette virker det som om denne metoden må initieres manuelt. Dette vil kanskje være mulig for et lite emnekart, men for et emnekart av en viss størrelse virker dette nærmest umulig. En emnekartapplikasjon vil ikke kunne fange opp at to emner representerer det samme temaet når de ikke bruker samme temaindikator.

### 6.4.5 Ulike problemstillinger knyttet til sammenflettingsmetodene

Metodene for sammenfletting av emnekart gir også grunnlag for andre spørsmål som bør drøftes.

• *Emner med samme navn i et gitt perspektiv, men som ikke har samme temaindikator.*

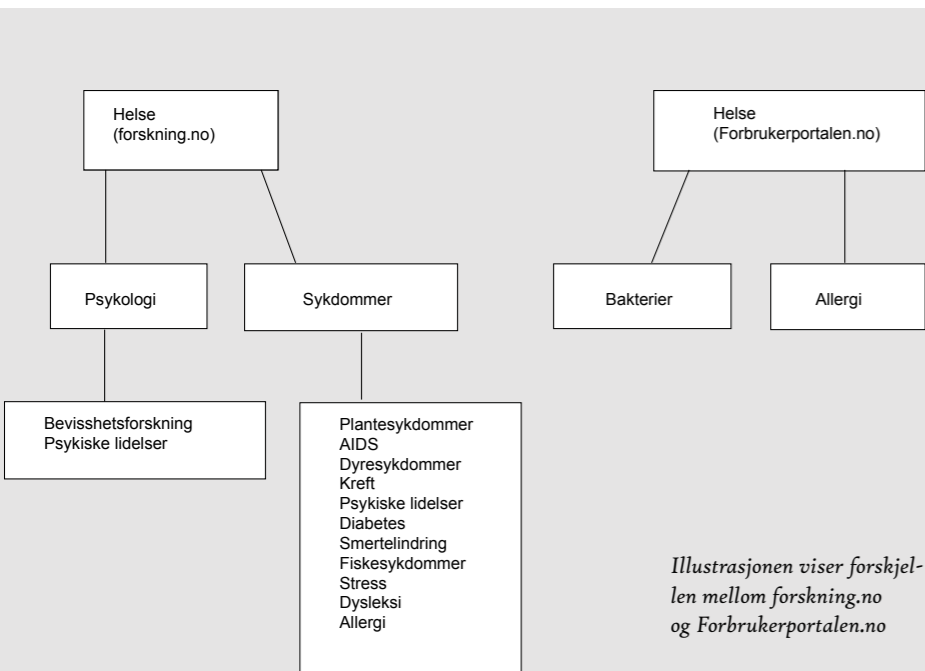
Brobyggingsmetoden lar seg som

nevnt bruke til å koble sammen emner som representerer det samme temaet, men som ikke bruker samme tema-indikator. Dette forutsetter at disse emnene blir fanget opp, og at emnene virkelig representerer det samme temaet. I og med at standarden ikke angir en bestemt rekkefølge for temabasert og navnebasert sammenfletting kan det virke som om metodene virker side om side. Det kan derfor skje at to emner som ikke har samme indikator, og derfor ikke vil bli slått sammen ved metoden for temabasert sammenfletting, likevel kan bli flettet sammen ved navnebasert sammenfletting. Dette forutsetter imidlertid at de to emnene har samme navn i samme perspektiv. Dette kan få konsekvenser for håndteringen av homonymer.

- *Emner som har samme indikator, men ikke samme basisnavn i et gitt perspektiv.* I navnebasert sammenfletting vil ikke

Helse forskning.no	Helse Forbrukerportalen.no
Psykologi	Allergi
Framtidens medisin	Bakterier
Vaksiner	Ernæring
Bakterier	Helsevesen
Bioteknologi	Kunstig befruktning
Medisiner	Lege
Sykdommer	Medisin
Menneskegruppen	Pasientrettigheter
Mat og helse	Smittevern
Medisin	Tannlege
Graviditet	
Mat	

Tabellen gir en oversikt over emnene som er med i sammenflettingen



Illustrasjonen viser forskjellen mellom forskning.no og Forbrukerportalen.no

disse emnene bli flettet sammen. I og med at de har samme indikator vil de kunne bli flettet sammen ved temabasert sammenfletting. To emner som har samme identifikator, men ulike navn i samme perspektiv, kan derfor bli flettet sammen. Spørsmålet som da gjør seg gjeldende er hvilket basisnavn det nye emnet skal ha. Det er to basisnavn å velge mellom, men standarden sier ingenting om hvilket av disse som bør eller skal velges.

Dette kan illustreres ved et eksempel. Et emnekart inneholder emnet *graviditet* mens et annet emnekart inneholder emnet *svangerskap*. Begge emnekartene er i det ubegrensede perspektiv, og begge emnene bruker den samme tema-indikatoren. Ved temabasert sammenfletting vil de to emnene bli flettet sammen. Emnekartstandardens sier imidlertid ikke noe om hvilket basisnavn det nye emnet får. De to emnene ville derimot ikke blitt flettet sammen dersom metoden for nav-

nebasert sammenfletting ble brukt.

Det er mulig at løsningen ligger i å benytte både navnebasert og temabasert sammenfletting. Dette for å sikre at emner som representerer det samme temaet faktisk blir slått sammen, og for å forhindre at emner som representerer forskjellig tema ikke blir flettet sammen. Det kunne imidlertid være ønskelig at standarden tydelig beskrev hvilke metoder som skal brukes når og i hvilken rekkefølge.

Emnekartstandardens er, som tidligere presisert, veldig åpen. Det er ikke et absolutt krav om bruk av indikatorer og det er heller ikke et absolutt krav om bruk av publiserte tema-indikatorer. Dette blir imidlertid anbefalt og standarden framhever viktigheten av å bruke publiserte tema-indikatorer nettopp av hensyn til sammenfletting. Dette vil være med på å sikre konsistente emnekart, noe som igjen ivaretar samlokaliseringssprinsippet.

## 6.5 Sammenfletting i praksis

Dette praktiske forsøket er ment å gi en bedre forståelse av sammenfletting av emnekart. Resultatet av sammenflettingen vil være med på å gi et bilde av hva brukeren kan forvente å møte når to emnekart er flettet sammen. Nedenfor følger en beskrivelse av hvordan forsøket ble gjennomført, hvilke begrensninger som ligger til grunn, før resultatet blir presentert og analysert. I likhet med analysen i kapittel 5, er blant annet samlokaliseringssprinsippet og navigasjonsmulighetene også her vektlagt.

### 6.5.1 Framgangsmåte

Utgangspunktet for forsøket er helsekategoriene i forskning.no og Forbrukerportalen.no. Nettstedene og emnene endres over tid, og grunnlaget for sammenflettingen er nettstedene slik de så ut 9. april 2003.

Emner refererer i denne sammenhengen til betydningen av termen slik den blir brukt i kunnskapsorganisasjon. Det innebærer at emner er begrenset til de emnene som representerer et tema i et emnekart.

Forskning.no er et større nettsted enn Forbrukerportalen.no, og følgelig er også helsekategorien for dette nettstedet større enn helsekategorien i Forbrukerportalen.no forskning.no har 23 underkategorier til *helse*, mens Forbrukerportalen.no har 10 underkategorier. Antall underkategorier for *helse* i forskning.no ble i dette forsøket kuttet ned til 12 kategorier for lettere å se resultatet av sammenflettingen. Kategoriene fra Forbrukerportalen.no ville da ikke for-

	forskning.no	forbrukerportalen.no
<b>Emnetyper</b>	Tema Fag	Emne
<b>Assosiasjonstyper</b>	Undertemaer Hører til under Beslektede temaer Beslektede fag	Overordnet Underordnet Beslektet
<b>Roller</b>	Overordnet Underordnet Beslektet tema Beslektet fag	Overordnet Underordnet Beslektet

Tabellen viser hvordan emnetyper, assosiasjonstyper og roller blir definert for forskning.no og Forbrukerportalen.no

svinne blant kategoriene til forskning.no. Dette var viktig for å kunne se hva som skjedde med emnene fra begge nettstedene, og hvordan resultatet av sammenflettingen ble.

Utvelgelsen av kategorier fra forskning.no er basert på gjennomgangen av nettstedet i kapittel 5. Emner hvor det ble identifisert problemer er særlig interessante i denne sammenheng. Sammenflettingen vil gi anledning til å se om problemene forsterkes eller endres i det sammenflettede emnekartet. Samtidig ble emner som hadde noe til felles med emnene i Forbrukerportalen.no også valgt ut. Dette for å se hvilke konsekvenser sammenflettingen ville ha for dem. Tabellen på neste side gir en oversikt over emnene som er med i sammenflettingen:

Helsekategorien i forskning.no inneholdt, som nevnt ovenfor, flere underkategorier av emner enn helsekategorien hos Forbrukerportalen.no. Det var med andre ord flere nivåer av kategorier. Det ble imidlertid ikke tatt hensyn til dette i sammenflettingen av emnekartene. For underkategoriene til helse i forskning.no ble deres underkategorier også tatt med. På neste side ses et eksempel hentet fra nettstedene som illustrerer forskjellen mellom de:

### 6.5.2 Utarbeidelse av emnekartene

Det ble laget to emnekart med utgangspunkt i henholdsvis forskning.no og Forbrukerportalen.no. Dette ble gjort ettersom emnekartfilene som nettstedene bygger på ikke var tilgjengelige i denne prosessen. Ulike hensyn måtte da tas og dette la igjen en del begrensninger i konstruksjonen av emnekartene.

Identifikatorer ble ikke brukt i dette eksperimentet. Årsaken til det er at det ikke finnes en oversikt over hvilke tema-indikatorer som er brukt hos forskning.no og Forbrukerportalen.no. Gjennom å ta i bruk identifikatorer ville det imidlertid ha vært mulig å illustrere tema-

basert sammenfletting. Ettersom det ikke var mulighet for dette, er det kun navnebasert sammenfletting som er testet. Hensikten med dette eksperimentet var å se på hva som møter brukeren etter at to emnekart er flettet sammen, og dette burde derfor ikke være av stor betydning. Den navnebaserte sammenflettingen gir et bilde av prosessen og resultatet av å flette sammen emnekart.

Det ble heller ikke tatt hensyn til forekomster i forsøket. Begrunnelsen for dette er at det først og fremst er det som skjer med emner og assosiasjoner i sammenflettingen som er fokus. Der som forekomstene hadde vært med ville de ha gitt et mer fullstendig bilde av hva som kom til å møte brukerne i det sammenflettede emnekartet, men de ville imidlertid ikke hatt innflytelse på emner og assosiasjoner.

Emner, emnetyper, assosiasjonstyper og roller som er brukt i dette forsøket, er definert på bakgrunn av hvordan de framstår på nettstedene. Terminologien fra nettstedene er brukt i et forsøk på å gjøre emnekartene så autentiske som mulig. Slik nettstedene er utformet bruker de forskjellige navn på henholdsvis emner, emnetyper, assosiasjonstyper og roller. Tabellen gir en oversikt over hvordan disse ble definert for forskning.no og Forbrukerportalen.no.

For begge emnekartene gjelder det ubegrensede perspektiv. Emnekartene ble kodet i henhold til XTM-formatet, og i møtekommer de syntaktiske kravene i XTM i.o.

Selve sammenflettingen ble gjort ed hjelp av *Omnigator*. Dette er en emnekartleser som viser emnekart, og er utviklet av *Ontopia*. Egendefinerte emnekart kan lastes inn i programmet som gir mulighet for å behandle dem på ulike måter, deriblant å flette dem sammen.

*Emnetyper, assosiasjoner og roller* Forskning.no opererer med *tema* og *fag* som emnetype. Forbrukerportalen bru-

ker på sin side emnetyper emne. Dette kan illustreres med et eksempel: *Helse og psykologi* er henholdsvis *tema* og *fag* i forskning.no:

```

<topic id="forskningHelse">
  <instanceOf>
    <topicRef xlink:href="#tema"/>
  </instanceOf>
  <baseName>
    <baseNameString>Helse</baseNameString>
  </baseName>
</topic>

```

```

<topic id="forskningPsykologi">
  <instanceOf>
    <topicRef xlink:href="#fag"/>
  </instanceOf>
  <baseName>
    <baseNameString>Psykologi</baseNameString>
  </baseName>
</topic>

```

Assosiasjoner og assosiasjonstyper er også valgt med utgangspunkt i nettstedene. Det vil si at når *helse* er overordnet tema for *sykdommer* hos forskning.no, så blir assosiasjonen seende slik ut:

```

<association id="helse-sykdommer-
  assosiasjon">
  <instanceOf>
    <topicRef xlink:ref="#relasjon" />
  </instanceOf>
  <member>
    <roleSpec><topicRef xlink:href="#overordnet"/></roleSpec>
    <topicRef xlink:href="#forskningHelse"/>
  </member>
  <member>
    <roleSpec><topicRef xlink:href="#underordnet"/></roleSpec>
    <topicRef xlink:href="#forskningSykdommer"/>
  </member>
</association>

```

Her defineres også medlemmer og deres roller i assosiasjonen. Som eksemplet viser, er *helse* og *sykdommer* medlemmer i assosiasjonen. *Helse* har rollen *overordnet*, mens *sykdommer* har rollen *underordnet*. *Sykdommer* er dermed et underordnet tema til *helse*.

## 6.6 Resultatet av sammenflettingen

I innledningen av denne oppgaven ble det sagt at emnekart er et navigasjons-

middel for samlinger av ressurser på nett. Resultatet av å flette sammen to emnekart er et sammenflettet emnekart hvor brukeren får flere navigasjonsmuligheter. Dette fordrer imidlertid at det blir gjort enklere for brukeren å finne en navigeringsmulighet som er relevant for det hun er på jakt etter.

Det mest i øyenfallende når helse fra henholdsvis forskning.no og Forbrukerportalen.no blir flettet sammen, er at det sammenflettede emnekartet åpner for nye navigasjonsmuligheter. Det oppstår nye assosiasjoner mellom emnene og dersom forekomstene hadde vært tatt med, ville hvert enkelt emne bli knyttet opp mot flere forekomster.

Det sammenflettede emnekartet ble studert ut fra tanken om samlokaliseringssprinsippet og ut fra hvilke navigasjonsmuligheter det tilbyr. Funnene fra analysen kan oppsummeres slik: Problemer

- knyttet til emnetyper

- knyttet til assosiasjoner
- knyttet til plassering i hierarkier
- på ressursnivå

I det følgende blir hvert av disse punktene gjennomgått og eksemplifisert

#### 6.6.1 Problemer knyttet til emnetyper

I det sammenflettede emnekartet er alle emnetyper fra de opprinnelige emnekartene representert. På den måten blir for eksempel helse både definert som emnetype emne og emnetype tema.

Det virker forvirrende at emnekartet opererer med emnetyper emne, tema og fag. Det er naturlig å sette spørsmål ved hva som skiller disse emnetyperne fra hverandre. Emnekartstandardene inneholder ikke metoder for å fange opp og samle emnetyperne. I det sammenflettede emnekartet blir det vanskelig for en bruker å vite hva som skiller emner, tema og fag.

Dette kan eksemplifiseres ved å se

på hva som skjedde med emnet bakterier i sammenflettingen. Både forskning.no og Forbrukerportalen.no hadde opprinnelig et emne som het bakterier. Bakterier ligger på samme nivå hos både forskning.no og Forbrukerportalen.no. De to forskjellige bakterieemnene ble flettet sammen i det nye emnekartet. Se illustrasjon på neste side.

Bakterier er her definert både som emnetype *tema* og *emne*. Dette innebærer at dersom en bruker klikker på *emne*, får hun opp en oversikt over alle emner av emnetype *emne*. Her vil hun blant annet finne *bakterier*. Dersom hun hadde klikket på *tema* ville hun også ha funnet *bakterier*. Dette ødelegger dermed for samlokaliseringssprinsippet.

De ulike emnetyperne representerer riktignok ikke fundamentale problemer knyttet til sammenfletting av emnekart, men er med på å gi brukeren et rotete bilde av nettstedet.

#### 6.6.2 Problemer knyttet til assosiasjoner

I utgangspunktet var det mulig å identifisere assosiasjoner som var like i de to opprinnelige emnekartene. Det vil si at assosiasjonene bestod av samme type relasjoner. Dette kan illustreres ved sammenflettingen av bakterier. Når bakterier er underordnet helse, er det samme relasjon som bakterier hører til under helse. Assosiasjonene består altså av samme relasjoner, men har forskjellige navn i de to opprinnelige emnekartene. Navnene beholdes i det nye emnekartet og dermed blir samme type assosiasjon uttrykt på forskjellige måter.

I eksemplet med *bakterier* vises dette også ved at *bakterier* både har beslektede temaer og er beslektet med andre emner. I realiteten er *bakterier* relatert til disse emnene med den samme relasjonen. Det er derfor ingen grunn til å gi assosiasjonene forskjellige navn. Det at assosiasjonene uttrykkes på forskjellig måte gir imidlertid et uoversiktlig bilde for brukeren.

Ettersom samme type assosiasjon kan uttrykkes på flere måter, er det naturlig å spørre seg om ulike typer assosiasjoner kan uttrykkes på samme måte. Dette kan for eksempel gjelde for assosiasjoner som bygger på assosiative relasjoner.

Ettersom det er to små emnekart som ligger til grunn for sammenflettingen er det ikke mulig å eksemplifisere alle muligheter som ligger i emnekart. Det kan også være flere potensielle problemer knyttet til nettopp assosiasjoner som ikke kommer fram i forsøket. Dette kan gjelde for større emnekart som inneholder flere og kanskje mer spesia-

liserte assosiasjoner. Da kan emnene i det sammenflettede emnekartet bli medlemmer av assosiasjoner som er vidt forskjellige og som på den måten kan virke forvirrende på brukeren. Dette kan særlig være tilfellet dersom et spesialisert emnekart blir flettet sammen med et generelt emnekart.

#### 6.6.3 Plassering i hierarkiet

Både forskning.no og Forbrukerportalen.no har et emne som heter *allergi*. Emnene ble flettet sammen og resultatet ble som i illustrasjonen lengre opp på siden.

I det sammenflettede emnekartet fikk *allergi* emnetyperne *emne* og *tema*. I tillegg ble *allergi* beslektet med *mat*, og ble underordnet *helse*, samt at det hørte til under *sykdommer*. Dette eksemplet er ment å illustrere problemer knyttet til antall nivåer i hierarkiene til de opprinnelige emnekartene. I forskning.no ligger *allergi* under *sykdommer* som igjen ligger under *helse*. I Forbrukerportalen.no derimot, ligger *allergi* direkte under *helse*. Derfor blir *allergi* plassert under to emner i det sammenflettede emnekartet. Det er ikke nødvendigvis forståelig for brukeren hvorfor *allergi* ligger både under *helse* og under *sykdommer*. Det positive ved dette er imidlertid at brukeren kan få inngang til samme emnet fra ulike steder, men dette kan også ødelegge for samlokaliseringssprinsippet dersom ikke alle forekomster knyttet til *allergi* kan nås gjennom et enkelt emne.

I forlengelsen av problemer knyttet til plassering i hierarkiet kan man tenke seg at det for store emnekart, med dype hierarkier vil oppstå problemer når dette flettes sammen med et emnekart med grunnere hierarki. Da kan det oppstå situasjoner hvor det blir vanskelig å se hvordan et emne langt nede i hierarkiet kan stå i sammenheng med et emne langt opp i hierarkiet.

#### 6.6.4 Problemer på ressursnivå

Som tidligere nevnt ble ikke forekomster tatt med i dette forsøket. Likevel er det mulig å tenke seg situasjoner hvor det oppstår problemer som berører både emner og forekomster.

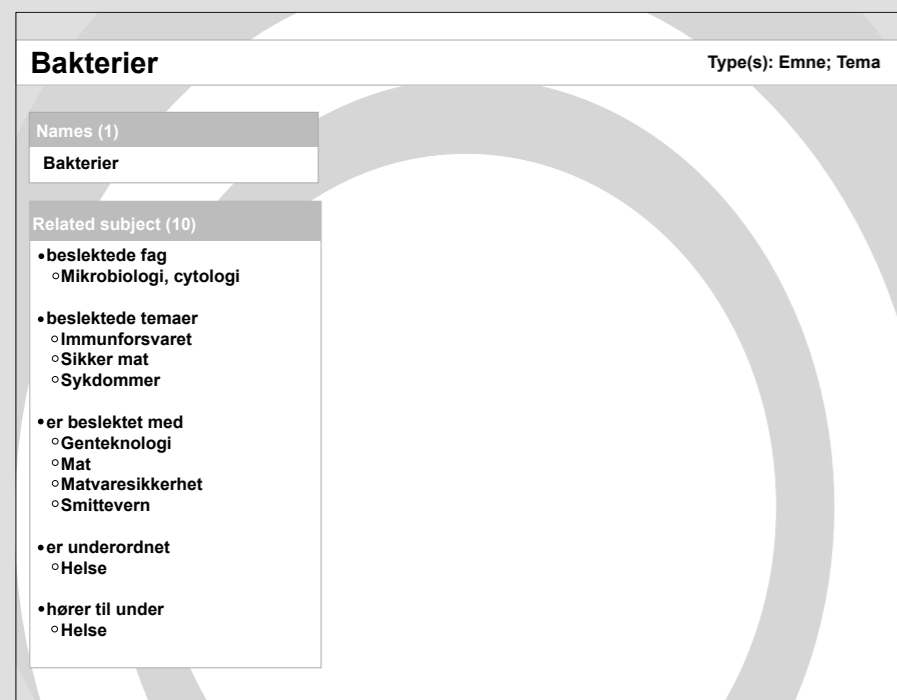
Dersom det ene emnekartet inneholder mer spesifikke emner enn det andre, kan det oppstå problemer angående enkelte forekomster. Forbrukerportalen.no har et emne kalt *kunstig befruktning*, mens forskning.no har to emner kalt *graviditet* og *svangerskap*. Kategoriene hos forskning.no er knyttet opp til en forekomst som er en artikkel om kunstig befruktning. Når de to emnekartene flettes sammen vil det være naturlig å forvente å finne alt om kunstig befrukt-

ning gjennom emnet *kunstig befruktning*. Ettersom verken *svangerskap* eller *graviditet* vil bli flettet sammen med *kunstig befruktning* vil ikke artikkelen kunne finnes gjennom dette emnet. Det finnes ingen mekanismer i emnekartstandarden for å fange opp slike situasjoner. Dersom dette skal løses må det gjøres manuelt.

Dette vil også kunne gjelde dersom et emnekart har et emne som heter *pollenallergi* og har ressurser som er relevante for temaet knyttet til seg. Dersom et annet emnekart kun har et emne som heter *allergi*, vil ressurser som omhandler pollenallergi være knyttet til dette emnet sammen med andre ressurser som omhandler andre typer allergier. Dersom disse to emnene blir flettet sammen kan man risikere å ha ressurser om pollenallergi knyttet opp mot *allergi* i stedet for pollenallergi.

#### 6.6.5 Oppsummering av forsøket

Problemene som er identifisert ved dette forsøket er i all hovedsak knyttet opp mot brukervennlighet. Analysen av det sammenflettede emnekartet har vist at logikken i navngiving for emner og assosiasjoner brytes. Det er også pekt på problemer som får konsekvenser for samlokaliseringssprinsippet, og som gjør at navigeringen blir noe forvirrende.



Sammenfletting av emne allergi hos forskning.no og Forbrukerportalen.no



Illustrasjonen viser hva som skjedde med emnet bakterier i sammenflettingen

# OPPSUMMERING

Sammenligningen av emnekart og tesauri har vist at de to modellene skiller seg fra hverandre på flere punkter. En kan derfor ikke snakke om en direkte sammenligning, men heller en sammenligning i form av at ulike aspekter ved de to modellene har blitt drøftet.

Emnekart er et forholdsvis nytt verktøy og derfor har oppgaven lagt vekt på en grundig presentasjon av modellen. Tesauri ble valgt som sammenligningsgrunnlag da dette verktøyet har lange tradisjoner i kunnskapsorganisasjon. I tillegg bygger emnekart på verktøy som bærer preg av å ha visse tesaurusaspekter i seg.

Den teoretiske sammenligningen har vist at emnekart og tesauri springer ut fra forskjellige tradisjoner, noe som igjen viser seg i forskjellene i strukturen til de to modellene. Den strukturelle forskjellen gir seg utslag i at emnekart har bedre navigasjonsmuligheter, mens tesauri hovedsakelig tilbyr hierarkisk navigering. En tesaurus er et kontrollert vokabular hvis hovedformål er å hjelpe indekserere og brukere til å velge de samme termene i henholdsvis indeksering og formulering av søkeuttrykk. Emnekart er på sin side et navigasjonsverktøy for framvisning av emner og assosiasjoner knyttet opp mot forekomster i World Wide Web.

Emnekartstandarden legger vekt på utformingen av emnekartets nettverksstruktur. Dette kommer til syne ved at standarden ikke tar utgangspunkt i en hierarkisk ordning, men heller fram-

hever betydningen av assosiative relasjoner. Emnekart er utviklet med sikte på implementering i World Wide Web, og har blant annet til hensikt å lede brukere til andre ressurser som henger sammen med emnet de er interessert i.

I forhold til gjenfinning og beskrivelse av nettressurser ligger det forskjellige muligheter i de to modellene. Tesaurusstandarden med sine regler, sørger for at en tesaurus inneholder forholdsvis forutsigbare relasjoner. En tesaurus gir på bakgrunn av disse reglene, samt de reglene som gjelder for valg av terminologi, et grunnlag for konsistente utvidelsesmuligheter. Ettersom det ikke ligger begrensninger i emnekartstandarden kan emnekart gi en rik beskrivelse av ressurser, emner, og relasjonene mellom dem. Todelingen gjør at emnekart blir en teknikk for kobling mellom beskrivelse og ressurs.

Opgaven har tatt for seg to implementasjoner av emnekart, og har drøftet disse i forhold til tesauri. Det har blitt pekt på enkelte problematiske forhold knyttet til bruk av emnekart. Det har også blitt pekt på at noen av problemene vil forekomme i tesaurisammenheng, og at problemene således ikke er direkte knyttet til emnekartstandarden. Dette gjelder for eksempel polyhierarkier. Det har også blitt vist at problemer knyttet til synonymmer blir løst på en sikrere måte i tesaurusstandarden enn i emnekartstandarden.

Et annet moment som har vært tatt opp er manglende navn på assosiasjoner.

I analysen av nettstedene kom det fram problemer knyttet til manglende forklaring på hvilken type relasjon der var mellom ulike emner. I en tesaurus blir det satt navn på de ulike relasjonstypene.

Flere av problemene som har kommet fram under analysen skyldes at emnekartstandarden er langt mer åpen enn tesaurusstandarden. Selv om dette gir større frihet i utformingen av emnekart, vil det samtidig gi større rom for misforståelse og navigasjonsvansker for brukeren. Det å oppdage eventuelle problemer og vurdere konsekvensen av friheten som emnekartstandarden gir, skyves dermed over på den som utformer nettstedet.

Sammenfletting av emnekart har vært drøftet og de ulike metodene for sammenfletting har blitt diskutert. Det kom fram problemer knyttet til bruk av perspektiv og temaindikatorer.

Den praktiske gjennomføringen av sammenflettingen av nettstedene forskning.no og Forbrukerportalen.no har vist ulike problemer knyttet til denne prosessen. Det har blitt observert at i det sammenflettede emnekartet ble emnene definert som flere emnetyper og samme assosiasjoner ble uttrykt på forskjellige måter. Dette ga et uoversiktlig inntrykk over navigasjonsmuligheten, og av hvordan emnene var relatert til hverandre.

Det sammenflettede emnekartet har også vist hvordan en bruker vil få tilgang til flere emner og flere navigasjonsmuligheter. Forsøket var imidlertid

for lite til at alle mulighetene som ligger i sammenfletting av emnekart kom fram. Det er derfor heller ikke mulig å trekke noen konklusjoner når det gjelder emnekart og sammenkobling av vokabularer, og en anbefaling kan heller ikke gis. Oppgaven har imidlertid pekt på noen aspekter ved sammenfletting som gir en pekepinn på hva man kan vente seg.

Både analysen og sammenflettingen har vært gjennomført i den hensikt å se om ulike aspekter kunne ødelegge for samlokaliseringssprinsippet.

Dersom det hadde vært tilstrekkelig med tid og ressurser, kunne det ha vært interessant å gjennomføre en større analyse, og en sammenfletting av flere og større emnekart.

For å tydeliggjøre forskjellene mellom emnekart og tesauri kunne det ha vært gjort praktiske forsøk som omfattet begge modellene. Dersom en tesaurus fra samme fagområdet som et emnekart, hadde vært tilgjengelig via World Wide Web, ville det vært enklere å se hvordan modellene fungerer i forhold til hverandre når de er implementert. Dette ville ha gjort den praktiske sammenligningen lettere, og ville klarere ha illustrert fordeler og ulemper ved modellene.

For å få et mer fullstendig bilde av sammenfletting kunne det også vært interessant å se på hva som skjer dersom flere, og større emnekart blir flettet sammen. Særlig ville det ha vært interessant å se hva som skjer med assosiasjonene og hvilke konsekvenser dette får

for navigeringsmulighetene. En kunne da ha forsøkt å flette sammen emnekart som tar i bruk et bredere spekter av forskjellige assosiasjoner enn de som ble brukt i denne oppgaven.

# LITTERATURLISTE

- Ahmed, Kal 2003: *Introducing topic maps : a powerful, subject-oriented approach to structuring sets of information*. I: XML-journal. – Vol. 03, issue 10
- Aitchison, Jean 2000: *Thesaurus construction and use : a practical manual* / Jean Aitchison, Alan Gilchrist, David Bawden. – 4th ed. – London : Aslib. – XIV. 218 s.
- Berners-lee, Tim 1998: *Semantic web road map* [online]. – URL: <http://www.w3.org/DesignIssues/Semantic.html> (lesedato 2003-03-21)
- Berners-Lee, Tim 2000: *Weaving the web : the past, present and future of the world wide web by its inventor*. – London: Texere. – 272 s.
- Berners-Lee, Tim 2001: *The semantic web : a new form of web content that is meaningful to computers will unleash a revolution of new possibilities* / Tim Berners-Lee, James Hendler James and Ora Lassila. - (May 17, 2001) [online]. – URL: [http://www.sciam.com/print\\_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21](http://www.sciam.com/print_version.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21) (lesedato 2003-01-30)
- Berners-Lee, Tim 2003: *The semantic web lifts off / Tim Berners-Lee and Eric Miller*. – [online]. – URL: [http://www.ercim.org/publication/Ercim\\_News/enw51/nerners-lee.html](http://www.ercim.org/publication/Ercim_News/enw51/nerners-lee.html) (lesedato 2003-03-25)
- Biezunski, Michel 2001: *XML topic maps : finding aids for the web* / Michel Biezunski and Steven R. Newcomb. – S. 2-6. – I: IEEE multimedia. – april-june edition
- Bray, Tim 2001: *What is RDF?* [online]. – URL: <http://www.xml.com/lpt/a/2001/01/24/rd.html> (lesedato 2003-11-10)
- Brenne, Marte 2002: *Tesaurus kontra emnekart : temaoppgave i brukerorientert gjenfinning* / Marte Brenne og Marit Kristine Ådland. – Oslo : Høgskolen i Oslo, avd. for journalistikk-, bibliotek- og informasjonsfag. Sivilbibliotekarstudiet. – 24 s.
- BRODD 2003: *Seminar om emnekart : presentasjoner* [online]. – URL: <http://www.brodd.no/kurs/emnekart/emnekart.htm> (lesedato 2003-11-14)
- Clarke, Stella G. Dextre 2001: *Thesaurial relationships*. – S. 37-52. – I: Relationships in the organization of knowledge / edited by Carol A. Bean and Rebecca Green. – Dordrecht : Kluwer.- 232 s.
- Connolly, Thomas M. 2002: *Database systems : a practical approach to design, implementation, and management* / Thomas M. Connolly, Carolyn E. Begg. – London : Addison-Wesley. – S. 998-1044.
- Draft requirements, examples, and a “low bar” proposal for topic map constraint language (TMCL) 2001.[online]. – URL: <http://www.y12.doe.gov/sgml/sc34/document/0226.htm> (lesedato 2003-01-30)
- Emnekart.no 2002. [online]. – URL: <http://www.emnekart.no> (lesedato 2003-11-14)
- Forbrukerportalen.no 2003. [online]. – URL: <http://www.forbrukerportalen.no> (lesedato 2003-11-01)
- Forskning.no 2003. [online]. – URL: <http://www.forskning.no> (lesedato 2003-11-01)
- Garshol, Lars Marius 2001: *Topic maps : orden i informasjonskaos* [online]. – URL: <http://www.it-forum.no/arkiv/itf2001/foredrag/TopicMaps.pdf> [lesedato 2003-01-30]
- Gilchrist, Alan 2003: *Thesauri, taxonomies and ontologies : an etymological note*. – S. 7-18. – I: Journal of documentation. – Vol. 59, no. 1
- Green, Rebecca 2001: *Relationships in the organization of knowledge : an overview*. – S. 3-18. – I: Relationships in the organization of knowledge / edited by Carol A. Bean and Rebecca Green. – Dordrecht : Kluwer.- 232 s.
- Greenberg, Jane 2003: *The semantic web : more than a vision*. – S. 6-11. – I: Bulletin of the American society for information science and technology. – April/May, 2003
- Gruber, Tom 2003: *What is an ontology?* [online]. – URL: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html> (lesedato 2003-10-15)
- Grønmo, Geir Ove 2000: *Creating semantically valid topic maps* [online]. – URL: <http://www.gca.org/papers/xml-europe2000/papers/s29-02.html> (lesedato 2003-02-11)
- Guide to the topic map standards 2002. [online]. – URL: <http://www.y12.doe.gov/sgml/sc34/document/0323.htm> (lesedato 2003-01-30)
- Hjørland, Birger 1993: *Emnerepresentasjon og informasjonssøgning : bidrag til en teori på kundskabsteoretisk grundlag*. – Göteborg : VALFRID. – 258 s.
- Hjørland, Birger 1995: *Informationsvidenskabelige grundbegreper*. – 2. rev. utg. – København : Danmarks Biblioteksskole. – 2 b. – Innhold: B. 1 : 80/20 reglen – informationsstrukturer. B. 2 : informationssystem – zoom.
- Ianella, Renato 1998: *An idiot's guide to the resource description framework* [online]. – URL: [http://www.dstc.edu.au/RDU/reports/RDF-Idiot/](http://www.dstc.edu.au/cgi-bin/redirect/rd.cgi?http://archi-ve.dstc.edu.au/RDU/reports/RDF-Idiot/) (lesedato 2003-11-11)
- ISO 5963, dvs.: International Organization for Standardization. Technical Committee ISO/TC 46 1985: *Documentation : methods for examining documents, determining their subjects, and selecting indexing terms*. – [Geneve] : The Organization – iii, 5 s. – (International standard. ISO ; 5963)
- ISO 2788, dvs.: International Organization for Standardization. Technical committee ISO/TC46 1986: *Documentation : guidelines for the establishment and development of monolingual thesauri*. – 2nd ed. - [Geneve] : The Organization. – iii, 32 s. – (International standard. ISO ; 2788)
- ISO/IEC 13250 2002: *Topic Maps : information technology document description and processing languages* 2002: [online]. – URL: <http://www.y12.doe.gov/sgml/sc34/document/0129.pdf> (lesedato 2002-12-12)
- ISO/IEC 19756 2003: *Requirements for a Topic Map Constraint Language* [online]. – URL: <http://www.isotopicmaps.org/tmcl/requirements.html> (lesedato 2003-04-03)
- Lancaster, F. W. 1986: *Vocabulary control for information retrieval*. – 2nd ed. – Arlington : Information resources press. – 270 s.
- Lancaster, F. W. 1991: *Indexing and abstracting in theory and practice*. – Illinois : University of Illinois. – S. 1-85.
- Marchionini, Gary 1995: *Information seeking in electronic environments*. – New York : Cambridge university press. – S. 27-60, 100-138.
- Milestead, Jessica L. 2001: *Standards for relationships between subject indexing terms*. – S. 53-66. – I: Relationships in the organization of knowledge / edited by Carol A. Bean and Rebecca Green. – Dordrecht : Kluwer.- 232 s.
- Norsk digitalt bibliotek 2003: [online] / fABM-utvikling. – Oslo : ABM-utvikling, 2003. – URL: [http://www.abm-utvikling.no/Digbib\\_5.pdf](http://www.abm-utvikling.no/Digbib_5.pdf) (lesedato 2003-02-13)
- Opsahl, Per 1987: *Emneordbasert kunnskapsorganisasjon i konstruksjon av tesauri*. – Oslo : Statens bibliotekhøgskole. – 30 s.
- Parsia, Bijan 2003: *Semantic web services*. – S. 12-15. – I: Bulletin of the American society for information science and technology. – April/May, 2003
- Pepper, Steve 1999: *Navigating haystacks and discovering needles : introducing the new topic map standard*. I: Markup Languages : theory & practice 1.4 (1999). s. 41-68

Pepper, Steve 2002: *The XML papers : lessons on applying topic maps* [online] / Steve Pepper and Lars Marius Gars-hol. – URL: <http://www.ontopia.net/topicmaps/materials/xmlconf.html> (lesedato 2002-12-16)

Pepper, Steve 2002: *The TAO of topic maps : finding the way on the age of infoglut* [online]. – URL: <http://www.ontopia.net/topicmaps/materials/tao.html> (lesedato 2002-12-16)

Pepper, Steve 2002: *Ten theses on topic maps and RDF* - URL: <http://www.ontopia.net/topicmaps/materials/tdf.html> (lesedato 2003-01-14)

Pepper, Steve 2002: *Towards a general theory of scope*. [online] / Steve Pepper [and] Geir Ove Grønmo. – URL: <http://www.ontopia.net/topicmaps/materials/scope.htm> (lesedato 2003-04-04)

Pharo, Nils 1995: *Dokumentproduksjon i world wide web : retningslinjer og en brukerundersøkelse*. – Oslo : Høgskolen i Oslo, avd. for journalistikk, bibliotek og informasjonsfag. – 139 s.

Rice, Ronald E. 2001: *Accessing and browsing information and communication* / Ronald E. Rice, Maureen McCreddie, and Shan-Ju L. Chang. – Cambridge; Massachusetts : MIT press. – S. 173-215.

Schwebs, Ture 2001: *Tekst.no : strukturer og sjangrer i digitale medier*. – Oslo : Landslaget for norskundervisning : Cappelen akademisk forl., c2001. – S. 65-94.

Soergel, Dagobert 1974: *Indexing languages and thesauri : construction and maintenance*. – Los Angeles : Melville. – 632 s.

Sowa, John F. 2003: *Ontology, metadata, and semiotics* [online]. – URL: <http://users.bestweb.net/~sowa/peirce/ontology/meta/htm>

*The Standard Application Model for Topic Maps* 2002: [online]. – URL: <http://www.isotopicmaps.org/sam/> (lesedato 2003-02-10)

Steigan, Pål 2001: *En åpen, demokratisk nasjonal kunnskapsbase* / Pål Steigan og Steve Pepper [online]. – URL: <http://www.aftenposten.no/meninger/kronikker/article.jhtml?articleID=242930>

Svenonius, Elaine 2000: *The intellectual foundation of information organization*. – Cambridge, Massachusetts : MIT press. – 255 s.

*TMQL requirements (1.0.0)* 2001: [online]. – URL: <http://www.y12.doe.gov/sgml/sc34/document/0249.htm> (lesedato 2003-02-10)

*The topic map query language (TMQL)* [online]. – URL: <http://www.isotopicmaps.org/tmql> (lesedato 2003-01-30)

Vogt, Yngve (2002). *Setter emnekart på dagsordenen : norske bedrifter leder an internasjonalt* [online]. – URL: <http://www.computerworld.no/index.cfm?fuseaction=artikkel&id=F62EDE65-E382-5B2D-B56D15323D3645D4> (lesedato 2002-11-18)

W3C, dvs. World Wide Consortium 1994: [online]. – URL: <http://www.w3.org> (lesedato 2003-11-14)

Wersig, Gernot 1971: *Information – Kommunikation – Dokumentation : ein Beitrag zur Orientierung der Informations- und Dokumentationswissenschaften*. – München-Pullach : Verlag Dokumentation. – S. 52-107

*XTM 1.0* dvs. *XML topic maps (XTM) 1.0* 2001: [online]. – URL: <http://www.topicmaps.org/xtm/1.0/> (lesedato 2003-01-24)

*XML topic maps : creating and using topic maps for the web* 2003/Jack Park, editor and Sam Hunting, technical editor. – Boston : Addison-Wesley. – 605 s. – ISBN 0-201-74960-2

## TIDLIGERE UTGIVELSER AV ABM-SKRIFT

#1 Digitalisering av fotosamlinger

#2 Reform 94 og museene

#3 Befolkningens vurdering av folkebibliotekene

#4 Statistikk for arkiv, bibliotek og museum 2002

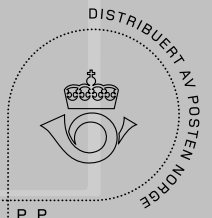
#5 Vei i vellinga. Håndbok i dokumentasjon av museumsgjenstander

#6 Sømløs biblioteklov for sømløse bibliotek tjenester

#7 Museumsarkitektur. En studie av nyere norske museumsbygg

skriftserien ligger tilgjengelig på [www.abm-utvikling.no](http://www.abm-utvikling.no) eller fås ved henvendelse til ABM-utvikling

**B**



NORGE

P. P.

**Returadresse:** ABM-utvikling  
Postboks 8145 Dep.  
N-0033 Oslo

ISBN 82-8105-009-8