

Kandidatnummer 121

**Emneindeksering og gjenfinning av
tverrfaglig forskning
-muligheter og utfordringer**

*As publishing is easier than ever before, it becomes more and more
difficult to find appropriate and important publications
and to keep track of the latest research results
(Kirchhoff, 2010, s. 8)*

Forord

Stor takk til Anders Olof Larsson, forsker ved Westerdals Oslo ACT, som gav meg ideen til oppgaven.

Likeså stor takk til Tine Lodberg Frost, som veiledet meg på rett kurs når kartet ikke lenger stemte overens med terrenget.

Men aller mest takk til mine tålmodige gutter, store og små. Endelig!

Sammendrag

Den voksende mengden med forskningsutgivelser gjør at det blir stadig vanskeligere å få oversikten over et fagområde. Samtidig fører digitaliseringen i samfunnet til at flere fagområder overlapper hverandre, men kanskje med ulike perspektiver. I denne oppgaven drøfter jeg hvilke hensyn som bør tas for å bedre tilgangen til, og gjenfinningen av forskning, ved å ta utgangspunkt i et prosjekt ved Westerdals Oslo ACT. Forskerne er bedt om å beskrive forskningen sin ved hjelp av nøkkelord, som deretter er planlagt brukt på nettsidene som en oversikt over, og en inngang til forskningen som foregår ved høyskolen. Flere av forskningsområdene er tverrfaglige og publikasjonene er på ulike språk. Jeg drøfter fordeler og ulemper med frie emneord, sammenlignet med et kontrollert vokabular. Jeg går også igjennom anbefalinger for autoritetskontrollerte termer, og ulike løsninger for å koble sammen begrep som har samme betydning, men som er på ulike språk. Deretter legger jeg frem et utvalg alternativer for å lenke forskning fra ulike forskere og avdelinger sammen, hovedsakelig basert på idéen om å ta i bruk -og bidra til- den semantiske web-en.

Min konklusjon er at en ekspertgruppe innen hvert fagområde, helst fra flere avdelinger, bør gjennomgå nøkkelordene, slik at relatert forskning blir beskrevet med de samme termene. Nøkkelordene bør deretter kontrolleres av bibliotekarene før de publiseres, slik at de følger formelle indekseringsregler og dermed lettere kan legges til rette for lenket data. Videre forskning kan være å undersøke en mulig integrering med bibliotekskatalogen.

Innholdsfortegnelse

1	Innledning	1
1.1	Bakgrunn og motivasjon	2
1.2	Problemstilling.....	4
1.3	Avgrensninger.....	5
1.4	Sentrale begreper	6
1.5	Oppgavens utforming.....	8
1.6	Etiske refleksjoner	8
2	Teoretisk grunnlag	9
2.1	Tidligere forskning	9
2.2	Litteratur i studien.....	10
3	Drøfting	12
3.1	Kontrollerte vs. frie emneord (naturlig språk).....	12
3.2	Utforming av et kontrollert vokabular.....	15
3.3	Forskning og publikasjoner på ulike språk.....	18
3.4	Flerspråklige verktøy.....	19
3.5	Søking i fulltekst	20
3.6	Automatiske forslag til termer	21
3.7	Forfatterens motiver	22
3.8	XML/RDF og Linked Data.....	23
3.8.1	Web-baserte semantiske verktøy	26
3.8.2	Ontologier.....	27
3.8.3	DBpedia	28
4	Konklusjon	29
4.1	Studiens relevans.....	31
5	Muligheter for videre forskning	31
6	Litteraturliste	33
6.1	Personlig kommunikasjon	38
7	Vedlegg	39
7.1	Vedlegg 1 Nøkkelord fra forskergruppen FACT	39
7.2	Vedlegg 2 Nøkkelord fra forskningsgruppen Information Systems.....	40
7.3	Vedlegg 3 LOD-sky.....	41
7.4	Vedlegg 4 Ordsky	42
7.5	Vedlegg 5 DocuBurst	43
7.6	Vedlegg 6 OntoFrame	44
7.7	Vedlegg 7 Samtykkeskjema	45

Figurer:

Figur 1: Taggmodell, s. 24

Figur 2: RDF-trippel, s. 24

1 Innledning

Det er ikke mer enn en generasjon siden at all litteratur som ble publisert var i trykt form. I det 21. århundre begynte informasjonsmengden å øke dramatisk, og en stor del av dette var forskning og faglige publikasjoner. På tross av de utallige mulighetene og verktøyene som finnes i moderne tid, har behovet for å effektivisere organiseringen av denne informasjonen bare blitt større.

Melvil Dewey formaliserte klassifikasjonen allerede i 1876 med sitt desimalklassifikasjonsskjema, DDK (Broughton, 2004, s. 5), og ble med tiden etterfulgt av andre som forsøkte å bøte på det de så som begrensninger med dette skjemaet.

Ulempen med disse skjemaene er at de ofte har kompliserte regler for å sette sammen en klassifikasjonskode, det kan være vanskelig å bedømme hva et dokument handler om, og kodene kan for lekfolk fremstå nærmest som hemmelig yrkeslingo som må dechiffreres for å gi mening. Et klassifikasjonsskjema kan heller ikke være statisk, men må stadig oppdateres slik at ny kunnskap og nye begreper kan implementeres, og utdatert kunnskap fases ut. Dette er en kognitiv arbeids- og tidkrevende prosess som krever stor interesse og kunnskap hos de som vedlikeholder skjemaet.

Parallelt med fremveksten av internett på begynnelsen av 1990-tallet, begynte brukere å sette stikkord på dokumentene sine (Vander Wal, 2007). Dette var en naturlig utvikling, da man raskt så behovet for å merke dokumenter slik at flest mulig som var interessert i det samme, lettere kunne finne frem til relevante nettsider og dokumenter. Men kildene til feilmerking er flere. Den kanskje mest åpenbare er ulike synspunkt om hvilke begrep som er beskrivende for et dokument, og dermed usikkerhet om dokumentet har en reell mulighet til å bli gjenfunnet av de(n) som har bruk for det. Andre utfordringer kan være språkproblemer som skrivefeil, «ord delings feil», bruk av både entall og flertall, ulike bøyninger og andre ordformer, samt inkonsekvent bruk av synonymer og ikke-definerte homonymer. Et annet problem kan være støy, det vil si for mange irrelevante treff som resultat av for mange -kanskje lite spesifikke- termer brukt for å beskrive et dokument. Det motsatte, en lite fullstendig treffliste, kan også bli resultatet i tilfeller der man har brukt få -kanskje også lite beskrivende eller generelle- termer.

Likevel er brukergenererte nøkkelord helt tydelig blitt en suksess, om man skal dømme ut fra hvor utbredt tagge-funksjonen er blitt, og bruken av begrepet «hashtag»/#/emneknagg i mediene og samfunnet ellers. De brukes hovedsakelig i sosiale medier for å beskrive hva et bilde, en tweet, eller andre dokumenter handler om, og kan dermed brukes til å finne andre dokumenter innen det samme temaet. Men eksempler viser at emneknagger også kan brukes i skjønnlitterære tekster som et språklig virkemiddel for å - gjerne humoristisk- beskrive en følelse eller en situasjon. Et eksempel er boka *#alfahann* av Jan Tore Noreng¹. Andre eksempler er den satiriske stripen *Dollhouse*² av Lene Hval, bokomtaler³ og avisartikler⁴.

Disse emneknaggene, på engelsk kalt tags eller tagger/tægger som en forforskning⁵, er altså blitt et godt etablert begrep i populærkulturen. Men hva er tendensen i academia?

1.1 Bakgrunn og motivasjon

I forbindelse med arbeidet med andre eksamensoppgaver ved HiOA, blant annet i *Research og analytisk referanseteknikk*, har jeg vært gjennom omfattende informasjonssøkeprosesser for å finne relevant faglitteratur, og blitt nokså godt kjent med flere av de større akademiske tidsskriftforlagene og databasene. Disse forlagene, i tillegg til de enkelte tidsskriftene, bruker ulike metoder for å emnebeskrive publikasjonene. Dette kan være tesauruser, som man blant annet finner i databasen *Library, Information Science & Technology Abstracts (LISTA)*, og emneordslister, eksempelvis hos *Library & Information Science Source (LISS)*, begge baser driftet av EBSCO. I tillegg kommer forfatterens egne nøkkelord, for eksempel EBSCOs «Author Supplied Keywords» og Web of Sciences «Author Keywords». Mange forlag og baser produserer også sine egne emneord. Både manuelt -for eksempel EBSCOs «Subjects» og IEEEs «INSPEC: Controlled Indexing»-, og automatisk indeksering, blant andre WoS' «KeyWords Plus», der termene er hentet fra artikkeltitler forfatteren av den gjeldende artikkelen har sitert (Thomson Reuters, 2015).

¹ <http://www.gyldendal.no/Barn-og-ungdom/Boeker-for-ungdom/alfahann2>

² <http://www.aftenposten.no/meninger/Avskaffe-au-pair-ordningen-Hvem-skal-da-ligge-i-ko-nar-det-er-burgunderslipp-pa-polet-615296b.html>

³ <http://www.nocomprendopress.com/product/flu-hartberg-moderator>

⁴ <http://www.aftenposten.no/familieogoppvekst/denfolelsen-nar-du-oppdager-du-er-gravid-i-uke-264-7898b.html>

⁵ Se oversikt over termer og begreper i kap. 1.5.

Å finne lignende artikler ved å klikke på et relevant dokumentets emneord eller nøkkelord er en måte å finne andre relevante publikasjoner på, men min erfaring var at dette ofte ble bomskudd. Jeg forundret meg flere ganger over stikkordene forfatterne hadde brukt for å beskrive artikkelen de hadde skrevet. De kunne være -etter min oppfatning- altfor snevre, for generelle, for faglige spesifikke og kanskje også strategisk valgte stikkord som et virkemiddel for å bli «gjenfunnet» oftere, og dermed øke sjansen for å bli sitert. Et eksempel på bruk av lite konkrete emneord som ikke skiller emnet fra et hvilket som helst annet fagområde, er en artikkel som omhandler undervisning i bibliotek- og informasjonsvitenskap ved et universitet i Botswana. Den er blant annet indeksert med forfatter-nøkkelordene «opportunities» og «challenges»⁶. Disse termene eksisterer også i tittelen, og kan være en annen utfordring om man ønsker en så fullstendig treffliste som mulig.

Et annet eksempel på et lite spesifikt nøkkelord, er en artikkel som handler om å heve de tekniske ferdighetene til bibliotekarer, blant annet indeksert med forfatter-nøkkelordet «Technology»⁷. Det er vanskelig å avgjøre om dette er grunnet forfatterens og mins ulike oppfatninger av hva et begrep skal beskrive, for eksempel grunnet ulikt morsmål eller ulik fagbakgrunn, eller om begrepet som er brukt faktisk er for generelt. Andre eksempler er de mye brukte begrepene «IT», «information» og «communication».

Nærmest ved en tilfeldighet kom jeg over et prosjekt som skulle gjennomføres ved Westerdals Oslo School of Arts, Communication and Technology (WOACT). Lederen for forskergruppa fACT har tatt initiativ til å samle inn nøkkelord fra alle de faglige ansatte som skal beskrive den forskningen de arbeider med og publikasjonene de har skrevet, først og fremst for at det skal bli enklere for eksterne å se hva det forskes på. Jeg fattet raskt interesse for dette prosjektet, og ba om å få bli med som observatør under prosessen, noe de sa seg villige til.

Flere av de ansatte forsker på emner som ikke bare er relevante for forelesernes egne studenter, men også for studenter på andre avdelinger, kollegaer eller eksterne studenter og andre interesserte, for eksempel journalister. Denne forskningen burde ideelt sett vært

⁶ Tilgjengelig via abonnementsbaserte databaser; Web of Science eller EBSCOhost:

Jain, P. & Jorosi, B. N. (2017). LIS Education at University of Botswana: Evolution, Opportunities and Challenges. *Desidoc Journal of Library & Information Technology*, 37(1), 42-50.
doi:10.14429/djlit.37.1.108643

⁷ Tilgjengelig via abonnementsbasert database; Web of Science:

Morato, J., Sanchez-Cuadrado, S. & Fernandez-Bajon, M. T. (2016). Trends in the technological profile of information professionals. *Profesional De La Informacion*, 25(2), 169-178.
doi:10.3145/epi.2016.mar.03

tilgjengelig fra alle relaterte studie- og forskningssider på WOACTs nettsider, ikke kun den enkelte avdeling hver forsker jobber på, eller på den enkelte ansatt-siden.

Min første magesfølelse da jeg hørte om prosjektet, var at «dette burde være en jobb for «kunnskapsorganisasjonsarbeidere». Men samtidig kom tvilen, da tankene om hvor arbeidskrevende det er å lage, og ikke minst vedlikeholde, en godt fungerende emneordliste kom snikende. I tillegg, tesauruser –om det skulle bli aktuelt- blir som regel til mer planlagt med den hensikt å dekke ett spesifikt fagområde (Broughton, 2006, s. 40, 43), og her er det jo snakk om tverrfaglig samarbeid. Dersom målet er å finne nye, kompletterende måter å formidle kunnskap på, bør da forskning alltid indekseres av en som har kunnskap om kunnskapsorganisasjon og formidling? Eller kan ikke forfatterens beskrivelser av egen forskning og egne publikasjoner være vel så gode? Har jeg noen forutinntatte antagelser om at beskrivende emneord basert på en informasjonsvitenskapelig vinkling er mer effektivt og hensiktsmessig, enn stikkord formulert av forfattere med åpenbart stor faglig kunnskap, men kanskje ikke innen klassifikasjon og gjenfinningsteori? Dette var noe jeg ønsket å undersøke videre.

1.2 Problemstilling

Den økte veksten i vitenskapelige publikasjoner har gjort det nødvendig å stadig forbedre indeksering og lage nøyaktig deskriptiv og innholdsmessig korrekt metadata. Målet er å legge bedre til rette for gjenfinning av ressurser og gjenbruk av forskningsdata. Denne informasjonen har lenge vært basert på menneskelig intellektuelt arbeid, men stadig mer forskning gjøres på å automatisere denne prosessen, for eksempel ved bruk av dataprosessert tekstanalyse av naturlig språk. Målet er at formaliserte emnebeskrivelser kan tilrettelegge for utnyttelse av den semantiske web-en, slik at relaterte emner kan kobles sammen, og ny kunnskap oppdages.

En oversikt over nøkkelord fra forskere fra Westerdals Oslo ACT viser at nesten ingen av nøkkelordene er like. Dersom dette er overlappende fagområder, hvordan bør dette kobles sammen? Min overordnede problemstilling blir derfor følgende:

På hvilken måte kan man best synliggjøre og tilrettelegge for gjenfinning av forskning, på tvers av ulike fagområder og på ulike språk?

For å undersøke dette vil jeg drøfte hvilke utfordringer eller fordeler frie emneord har sammenlignet med et kontrollert vokabular. Jeg vil også se på ulike løsninger for hvordan relaterte emner kan kobles sammen.

1.3 Avgrensninger

Det er utgitt mye vitenskapelig litteratur om sosial klassifikasjon (det vil si brukernes bidrag), blant annet av Spiteri (2007), men jeg kommer til å avgrense oppgaven til å omhandle forfatteres emnebeskrivelser, såkalt «narrow folksonomy» (Peters, 2009, s. 166), da det kun er dette som skal brukes i WOACT-prosjektet. En del av faglitteraturen om folksonomier vil likevel kunne ha en overføringsverdi i studien. En samling av forfatternes nøkkelord (i en ordsky eller lignende) kan for eksempel bli sett på som en folksonomi om man bruker den smale definisjonen, det vil si en samling av nøkkelord fra flere bidragsytere, men der kun opphavspersonen kan tagge egne ressurser (s. 166).

Opprinnelig var emnebeskrivelsen av forskningen på WOACT et prosjekt som skulle foregå i løpet av høstsemesteret 2016, og ferdigstilles før jul. Fordi jeg på grunn av studier ikke kunne følge dette prosjektet i så stor grad som jeg ønsket, var tanken at jeg skulle undersøke hvordan beskrivelsen og lenkingen var gjennomført etter at arbeidet var gjort. Jeg skulle deretter gjøre mine betraktninger om det -etter mitt syn- hadde blitt gjort på en best mulig måte, med tanke på muligheten for gjenfinning og eventuelt oppdagelse av relatert forskning og litteratur. Av ulike årsaker har prosjektet blitt forsinket, og fremdriften er kommet til innsamling av nøkkelord fra de faglige ansatte. Ulempen med denne forsinkelsen er at jeg nå ikke får sett en modell i virksomhet, og får dermed heller ikke undersøkt hvordan en valgt løsning fungerer. På den annen side vil jeg nå være mindre farget av en ferdig løsning, og kanskje mer åpen for alternative løsninger.

Andre avgrensninger jeg har gjort er å utelate avanserte matematiske modeller, slik som for eksempel i *Towards robust tags for scientific publications from natural language processing tools and Wikipedia* (Łopuszyński, M. & Bolikowski, Ł. (2015)).

2.2 Om Westerdals Oslo ACT

Høgskolene Nordisk Institutt for Scene og Studio (NISS), Norges Informasjonsteknologiske Høgskole (NITH) og Westerdals Høgskole fusjonerte i 2014 til Westerdals Oslo School of

Arts, Communication and Technology (Westerdals Oslo ACT, udatert). Den nye høyskolen har nå fått en faglig spennvidde som strekker seg fra informasjonsteknologi, skuespillteknikk, tekstforfatter, maske- og hårdesign, regi, app-utvikler, kulturledelse, musikkprodusent, lystekniker, strategi, lydtekniker, fotograf og mange, mange fler. Studentene fra de ulike avdelingene samarbeider i skoleprosjekter som konserter, teaterstykker og lignende, og mange fagområder er involvert i hvert fellesprosjekt.

Enkelte av de fusjonerte høyskolene er nokså ferske som en høyere utdanningsinstitusjon. Mange av fagretningene er svært praktiske lagt opp, og har tidligere hatt godkjenninger som fagskole eller videregående skole, evt. årsheter i høyere utdanning. En akkreditering som høyere utdanning fører med seg noen krav, for eksempel at institusjonen må ha fagbibliotek, minimum antall ansatte med førstekompetanse m.m. Denne «akademiseringen» har vært en omstilling for mange, blant annet med søknadsskriving om opprykk, samt forskning.

Mange av studieretningene har enkeltemner som er svært like, og det er ikke uvanlig at faglig tilsatte underviser på flere avdelinger. Dette viser seg også i den faglige bredden de ansatte har. Mange som underviser i lyd- og musikkteknikk, har bakgrunn i utøvende musikk. En foreleser i prosjekt- og kulturledelse skrev sin doktoravhandling om humor i film og TV. En annen foreleser på samme avdeling har skrevet flere publikasjoner om informatikk. Dette øker sannsynligheten for at de fagansattes forskning er relevante for flere enn én avdeling, og bør formidles tverrfaglig.

1.4 Sentrale begreper

Siden prosjektet fortsatt er i startfasen, er det ikke fastsatt nøyaktig hva resultatet av de innsamlede nøkkelordene skal være. I drøftingen min kommer jeg derfor til å benytte flere termer fra litteraturen om folksonomier, selv om oppgaven min kun omhandler opphavspersoners nøkkelord. Blant annet sidestiller jeg tagg og nøkkelord. Der ikke annet er referert, er definisjonene hentet fra Ordnett.no⁸.

- **Folksonomy (på norsk «folksonomi»):**

Et begrep satt sammen av «folks» og «taxonomy», først brukt av Thomas Vander Wal; «A user-generated system of classifying and organizing online content into different

⁸ <https://ordnett.no/>

categories by the use of metadata such as electronic tags» (2007). Begrepet ble først brukt om personlig tagging for eget bruk, men er senere blitt utvidet til å inkludere system der brukerne bidrar kollektivt til samlingen av tagger (Vander Wal, 2007). Blir også kalt social tagging, folk classification, ethnoclassification, collaborative tagging, social classification, distributed classification, social indexing, social annotation, emergent semantics, open tagging, free tagging, grassroot classification, mob indexing, shared metadata og Wisdom of the crowds.

▪ **Homonymer:**

To eller flere ord som lyder og/eller skrives likt, men har forskjellig opprinnelse og betydning. Innen klassifikasjon må homonymer ha kvalifikatorer for å skille én betydning fra en annen (Broughton, 2006, s. 214).

For eksempel: myse (se med sammenknepne øyne), myse (melkeprodukt).

- **Indeksering:** I tradisjonell forstand betyr begrepet indeksering å klassifisere informasjon ved hjelp av nøkkelord som stilles opp i en indeks (Broughton, 2004, s. 302). I denne oppgaven kommer jeg til å legge en mer utvidet betydning i begrepet, slik at det også omfatter forskernes beskrivelse av innholdet i egen forskning og publikasjoner, og nøkkelordene de har valgt ut for å beskrive denne. Av hensyn til tekstflyt har jeg også valgt å bruke verbbegrepet i substantivform, *en indekserer*, for å betegne en som indekserer innhold i ressurser.

▪ **Ontologi:**

En modell eller oversikt over et fagområde der relasjonene mellom konseptene i feltet er spesifiserte (Broughton, 2006, s. 218).

▪ **Polysemi:**

Flertydighet, at et ord har flere betydninger.

▪ **Ressurs:**

I denne sammenhengen brukt om et hvilket som helst dokument i en hvilken som helst form som en bruker kan søke etter.

- **Tag/tagg/tægg:**
 - a) (I IT) identifiserende element; «etikett».
 - b) Merkelapp i form av ord eller fraser som beskriver en ressurs (Glushko, 2013, s. 144).
 - c) Brukergenererte, gjenbrukbare merknader og søkbare emneord som kan brukes på alle mediatyper (Voß, 2007).Brukes i denne oppgaven jevnbyrdig med nøkkelord eller stikkord.

- **Term:**
 - a) Fagord med fast betydning; faguttrykk.
 - b) Ord eller sammensatt uttrykk brukt til å representere et begrep (Almo, 2011).

1.5 Oppgavens utforming

Oppgaven følger oppbygningen til en litteraturbasert oppgave. I innledningen skriver jeg litt om motivasjon for studien, problemstilling, oppgavens omfang og avgrensning, og kort om sentrale begreper. Deretter følger en utgreiing om min tidligere tilknytning til høyskolen, og litt om tidligere forskning innen området frie emneord og kontrollerte vokabular. Videre gir jeg en beskrivelse av det teoretiske grunnlaget jeg har brukt i oppgaven, før jeg drøfter ulike muligheter og utfordringer det ligger i tverrfaglig forskningsformidling, basert på en samordning av emnebeskrivelser. Til slutt trekker jeg konklusjoner og foreslår muligheter for videre forskning.

I kapitler der jeg drøfter ulike verktøy eller teknologi rettet mot lenket data eller flerspråklighet, har jeg fokusert på prinsippene bak ideen eller teknologien som brukes til å gjøre data tilgjengelige, mer enn en fullstendig beskrivelse av hvert enkelt verktøy.

1.6 Etske refleksjoner

Jeg har tidligere vært ansatt (≈13 år) ved Nordisk Institutt for Scene og Studio (NISS) som høyskolelærer i avdeling for lydproduksjon, senere i WOACT som bibliotekar i et 9 måneders vikariat i en stilling som hovedbibliotekar på Campus Vaterland. Min tilknytning til prosjektet er derfor ikke nøytral, men jeg har så langt det har latt seg gjøre forsøkt å holde en profesjonell distanse. En positiv konsekvens av mine tidligere arbeidsforhold, er at jeg har en nokså god innsikt i de ulike avdelingene og hvilke fagområder som er overlappende på tvers

av de ulike studieretningene. Min erfaring som bibliotekar hos WOACT, og med dette observasjoner av hvordan det store flertallet av studentene bruker biblioteket, tror jeg også kan hjelpe meg i undersøkelsen av hvordan den beste løsningen bør være for å best kunne formidle institusjonens forskning. Samtidig utviklet WOACT nye nettsider under mitt vikariat, og biblioteksidene ble ikke publisert før etter vikariatet var avsluttet. Jeg har derfor ingen forhåndskunnskaper om hvordan de nye sidene blir brukt i søk etter faglitteratur.

2 Teoretisk grunnlag

2.1 Tidligere forskning

Det er de siste 10 årene skrevet mye om folksonomier eller sosial tagging, blant annet av Louise F. Spiteri (2007), Gene Smith (2008) og Isabella Peters (2009), men få studier har tatt for seg forfatteres nøkkelord (Gil- Leiva & Alonso- Arroyo, 2007, s. 1181; Lu & Kipp, 2014, s. 483). Av nyere forskning er særlig Margaret E. I. Kipps forskning relevant i min studie, da hun blant annet har gjennomført undersøkelser der hun sammenligner tagger påført av brukere, forfattere og profesjonelle indekserere. Hennes funn viser at brukernes tagger og forfatterens nøkkelord ofte var likere, sammenlignet med nøkkelord og kontrollerte emneord, men at termene fra alle tre kildene alt i alt ofte overlappet hverandre (2005, 2011a).

Gil- Leiva and Alonso-Arroyo har undersøkt forfatteres nøkkelord på vitenskapelige publikasjoner, og fant at nesten halvparten overlappet med de kontrollerte emneordene. De konkluderte dermed med at nøkkelordene kan være en viktig kilde ved indeksering av vitenskapelige artikler (2007, s. 1179).

En tidlig studie utført av Voorbij sammenlignet emnebeskrivelser angitt av bibliotekarer og nøkkelord fra titler i 475 monografier tilgjengelig fra Nederlands folkebibliotekskatalog (1998). Resultatene viste at flere dokumenter ble gjenfunnet ved bruk av emnebeskrivelser, enn ved titler. Årsaken ble blant annet begrunnet med at titlene ikke alltid avslører nok informasjon om innholdet, men en viktigere grunn var alle de ulike måtene et emne var beskrevet på i titlene. Man så derfor viktigheten av å bruke et kontrollert vokabular (s. 468).

En studie fra 2005 av Gross & Taylor, replisert av Gross, Taylor & Joudrey i 2015, viste at gjennomsnittlig 35,9% (2005) og 27% (2015) av dokumentene ville manglet i trefflisten ved

søk i en bibliotekskatalog, dersom det kun ble søkt med nøkkelord, sammenlignet med et søk med kontrollert vokabular (Gross & Taylor, 2005; Gross, Taylor & Joudrey, 2015).

2.2 Litteratur i studien

Som nevnt i kapittel 2.1, er det ikke forsket mye på forfattertagger. Kipp er én av få som har utført studier om taggepraksiser på forskningsorienterte sosiale bokmerkingssider som CiteULike. Selv om hun hovedsakelig har undersøkt brukeres (lesernes) og ikke forfatternes nøkkelord, anser jeg forskningen hennes som relevant, da jeg antar at det er stor sannsynlighet for at brukere av vitenskapelige bokmerkingssider i mange tilfeller også er vitenskapelige forfattere. Hun har også samlet inn data der hun sammenligner tagger, forfatternes nøkkelord, og profesjonelt indekserte emneord, som blant annet har resultert i artiklene *Tagging of biomedical articles on citeulike: A comparison of user, author and professional indexing* (2011a) og *User, Author and Professional Indexing in Context: An Exploration of Tagging Practices on CiteULike* (2011b). Et argument mot å legge for stor vekt på spesielt den sistnevnte artikkelen, er at innsamlingen av forfatternes nøkkelord ble gjort fra artikler som omhandler bibliotek- og informasjonsvitenskap. Om man går ut fra at mange av forskerne bak slike artikler selv er bibliotekarer, er det naturlig å tenke seg at forfatternes nøkkelord var av høyere kvalitet, sett fra et profesjonelt indekseringsståsted, enn det som er vanlig. Kipps konklusjon er at profesjonelt indekserte emneord oftere la vekt på stedsinformasjon i emne- eller innholdsbeskrivelsen («aboutness») enn brukerne og forfatterne (2011b, s. 43). I tillegg kan vokabularet i tradisjonelle systemer forbedres og utvides av bidrag fra artikkelforfattere og taggere, på grunn av taggenes eller nøkkelordenes nærhet til fagfeltet (Kipp, 2011a, s. 245). Brukerne la også, naturlig nok, større vekt på personlige tagger som informasjon om organisering av tid eller oppgaver (2005, s. 434).

En sentral del av oppgaven er sammenligningen av kontrollerte vokabular med frie emneord. Her har jeg brukt Bürger et al. (2010) som en sentral referansekilde, grunnet den omfattende rapporten om bruk av stikkord på den semantiske web-en. Formålet er å motivere sluttbrukerne til å bidra med mer formelle nøkkelord på et høyere uttryksnivå, og med en bedre struktur (2010, s. 11). Universitetsbiblioteket i Oslo og deres arbeid med *Realfagstermer* (2016) er også relevant i en undersøkelse om bruken av et kontrollert vokabular sammenlignet med frie emneord.

Innen feltet klassifikasjons- og indekseringsteori, har jeg brukt sentrale verk av Broughton (2004, 2006) og Lancaster (2003), i tillegg til Spiteri (2007) og hennes anbefalinger om bruk av tagger i en bibliotek katalog. I tillegg har jeg brukt Almo (2011), Suonuuti (2015) og Marshall (2003), samt ulike internasjonale standarder når det gjelder oppbygning av kontrollerte vokabular og arbeid med termer. Robertson & Greenberg (2002) sin studie har jeg brukt som kilde i drøftingen av forfatteres indeksering av eget innhold. Michael Buckland (1999), Jane Greenberg (2002) har jeg brukt i mine refleksjoner om mulige avvik mellom brukernes og forfatterens valg og bruk av termer, og konsekvensene av disse ulike tolkningene.

Fordelene med brukernes tagger og hvordan disse kan oppdatere og forbedre emnebeskrivelsene og søkeinnngangene i bibliotek katalogene, er blitt undersøkt av Kakali & Papatheodorou (2010), i tillegg til Kipp (2011b). Samtidig er frie emneord blitt beskrevet som både rotete og ustrukturert av Gene Smith (2008) og Hammond, Hannay, Lund & Scott (2005). Denne mangelen på semantisk struktur og språkkontroll, og konsekvenser av ulike tolkninger av tagger er undersøkt av Huang & Chuang (2009) og Kolbitsch (2007). Rosenfeld, Morville, & Arangos *Information architecture: for the web and beyond* (2015) gir fornuftige råd om oppbygning, struktur og navigering innen digital informasjon, og inneholder også et kapittel om tesauruser, kontrollerte vokabular og metadata (kap.10). Jeg anser denne kilden som relevant, grunnet grensesnittet nøkkelordene er tenkt å fungere i.

Drøfting om søking i fulltekst og utfordringer knyttet til dette, er basert på forskning av Pera, Lund & Ng (2009) og Jeffrey Beall (2008).

I kapittelet om taggeforslag, har jeg tatt utgangspunkt i Matthews et al.s undersøkelse om gjenfinningskonsistens på litteratur innen IT og datavitenskap, der de sammenlignet publikasjoner med kun sosiale tagger, og publikasjoner med tagger kombinert med et autorisert vokabular (2010). Fordelene med et taggesystem utdypes også av Marlow, Naaman, Boyd & Davis (2006) og Golder & Huberman (2006). Van Damme, Hepp & Sioarpaes (2007) presenterer prinsippet om «kollektiv visdom»; om hvordan flertallets bruk fører at en tags betydning befestes. Mathes (2004), Peters (2009) og Guy & Tonkin (2006) påpeker derimot faren for at tagger kan bli knyttet til ressurser som i utgangspunktet ikke ville blitt assosiert med disse. Mulighetene for å kombinere taggeforslag med kontrollerte termer, er blitt

undersøkt av Golub, Lykke & Tudhope, og Golub, Moon, Tudhope & Nielsen (2014; 2009). Marlow et al. (2006), i tillegg til Arch (2007), er kilder i omtalen av «spam tagger».

Innen temaet semantisk web, er det naturlig å henvise til web-ens «far», Tim Berners-Lee (i Bizer, Heath & Berners-Lee, 2009) og World Wide Web Consortium, særlig angående formatet RDF/XML (2009a, 2009c, 2016). Jeg referer også her til tidligere nevnte Bürger et al. (2010).

Mye er skrevet om lenket data, og i en oppgave som denne er det vanskelig å plukke ut kun noen få kilder, men Miao, Meng, Fang, Nishino & Igata har skrevet en høyst relevant artikkel i denne sammenhengen, kalt *Link scientific publications using linked data* (2015). Fra en lignende kontekst, nærmere bestemt utdanning og opplæring, referer jeg til Dietze et al. (2012) og deres forskningsprosjekt der målet er å kunne dele ressurser innen utdanning på en semantisk sosial plattform.

Utfordringene med lenket data som hensynet til personvern og andre problemer som oppstår når data slettes, flyttes eller på andre måter endres, drøftes av Bizer, Heath & Berners-Lee (2009). Løsninger på disse utfordringene har blitt foreslått av Giasson (2006), med verktøyet «Ping the Semantic Web».

Ulike teknologier og verktøy er valgt ut som eksempler i oppgaven for å beskrive bredden på forskningen rundt lenket data. Blant de utvalgte er WordNet (Princeton University, 2015), beskrevet i en artikkel av Bailin & Truszkowski (2002), og DBpedia (2017a) omtalt av Miao et al. (2015). De Keyser (2012), Noy & McGuinness (2001) og McGuinness (udatert) har gitt meg bakgrunnsstoffet om ontologier, i tillegg til eksempelet om OntoFrame (de Keyser, 2012).

3 Drøfting

3.1 Kontrollerte vs. frie emneord (naturlig språk)

Klassifisering av konsepter som fagområder og emner kan være utfordrende dersom indekseringen skal utføres av noen som ikke selv er eksperter på emnet. Faren ligger i at termene som blir valgt for å beskrive ressursen hverken reflekterer forfatterens intensjon, eller brukerperspektivet (Buckland, 1999). I prinsippet burde det derfor være større sannsynlighet for at man får en mer korrekt beskrivelse av et emne dersom forfatterne selv indekserer sine

egne publikasjoner. Jane Greenberg formulerer det slik: «Scientists are intimate with their research, they want their research to be discovered and consulted, and they are likely to have a grasp of the natural language searching vocabulary used by specific audiences» (Miksa et al., 2002, s. 528). Men dette naturlige språket som ofte blir brukt i søk, inkludert nye konsepter og begreper, bruker ofte lang tid på å implementeres i kontrollerte emneordslister som Library of Congress Subject Headings (LCSH)⁹ eller tesauruser som Humord¹⁰. Det er kostbart å bruke dedikerte eksperter til å organisere ressurser eller bygge kontrollerte vokabularer, og ofte ser man også forsinkelser fra publikasjonene utgis, til vokabularet blir oppdatert. Dette kan komme av indeksererens arbeidsmengde, eller at fagområdet er i rask endring og at det må gjøres grundige undersøkelser før man kommer frem til hvilke begreper som er de korrekte (Bürger et al., 2010, s. 27). Som et eksempel har verken LCSH eller Humord implementert «folksonomy»/«folksonomi» i listene sine, selv om et søk på «folksonomy ELLER folksonomi» i Oria gir 2423 treff på alt fra indonesiske til tsjekkiske utgivelser, hvorav engelskspråklige utgjør 2347 publikasjoner¹¹. Kakali & Papatheodorou foreslår at bibliotekene bør evaluere og fornye innholdet i emneregistrene sine ved å legge til nye termer og eventuelle relasjoner, og erstatte eksisterende emneord med noen som er mer formålstjenlige. Årsaken er at de mener frie nøkkelord som sosiale tagger direkte uttrykker utviklingen av et vitenskapelig område (2010, s. 200). En metode som blir foreslått, er at bibliotekarene (eller andre som vedlikeholder vokabularet) undersøker sosiale taggesystemer, slik at man får en oversikt som viser hvilke termer og begrep som faktisk blir brukt til å beskrive ulike konsepter. Dette kan deretter bli innlemmet i det kontrollerte vokabularet (Mathes 2004; Hammond et al. 2005; Morville 2005, gjengitt av Kipp 2011, s. 45). Denne kombinasjonen av brukerspråk og standardisert språk vil dermed kunne samle relaterte ressurser.

Samtidig har frie emneord eller tagger en svakhet ved at det vanligvis ikke finnes noen autoritetskontroll som kan sile ut inkonsekvent bruk av ordformer og stavefeil (da hadde de jo heller ikke vært «fri»), lage referanser og lenker mellom relaterte begrep og synonymer, eller definere homonymer og polysemi (Spiteri, 2007). Forskningsresultater viser at samarbeidende emnebeskrivende systemer som mangler en viss form for semantisk språkkontroll, gjør indeksering, søk og navigasjon vanskelig (Kolbitsch, 2007, s. 28). Et eksempel er de gamle

⁹ <http://www.loc.gov/aba/publications/FreeLCSH/freelcsh.html#About>

¹⁰ <http://www.bibsys.no/files/out/humord/>

¹¹ Søk utført 20. april 2017.

instituttbibliotekene under Realfagsbiblioteket, som tidligere benyttet frie emneord. Disse gav hverken god presisjon eller fullstendighet i gjenfinningen, og det ble derfor satt ned en gruppe i 2009 som skulle jobbe med å utvikle et kontrollert vokabular. Resultatet ble vokabularet «real-fagstermer» (Universitetsbiblioteket i Oslo, 2016).

Mangelen på konsistens, sammen med ulike tolkninger om hva de frie emneordene skal representere (Huang & Chuang, 2009, s. 341), reduserer mulighetene emneordene har til å hjelpe brukerne finne relevante ressurser. Problemet er blitt beskrevet som «taggesuppe» (Hammond et al., 2005) og «messy metadata» (Smith, 2008, s. 56-57), det vil si innholdsbeskrivende metadata (i dette tilfellet tagger) uten en klar struktur, form eller bruksområde. I tillegg, -der frie emneord på den ene siden har en fordel grunnet deres umiddelbare implementering av nye begrep og uttrykk, er det på samme tid også deres ulempe. Kontrollerte emneordslister er laget for å vare lenge, har vært i bruk lenge, og er brukt på millioner av dokumenter. Tagger er i sin natur ofte kortvarige, og er lite brukt på ressurser som ikke er elektroniske, -det vil si det aller meste som hittil er utgitt. En konsekvens av denne flyktigheten kan man se på sosiale bokmerkingssider, der svært mange er blitt avviklet, eller ikke oppdateres. Eksempler på slike er Connotea.org, Simpy.com, Technorati.com, Furl.net, Spurl.net, Tagcloud.com, Shadows.com og Unalog.org. Dette kan være tegn på at mer kontrollerte emnestrukturer, såkalt «top-down»-strukturer har en verdi (Rosenfeld et al., 2015, s. 128).

En gjennomgang av nøkkelordene fra fACT viser at nesten ingen av ordene er like ([vedlegg 1](#)). Er det fordi ingen forsker på det samme eller har skrevet om det samme, eller er det fordi de beskriver fagområdene på ulike måter? Er det kanskje en fordel at de bruker flere ulike nøkkelord, slik at emnene får flere innganger? Forskning på gjenfinningssystemer har funnet at brukerdefinerte tagger kan beskrive alternative synsvinkler som ikke er til stede i et kunnskapsorganisasjonssystem, spesielt nye konsepter. De kan derfor bidra til å oppdatere kunnskapsorganisasjonssystemet (Matthews et al., 2010, s. 448). Ved å sjekke rettskrivning og gjøre bibliotekfaglige vurderinger av disse nøkkelordene, vil man få fordelene med å ha et autoritetskontrollert vokabular som samtidig beskriver publikasjonene innen gitte forskningsområder på en faglig korrekt måte. Denne løsningen har også blitt foreslått av Bürger et al.; å la de som tagger definere sine forskningsområder og nøkkelord, som deretter må godkjennes av spesialister på indeksering og av eksperter på fagområdet (2010, s. 27).

Et alternativ til den manuelle indekseringen, er å bruke WordNet, en engelskspråklig leksikalsk database med et kontrollert vokabular. Denne kan både brukes som et oppslagsverk¹², for eksempel for å sjekke staving, men kan også brukes til å automatisk analysere tekst med naturlig språk og gruppere eller «mappe»¹³ ord som har samme betydning (Princeton University, 2015). Den viser relasjoner som blant annet synonymy og hierarkiske relasjoner, og den kan skille homonymer fra hverandre. Bruksområdene kan være å trekke ut semantisk informasjon fra forskningsartikler, et tenkt scenario i en artikkel av Bailin & Truszkowski. Her utreder de om bruksområder for en egenutviklet kommunikasjonsprotokoll for informasjon- og teknologiutveksling (2002, s. 16-18). Ulempen er at vokabularet i WordNet er av generell art, og inneholder ikke fagspesifikke termer som for eksempel «inbetweenness», som en forsker fra WOACT har brukt ([vedlegg 1](#)). Man kan heller ikke søke på såkalte flerordstermer (Suonuuti, 2015), som «social media».

3.2 Utforming av et kontrollert vokabular

Kipp har sammenlignet brukeres tagger, forfatteres nøkkelord og kontrollerte termer, og funnet at forfatterens nøkkelord, blant annet i forbindelse med publisering av journalartikler, ikke er mye brukt utenfor det aktuelle, ofte smale, fagområdet. Det kan også være spesifikt for artikkelen (2007, s. 3). I tillegg viste en studie utført av Robertson & Greenberg at emne eller «aboutness» var den typen innholdsbeskrivende metadata vitenskapelige forfattere som oftest ønsket hjelp av katalogiseringsekspertene til å lage (2002, s. 51). Selv om man ikke har til hensikt å lage en thesaurus eller en kontrollert emneordliste, bør nøkkelordene derfor likevel formaliseres. Alle nøkkelordene vil da følge de samme formreglene, og kan på den måten hjelpe en bruker til å velge riktig form av en term i en søkeprosess. De aller fleste søkemotorer søker ordrett på søketermen, (bortsett fra at for eksempel Google kan gi andre forslag ved antatt feilstavede ord). Søker man etter «lese», får man for eksempel ikke treff på «lesing». En kontrollert liste med termer kan tilrettelegge bedre for en mulig fremtidig lenking mot autoritetsregistre, emneordlister, mapping av termer og lignende. Søking ved bruk av naturlig språk kan føre til mye «støy» på trefflistene, det vil si at man får for mange treff på dokumenter uten relevans (Broughton, 2004, s. 84). Ikke minst vil konsistente nøkkelord også se ryddigere ut.

¹² <http://wordnetweb.princeton.edu/perl/webwn>

¹³ Begrepet «mapping» har ingen tilsvarende ord på norsk, men jeg bruker det i denne sammenhengen som et begrep for å beskrive at man skal forbinde termer i én gruppe, med ett eller flere termer i en annen gruppe.

Broughton (2006, s. 70) anbefaler å gå gjennom de følgende punktene når man skal lage en indeksert liste med søkbare termer. En annen viktig kilde er *ISO 25964-1:2011 Information and documentation — Thesauri and interoperability with other vocabularies — Part 1: Thesauri for information retrieval*¹⁴:

- Reduser antall termer ved å gå bort ifra bruken av naturlig språk, til fordel for et begrenset sett av termer
- Når ulike termer har samme betydning (synonymi), må det avgjøres hvilken som skal brukes, såkalt «foretrukken» eller «anbefalt term».
- Dersom termene har forskjellige stavemåter, bør man bestemme seg for hvilken som skal brukes
- Det må avgjøres bruk av entalls- eller flertallsform av termene, og eventuell bruk av punktum. Som regel setter man alt som kan telles («hvor mange») i flertall, for eksempel «bøker», og «hvor mye» i entall, for eksempel «politikk» (Almo, 2011).
- Sammensatte termer («mediepolitikk»), og flerordstermer («politisk kommunikasjon») bør beholdes dersom de er registrert slik og ment å være ett begrep, da de ellers kan miste sin betydning, eller bli for generelle (Kipp, 2011b, s. 43). Bli enig om når og hvordan ord skal kombineres, samt ordrekkefølge. Denne prosessen er enda viktigere når automatisering er involvert, siden søkeprogramvaren vil kreve (nesten) nøyaktig samsvar mellom de indekserte termene og søkeordet.

(Broughton, 2006)

I tillegg kommer kontroll av homonymer, det vil si identiske termer som har ulik betydning, der meningen må gjøres utvetydig. Dette gjøres som regel med en forklaring (kvalifikator) i parentes (Broughton, 2006, s. 77). I prosjektet til WOACT anser jeg dette som unødvendig, da vokabularet vil være nokså emnespesifikt og entydig. Eventuelt kan man legge til flere termer og lage en flerordsterm.

Adjektiv og adverb bør ikke stå alene, da de sjelden gir mening som emnebeskrivende termer (s. 82). Som et eksempel bør «Big Data» og «Behavioural economics» stå som egne flerordstermer.

¹⁴ www.standard.no utgir ISO-sertifiserte standarder for terminologiarbeid, men dette er en betalingsløsning så et alternativ kan være NISO sin fritt tilgjengelige *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (Rosenfeld et al., 2015, s. 108-109); <http://www.niso.org/>. NISO representerer USAs interesser i International Organization for Standardization (ISO).

Infinitivform av verbet («å organisere») og fortids- og presensformer («organisert», «organiserer») brukes sjelden, unntatt der de brukes som et adjektiv i setninger («organisert kriminalitet»). Det vanlige er å bruke formen handlingssubstantiv, (norsk eksempel «organisering», engelsk eksempel «marketing») (International Organization for Standardization, 2011, s. 23-24).

Termer bør skrives med små bokstaver, men unntak kan gjøres for forkortelser, akronymer, egennavn eller termer som i vanlig bruk skrives på en særegen måte, eksempelvis «pH». Store forbokstaver kan også vurderes dersom dette brukes konsekvent, da dataprogrammer ikke lenger skiller mellom små og store bokstaver (s. 26). Forkortelser eller akronym bør kun benyttes unntaksvis, og bare dersom de er så veletablerte at det fulle navnet sjeldent er i bruk (s. 34-35). Som eksempler finnes det i forskernes nøkkelord forkortelsen «IS» (information systems) «IT» (informasjonsteknologi) og «ERP» (enterprise resource planning) ([vedlegg 2](#)). «IT» vurderer jeg som et veletablert begrep. Uten å være utdannet programmerer, går jeg ut fra at også IS er et godt innarbeidet begrep i fagmiljøet. Men siden nettsidene ellers bruker det fulle navnet «information systems»¹⁵, ville jeg anbefalt å bruke dette som foretrukket term, så kan man heller bruke forkortelsen i den løpende teksten. Det samme gjelder ERP.

Videre anbefaler NISO at man bør unngå neologismer¹⁶, slang og lignende, bortsett fra i tilfeller der det ikke finnes andre brukbare alternativer (International Organization for Standardization, 2011, s. 33). Det tidligere nevnte begrepet «inbetweenness» finnes ikke de engelskspråklige ordbøkene på Ordnett.no, men andre treff i web-søk viser at det er i bruk, også i formen «in betweenness». NISO anbefaler at «Terms should reflect the usage of people familiar with the domain of the controlled vocabulary as reflected in literary, organizational, and user warrant» (2005 (R2010), s. 30). Forskeren har en PhD i engelsk, og dersom denne termen er i bruk i faglitteraturen, vil jeg vurdere det som lurt å inkludere den. For å både kunne gi presise søkeinn ganger for brukere som ønsker en spesifikk ressurs, men samtidig tilrettelegge for brukere som ønsker å gresse, anbefaler Marshall å indeksere med både brede og smale termer. Man vil da ta bedre hensyn til de ulike tilnæringsmetodene og behovene til de ulike sluttbrukerne (Marshall, 2003, s. 85).

¹⁵ <https://www.westerdals.no/studieprogram/master-of-information-systems-spesialisering-digital-business-systems/>

¹⁶ Ordnett.no: språklig nydannelse; nyord

Parenteser, bindestrek, apostrofer og tall kan gjøre søk vanskelige og bør unngås, spesielt i foretrukne termer (International Organization for Standardization, 2011, s. 26).

3.3 Forskning og publikasjoner på ulike språk

Dersom man ønsker å utforme noen retningslinjer for valg av nøkkelord, bør valg av språk være ett av de viktigste. Det danner grunnlaget for andre vurderinger, for eksempel definering av målgrupper; om man skal rette seg mest mot egne studenter og ansatte, andre forskningsmiljøer og medier i Norge, eller sikte internasjonalt.

Nøkkelordene fra WOACT gjenspeiler at både forelesere og studenter har et annet morsmål enn norsk. Særlig viser det seg i nøkkelordene fra forskningsgruppen *Information Systems*, der alle er på engelsk ([vedlegg 2](#)). Jeg vil tro hovedårsaken er at masterutdanningen på avdeling for teknologi foregår på engelsk. I tillegg er majoriteten av all litteratur innenfor IT på engelsk, inkludert språket i forskernes publikasjoner.

På avdeling for kommunikasjon er det blant andre forskere fra Sverige og Tyskland, og enkelte av nøkkelordene gjenspeiler dette, for eksempel «kulturanalys». Mange av de engelske nøkkelordene som like gjerne kunne vært på norsk, for eksempel «media and ethics», vil jeg tro er valgt fordi forskeren ikke har norsk som morsmål (tysk), har studert i engelskspråklige land, og utgir som hovedregel publikasjonene på engelsk.

På nettsidene til WOACT har man har valget mellom norsk eller engelsk versjon. Dersom det skal bli mulig å søke på forskernes nøkkelord, noe jeg anser som svært sannsynlig siden dette relativt enkelt kan tilrettelegges for på nettsidene¹⁷, bør man ta hensyn til at det potensielt kan søkes på begge språkene. Jeg vurderer det som lite sannsynlig at det vil søkes på andre språk, om man da ikke er godt kjent med publikasjonene til en forsker. Min anbefaling vil være å bruke norske nøkkelord i den norske versjonen, bortsett fra på fagområder som mangler, eller i liten grad benytter seg av et godt norsk fagspråk. Som tidligere nevnt er IT et område med hyppig bruk av engelsk. Samtidig ville jeg oversatt termene til engelsk, for den engelske nettversjonen.

¹⁷ ved å utnytte mulighetene i publiseringssystemet (i dette tilfellet WordPress), eller i nettleserens markeringsspråk *html*, ved å bruke attributtverdiene for språk

3.4 Flerspråklige verktøy

Flerspråklig tilgang til informasjon er et eget fagområde under feltet informasjonsvitenskap, og omtales blant annet i retningslinjer for utarbeidelse av tesauruser og interoperabilitet med andre vokabular, utgitt av ISO (International Organization for Standardization, 2013).

Interoperabilitet er et begrep som brukes om systemer som kan samhandle med hverandre. I arbeid med termer betyr dette at vokabularer som i utgangspunktet var ment å stå alene kan relateres til hverandre, slik at søk i den ene, også kan gi treff fra et annet vokabular med tilsvarende begrep. Dette betyr også at ressurser som er beskrevet med ett vokabular, også kan beskrives med termer fra et annet (National Information Standards Organization, 2005 (R2010), s. 82). Om en bruker for eksempel søker etter «branding», kan dette mappes til «merkevarebygging».

Videre følger et knippe prosjekter som har forsket på flerspråklig tilgang. Jeg har inkludert avsluttede eller nedlagte prosjekter, da jeg har regnet selve konseptet som viktigere enn resultatet.

*DBpedia*¹⁸ trekker ut strukturert informasjon fra *Wikipedia*¹⁹, og har over 4,58 millioner navn på entiteter tilgjengelig på 125 språk (DBpedia, 2017b). Lenking til denne ontologien kan gjøre oversettelsesarbeidet lettere, berike vokabularet, og dermed øke inngangene til emnene. Ulempen er at de enorme datamengdene krever store mengder maskinkraft. (Mer om ontologier og DBpedia i kapittel [3.7.2](#) og [3.7.3](#))

The CLEF Initiative (Conference and Labs of the Evaluation Forum) har som formål å fremme forskning og stimulere til utvikling av flerspråklige og multimodale IR-systemer for europeiske og ikke-europeiske språk. CLEF etterfulgte *INEX (INitiative for the Evaluation of XML Retrieval)*²⁰ i 2014, og det har siden 2002 blitt avholdt årlige konferanser der ulike systemer blir presentert og testet ut (The CLEF Initiative, udatert).

CACAO (Cross-language Access to Catalogues and Online Libraries), et toårig program støttet av EU, var et prosjekt der man ønsket å lage et verktøy som kunne integreres i

¹⁸ <http://wiki.dbpedia.org/>

¹⁹ <https://www.wikipedia.org/>

²⁰ <http://inex.mmci.uni-saarland.de/>

bibliotekskatalogen, der brukerne skulle kunne søke på sitt eget morsmål og finne dokumenter på et hvilket som helst språk. Dette skulle gjennomføres ved koble sammen teknologi innen naturlig språkbehandling (NLP) med eksisterende informasjonssøkesystemer (Göttingen State and University Library, 2017). Det ferdige resultatet skulle integreres i *The European Library*²¹, senere *Europeana*²². Jeg har ikke funnet dokumentasjon på om dette ble gjennomført før prosjektet ble avsluttet.

3.5 Søking i fulltekst

I en tid der begrepet «å google» er blitt synonymt med «å søke» og til og med er innlemmet i rettskrivningsordboka²³, er det naturlig å spørre hvorfor man ikke alltid indekserer automatisk og lar brukerne søke i fulltekst. Et hinder kan være at det ikke finnes en felles søkeportal for å søke i alle typer kilder, noe blant annet biblioteksystemet *koha*²⁴ mangler, systemet til WOACTs bibliotek. Tilgang til og søking i tidsskrift må gjøres enkeltvis på de ulike databasenes nettsider, og mange baser krever innlogging på grunn av abonnementsrestriksjoner.

Om målet er å øke gjenfinnbarheten og tilgjengeligheten, bør flest mulig av ressursene være fritt tilgjengelig, såkalt Open Access (OA). Utfordringen er at enkelte publikasjoner kan være utgitt på forlag som ikke tillater OA, eller kun med visse begrensninger, og gjør det dermed vanskelig å koble sammen ressurser. Løsningen kan kanskje være å legge ut *preprints*, det vil si publikasjoner som ennå ikke er utgitt eller fagfellevurdert. I samtale med lederen av fACT, bekreftet han at dette var en løsning særlig han benyttet, også i tilfeller der publikasjonen var utgitt på et abonnementsbasert forlag. Han var klar over at dette kunne være i strid med vilkårene, men mente det var en utfordring å balansere ønsket om fri formidling på den ene siden, og å samle publikasjonspoeng, siteringsstatistikk og så videre på den annen.

Dersom forskningen ikke har slike hindringer, er det likevel ikke sikkert at fulltekstsøk er den beste løsningen. Årsaken til det, er at navnet på konseptet eller begrepet man søker etter ikke alltid er brukt i publikasjonen. Alternativt kan forfatteren ha brukt et begrep utradisjonelt. Sagt på en annen måte: «It is the ideas dealt with by an author, rather than the words used, that must be indexed» (Lancaster, 2003, s. 27). I tillegg viser forskning at det kan være

²¹ <http://www.theeuropeanlibrary.org/>

²² <http://www.europeana.eu/>

²³ Ordbnett.no

²⁴ <https://koha-community.org/>

vanskelig for en bruker å vite hvilke termer som bør brukes i et litteratursøk eller når man skal skaffe seg oversikt over et forskningsområde (Pera et al., 2009, s. 1405). Jeffrey Beall lister opp andre punkter som også kan være en utfordring ved et nøkkelordbasert fulltekstsøk i akademiske bibliotekataloger. Disse er blant annet varianter av synonymer, ulike stavemåter og ordformer, foreldede begreper, forskjeller i ulike fagdisipliner, homonymer, få muligheter til å sortere, figurativt språk, abstrakte emner, ukjent søkeord og ikke-tekstlige ressurser (Beall, 2008).

I slike situasjoner kan et register med termer eller en emneordsliste fungere som et hjelpemiddel til å finne relevante ressurser. I tillegg vil det gjøre brukeren oppmerksom på fagterminologien, og som et resultat kanskje også utføre mer presise søk.

3.6 Automatiske forslag til termer

Andre verktøy som kan bidra til å utvikle et mer konsistent vokabular, er å inkludere en funksjon, for eksempel et elektronisk skjema, som gir forslag til nøkkelord i det man starter å skrive. Forslagene kan komme fra en liste med tidligere brukte nøkkelord, en rettskrivningsordliste, eller fra et kontrollert vokabular, for eksempel en emneordsliste, en tesaurus, eller en ontologi (for eksempel WordNet). Flere av de vitenskapelige bokmerkingsverktøyene på web-en har tilrettelagt for dette, blant andre BibSonomy (udatert). Den klare fordelene med å tilby forslag til termer, er at brukere som begynner å skrive det samme begrepet sannsynligvis vil velge den samme termen (det vil si i samme form), og at man dermed lettere kan gruppere ressurser som omhandler det samme (fordi de blir indeksert med det samme begrepet). Dette vil igjen forbedre søk og gjenfinning for brukerne. Sjansen for skrivefeil vil også reduseres, basert på prinsippet om «kollektiv visdom»: hvis flertallet bruker den samme termen, er det akseptert som et eksisterende og godt stavet ord, og blir en foretrukket term (Van Damme et al., 2007, s. 6). Dersom nøkkelordene har gått gjennom en form for forhåndsgodkjenning eller autoritetskontroll av eksperter, reduseres sjansen for inkonsistens ytterligere. Denne ensrettingen har derimot den ulempen at bredden i nøkkelordene forsvinner, og det kan bli vanskeligere å skjelve ressursene fra hverandre fordi de samme nøkkelordene er brukt, ofte kalt «power tags» (Peters, 2009, s. 363). Sannsynligheten er også stor for at enkelte nøkkelord vil bli knyttet til ressurser som i utgangspunktet ikke ville blitt assosiert med disse (Marlow et al., 2006, s. 34; Mathes, 2004), det som Peters kaller «the self-fulfilling prophecy» [sic] (2009, s. 205), eller «the positive

feedback loop» (Guy & Tonkin, 2006).

Flere forskere har undersøkt om man kan forbedre kvaliteten på tagger eller nøkkelord ved å bruke verktøy som foreslår termer fra et kunnskapsorganisasjonssystem (KOS) eller fra et klassifikasjonssystem som DDK (Golub et al., 2014; Golub et al., 2009) under emneindekseringen. Resultatene viste at spesielt forslag fra kontrollerte vokabular veiledet frem til å finne de riktige taggene. Det gav bedre konsistens og ytterligere søkeinn ganger til ressursene utover søking i fulltekst (2014, s. 824-825; 2009 kap. 5).

Matthews et al. gjorde en undersøkelse i 2010, der 10 forskere innen IT og datavitenskap skulle tagge publikasjoner der noen av disse allerede hadde forslag til tagger fra det kontrollerte vokabularet *ACM Computing Classification Scheme*²⁵, mens andre manglet dette og derfor var åpne for tolkning. Formålet var å evaluere effekten på både indeksering og gjenfinning når man kun bruker sosial tagging, sammenlignet med å bruke sosial tagging i kombinasjon med et kunnskapsorganisasjonssystem. Konklusjonen var at informantene mente taggekvaliteten og -hastigheten ble bedre med taggeforslag, enn om man kun brukte egne tagger, og at gjenfinningskonsistensen ble bedre dersom man brukte autoritetskontrollerte termer (s. 460-461). Kvaliteten på ACM-skjemaet ble vurdert som lite tilfredsstillende av informantene, men skjemaet har i ettertid blitt oppdatert (i 2012), og det er derfor grunn til å tro at kvaliteten på skjemaet ville blitt vurdert annerledes i dag.

Et eksempel på et verktøy som kan brukes til å lage en liste med kontrollert vokabular og deretter komme med forslag til termer med grunnlag i dette vokabularet, er Unilexicon²⁶.

Andre muligheter er open-source programvare eller verktøy som API-er (hjelpesprogram), som kan bygges inn i programvare (som nettlesere) for å utvide programmene med flere funksjoner eller tjenester. API-er kan blant annet lastes ned på nettsiden *ProgrammableWeb*²⁷, en ressurs med over 17 000 API-er.

3.7 Forfatterens motiver

Man må anta at de fleste forskere som publiserer ønsker å få resultatene av forskningen sin

²⁵ <http://www.acm.org/about/class/class/2012>

²⁶ <https://unilexicon.com/>

²⁷ <https://www.programmableweb.com/>

lest, enten den er i form av en kronikk²⁸, eller i en mer formell karakter som vitenskapelige artikler, bokkapitler og lignende. Den økende akademiske konkurransen om midler og poeng fører nødvendigvis også til at man må tiltrekke seg oppmerksomhet til egen forskning.

Når nøkkelordene til forskerne ved WOACT etter hvert legges ut, vil emneoversikten bli synlig, med lenker til de fagansattes sider. Vil dette -bevisst eller ubevisst- kunne påvirke forskernes valg av nøkkelord? Resultater hos Marlow et al., peker mot at det er en forbindelse mellom sosial tilhørighet, sosiolingvistikken og taggevokabularet (s. 38), og en stor del av den faglige omgangskretsen er kollegaer og forskere innenfor samme fagområde. Om man ser hyppig brukte nøkkelord som dermed kanskje er et populært emne, vil det være fristende å velge disse fremfor andre, mindre brukte nøkkelord? En ordsky eller andre format som gjenspeiler termfrekvens vil fremheve de mest populære taggene. En forfatter kan da bli motivert til å bruke disse taggene, eller i ytterste konsekvens til og med lage såkalte «spam tagger» (Marlow et al., 2006, s. 35-36), også kalt «spagging» (Arch, 2007, s. 81), for å bli mer synlig og dermed kanskje tiltrekke seg flere lesere. På den annen side kan man tenke seg at det å skille seg ut fra mengden, det vil si å velge utradisjonelle eller oppsiktsvekkende tagger, gjør at man tiltrekker seg interesse og oppmerksomhet. Begrepet spam tagging brukes for det meste om «adware» og «spyware» som vil tiltrekke seg brukere, og er en metode jeg vil anta er mer tilstede i folksonomier og ikke-vitenskapelige taggeområder. En autoritetskontroll vil også kunne luke ut slike «lokke>tagger».

3.8 XML/RDF og Linked Data

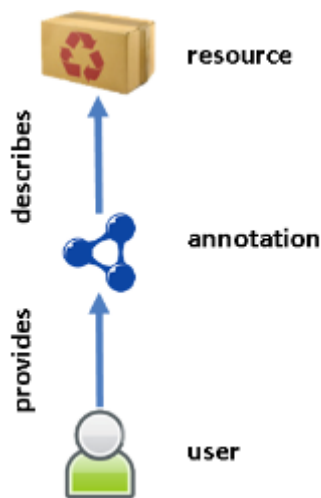
The holy grail of computing, solving the problem of universal data interchange between dissimilar systems

(Dr. Charles Goldfarb, gjengitt etter Papaleo, 2014, s. 109).

Siden utviklingen av den brukermedvirkende og interaktive «Web 2.0», har mye av forskningen vært rundt utviklingen av «Web 3.0», også kalt «semantisk web» eller «Linked Data» (World Wide Web Consortium, 2009b). Begrepet lenket data refererer kort oppsummert til hvordan den beste praksisen rundt publisering og tilrettelegging av strukturert data bør være, for å kunne koble sammen relatert data på web-en (Bizer et al., 2009, s. 205).

²⁸ Eksempel: http://www.aftenposten.no/meninger/kronikk/Humaniora-Krav_-frihet-og-egenart---Camilla-Serck-Hanssen_-Terje-Lohndal-og-Gunn-Elisabeth-Birkelund-619776b.html

Formatene Extensible Markup Language (XML) og Resource Description Framework (RDF), spesifisert av The World Wide Web Consortium (W3C), er to av de mest brukte standardene i dag for å gjøre informasjon maskinleselig. Dette gjør det mulig å utveksle store mengder informasjon på internett, og dermed bidra til utviklingen av den semantiske web-en (World Wide Web Consortium, 2009b). En RDF-trippel består av et subjekt, et predikat og et objekt, og er en modell for å uttrykke metadata. Informasjonen kan for eksempel være «Eva har skrevet artikkel A», der «Eva» er subjektet, for eksempel hentet fra ansatte-sidene i en institusjon, «har skrevet» er predikatet og «artikkel A» er objektet, hentet fra institusjonens arkiv. Figur 1 med trippelen *bruker, tagger og ressurs* illustrerer hvordan dette er tenkt:



Figur 1: A Generic Annotation Model (Bürger et al., 2010, s. 12)

Triplene må ha unike identifikatorer i form av en Uniform Resource Identifier (URI)²⁹ (Bizer et al., 2009, s. 207). URI-ene kan deretter uttrykkes i syntaksen RDF/XML og på den måten gjøre informasjonen maskinleselig. Denne metoden gjør at ikke bare html-dokumenter på web-en kan lenkes sammen, slik det som oftest gjøres i dag ved bruk av url-er (også kalt «Web of Documents»), men at selve innholdet i dokumentene kan utveksles og gjenbrukes, den såkalte «Web of Data». En trippel kan uttrykkes slik:

Subject: <http://data.linkedmdb.org/resource/film/77>
 Predicate: <http://www.w3.org/2002/07/owl#sameAs>
 Object: http://dbpedia.org/resource/Pulp_Fiction_%28film%29

Figur 2: Eksempel på trippel lånt fra (Bizer et al., 2009, s. 208).

²⁹ Subjekt og objekt kan også ha en blank node, det vil si en ressurs uten en URI, også kalt en anonym ressurs. Jeg går ikke dypere inn i dette da jeg vurderer det som utenfor denne oppgavens undersøkelseskontekst.

Mange tagger består av 2/3 av en RDF-trippel. Man vet subjektet, det kan for eksempel være en url til en nettside, og objektet, for eksempel et bilde eller en annen ressurs, men predikatet som skal koble disse sammen mangler ofte (World Wide Web Consortium, 2009d). I en artikkel fra 2015 ble det publisert en undersøkelse som hadde som formål å finne ut hvilken metode som egnet seg best til å lenke vitenskapelige publikasjoner sammen ved bruk av lenket data. Entitetene som skulle lenkes var i hovedsak forfattere, organisasjoner og nøkkelord. Som mange andre semantiske verktøy, ble de lenket mot de tilsvarende entitetene i DBpedia. Verktøyene de sammenlignet var *LOD Refine*³⁰, *SILK*³¹ og forfatternes egenutviklede prototype *FELinker*. Resultatene viste at det var nødvendig å legge til mer semantisk informasjon, for eksempel i form av relaterte nøkkelord, for at gjenfinningen skulle bli mer nøyaktig (Miao et al., s. 269).

Flere har forsøkt å utvikle verktøy for å utnytte de semantiske mulighetene på web-en. W3C tilbyr en oversikt over slike tjenester, som blant annet inkluderer RDF-konvertere, valideringsverktøy som kan feilsøke nettstedet og verifisere at publiserte data er i samsvar med lenket data-prinsippene, og visualiseringsverktøy (World Wide Web Consortium, 2016).

Man har også oppdaget mulighetene lenket data kan ha innenfor utdanningsområdet. Tanken er at blant annet bibliotekataloger, litteraturlister, opplæringsvideoer eller institusjonelle arkiv kan lenkes sammen på tvers av UH-sektoren, slik at flere får tilgang til forskning og faglitteratur. Blant utgivelser som omhandler dette, er *Linked Education: Interlinking educational Resources and the Web of Data* (Dietze et al., 2012). I dette forskningsprosjektet ble det blant annet utviklet verktøyet *MetaMorphosis+*³², «... a semantic social platform for sharing educational resources across the Web» (MetaMorphosis+, udatert). Det langsiktige målet var å etablere en felles portal på nettstedet *Linked Education*³³ der pedagogiske datasett kunne lenkes sammen (Dietze et al., 2012, s. 371).

Andre prosjekt som benytter seg av lenket data er *Linked Universities*³⁴, en allianse av europeiske universiteter engasjert i å gjøre sine offentlige data tilgjengelig for å kunne gi tilgang til informasjon om deres publikasjoner, kurs, pedagogisk materiale og lignende.

³⁰ <https://github.com/sparkica/LODRefine>

³¹ <http://silkframework.org/>

³² <http://metamorphosis.med.duth.gr/>

³³ <http://linkededucation.org/>

³⁴ <http://linkeduniversities.org/>

Universitetsbiblioteket i Oslo³⁵ er et annet eksempel som har tilgjengeliggjort kontrollerte emnevokabularer som åpne lenkede data (udatert).

Dersom disse offentlige datasettene oppfyller kravene til RDF/XML-formatet, kan man også lage visuelle grensesnitt av lenkene. Et hyppig brukt eksempel er Linked Open Data-skyen³⁶ (LOD-cloud), illustrert i [vedlegg 3](#). Diagrammet inneholder data fra flere store institusjoner som BBC, Thomson Reuters og Library of Congress. Innholdet er svært variert, og representerer blant annet steder, bøker, filmer, musikk, tv- og radioprogrammer, gener, medisiner, statistikk og vitenskapelige publikasjoner (Bizer et al., 2009, s. 209).

Samtidig som teknologien bak lenket data gir store muligheter til å koble kunnskap og informasjon sammen og på den måten legge til rette for ny viten, har den også sine ulemper. For eksempel er ikke data statisk. Innhold endrer seg, flyttes, eller slettes. Nye RDF-lenker må legges til og oppdateres (s. 221). Løsninger der du kan gi beskjed om og laste opp endret eller ny RDF-data, eksisterer, blant annet *Ping the Semantic Web* (Giasson, 2006).

En annen utfordring er hensynet til personvern, og problemene som kan oppstå når man integrerer data fra forskjellige kilder. Dette vil trolig kreve en kombinasjon av tekniske og juridiske virkemidler, sammen med en høyere bevissthet hos brukerne om hvilke data som skal oppgis i hvilken sammenheng (Bizer et al., 2009, s. 222).

3.8.1 Web-baserte semantiske verktøy

De siste årenes fremvekst av interaktive emnekart og taggeskyer antyder at det eksisterer et ønske om å visualisere informasjonssøkingen og utforske webinnholdet i et fleksibelt og intuitivt grensesnitt. Dette kan bidra til å heve blikket, og kanskje se sammenhenger der man før kun så faste hierarkier.

Den kanskje enkleste metoden er å lage en ordsky. Man bør i så fall sørge for at man bruker en ordskygenerator som kan vise flerordstermer, noe som ikke lar seg gjøre i CiteULike, blant annet. Et eksempel på en slik generator er WordArt³⁷, der man med bruk av tildesymbolet (~) kan beholde termene samlet. Resultatet med og uten hensyn til flerordstermer, og med og uten

³⁵ <http://data.ub.uio.no/skosmos/nb/?clang=>

³⁶ <http://lod-cloud.net/>

³⁷ <https://wordart.com/>

stoppord³⁸, kan sees i [vedlegg 4](#). Andre vurderinger er om man ønsker å vekte hyppig brukte emneord slik at disse øker i størrelse, eller om alle skal vektas likt. En ordsky der enkelte ord skiller seg ut i form eller farge er mer spennende å se på, men kan bidra til å favorisere enkelte emner slik at disse blir klikket på oftere.

Vialab³⁹ har utviklet flere verktøy for å visualisere informasjon. Ett av disse er DocuBurst⁴⁰, som analyserer tekstdokumenter for deretter å visualisere innholdet ved å bryte de opp i emner. Det viktigste emnet vil være sentrert, med mer underordnede emner mot kanten, formet som en sirkel. Disse emnene kan deretter vise relaterte emner, og man kan også sammenligne dokumenter. Om man vil, kan den også lenke til WordNet. Eksempel kan sees i [vedlegg 5](#).

Ulempen med slike visuelle modeller kan være at de ikke oppfyller kravene til universell utforming. Slike modeller kan derfor kun være alternativ, for å lage ekstra innganger til emnene. Tradisjonelle innganger må også være til stede, for eksempel ved å inkludere nøkkelordene på den enkelte forskers ansattside. Om ulike forskningsområder skal ha egne nettsider, kan nøkkelordene også være synlig der, over brødteksten, slik at de er synlig uten å måtte scrolle nedover siden.

3.8.2 *Ontologier*⁴¹

En ontologi definerer på en formalisert måte termene som brukes til å beskrive et emneområde i webdokumenter (de Keyser, 2012, s. 134). I tillegg inneholder den definisjoner på klassene, funksjonene og relasjonene mellom termene. Det finnes standarder for hvordan man bygger en ontologi, og den kan inneholde alt fra noen få, til flere tusen termer (World Wide Web Consortium, 2015). Et maskinleselig indekseringsspråk gjør at automatiserte dataprogrammer kan samle inn informasjon og utveksle ontologier (mapping). For at denne utvekslingen av data skal foregå enklest mulig, er det nødvendig med et felles «ontologispråk», dvs. et sett med regler for hvordan man bygger en ontologi. I følge de Keyser er det stor sannsynlighet for at World Wide Web Consortiums versjon Web Ontology

³⁸ Ordnett.no: ord som automatisk utelates fra et datamaskingenerert register.

Gir mindre støy på trefflister og raskere søk. Eksempler på engelske stoppord: and, is, about.. Norske stoppord: og, er.

³⁹ <http://vialab.science.uoit.ca/>

⁴⁰ <http://vialab.science.uoit.ca/portfolio/docuburst>

⁴¹ Innen informasjonsvitenskapen. Har en annen betydning innen filosofien.

Language (OWL) vil stå igjen som en standard, da den er tilrettelagt for den semantiske web-en på grunn av sin kompatibilitet med blant annet XML og RDF (de Keyser, 2012, s. 132). De standardiserte definisjonene på både vokabular og relasjoner gjør at man automatisk kan oversette ontologier til flere språk. Ontologier på weben omfatter alt fra kategorisering av websider, til kategorisering av varer, for eksempel produktkatalogen til Amazon (Noy & McGuinness, 2001, s. 1). Ved å inkludere en ontologi kan man forenkle organisering og browsing, tilrettelegge for mer avansert søk, og generelt gi en mer intelligent tilgang til informasjon og tjenester på internett (McGuinness, udatert). En ontologi vil ofte vises i XML-format, men det finnes også verktøy som kan gi et mer visuelt grensesnitt, for eksempel i form av et emnekart eller en tradisjonell «trestruktur». Det finnes egne editorer for ontologikonstruksjon, for eksempel «Protégé» (Stanford Center for Biomedical Informatics Research, 2015).

Særlig relevant i min studie, er prototypen *OntoFrame*, en kunnskapsdrevet semantisk inngangsport til akademisk forskningsinformasjon (de Keyser, 2012, s. 134). Utviklerne bak *OntoFrame* referer til den amerikanske forskningsstiftelsen *NSF*, som hevder at forskere tilbringer mer enn halvparten av forskningstiden sin på å samle informasjon (Jung et al., 2012). For å bøte på dette behovet, skal *OntoFrame* angivelig «...finne informasjonen som forskeren ønsker, og deretter koble, kombinere og analysere den for å gi så mye verdi tilbake til brukeren som mulig» (2012). Denne protoypen ser ikke ut til å ha blitt utviklet videre, men prinsippet ligger svært nær formålet til *WOACTs* ønske om å bedre formidlingen av tverrfaglig forskning (se [vedlegg 6](#)).

3.8.3 DBpedia

DBpedia er én av de største ontologiene, og er et brukersamarbeid der innsatsen er rettet mot å trekke ut strukturert informasjon fra Wikipedia og gjøre denne informasjonen tilgjengelig på nettet i form av RDF-tripler. På denne måten kan man koble sammen ulike datasett på nettet til Wikipedia-data (2012). Utviklerne har satt seg som mål å bøte på mange av utfordringene andre ontologier har, som at de ofte kun dekker et spesifikt fagområde, og at kunnskapen er samlet av en liten gruppe mennesker. Den er heller ikke like kostnadskrevende som andre ontologier, i og med at kunnskapen er gratis tilgjengelig og koblet opp mot Wikipedias kunnskapsbase.

I en undersøkelse av tagger i et KOS og om disse kunne forbedre ressursenes emnebeskrivelser, ble konklusjonen at Wikipedia var mye bedre enn for eksempel WordNet, da Wikipedias struktur er basert på en lenking mellom konsepter. I tillegg gjør brukersamarbeidet at vokabularet utvikler seg raskt (DBpedia, 2017a). På den annen side har Wikipedia en dårligere datakvalitet, sammenlignet med WordNet, ved at den har mindre formalisert struktur og inkluderer inkonsistenser.

Som med andre digitale verktøy, har ontologier den ulempen at den krever en god del teknisk kunnskap av de som skal lage og vedlikeholde den, og det forutsetter også at den er brukervennlig. Flere har derfor testet ut ulike visuelle modeller for å forbedre søkegrensesnittet til det lenkede åpne datasettet.

Erfaring viser dessverre at slike nyskapende prosjekter i praksis ofte renner ut i sanden. Det vil være nødvendig å undersøke behov, brukernes rutiner og kjøre brukertester, og ikke minst må det vedlikeholdes på jevnlig basis. Et dårlig fungerende system kan føre til at brukerne blir frustrerte og avskrekket fra å bruke det.

4 Konklusjon

Jeg har i denne oppgaven drøftet hvordan forskning best kan formidles når forskningen er tverrfaglig, publisert ulike steder, er skrevet av ulike forskere og er på ulikt språk. For å besvare problemstillingen min, har jeg tatt utgangspunkt i et emneindekseringsprosjekt ved Westerdals Oslo ACT, og drøftet dette i lys av relevant forskning innen fagområdet indeksering, kontrollerte vokabular og semantiske verktøy. Jeg har også sett på ulike alternativer til flerspråklige og tekniske løsninger.

Det har vært skrevet om og testet ut mange ideer og verktøy for å knytte sammen kunnskap og metadata, men få, om noen har så langt klart å lage en varig og fungerende løsning som både tar hensyn til at emnebeskrivelsen skal være lavterskel, og samtidig tar hensyn til det som oppfattes som god indekseringsskikk. Dette kan tyde på at mer kontrollerte emnebeskrivelser har en verdi. Men en formell bibliotekfaglig indeksering er kostbart og tidkrevende. Det kan også være en utfordring å indeksere når den som utfører indekseringen ikke er ekspert på emnet. Samtidig viser undersøkelser at forskeres valg av nøkkelord ofte er så nye begrep at de ennå ikke er blitt etablerte emneord, eller i utbredt bruk. Det er derfor en fare for at forfatterne

velger andre begreper enn brukerne vil søke på. Andre undersøkelser viser at forfattere velger nøkkelord av god kvalitet, men at de ofte er gjentakelser av termer brukt i titlene. En emneoversikt over aktuell forskning bør derfor følge en felles, skriftlig anbefaling om valg av og form på nøkkelord, basert på etablerte indekseringsregler. Flere studier har kommet frem til at det bør opprettes en gruppe med fagekspertter som lager liste over vokabularet, som deretter kontrolleres av indekseringskyndige. Dette vil også være min anbefaling.

Dersom tilstrømmningen av ny forskning er stor, eller det ofte blir behov for nye nøkkelord, kan det vurderes om man skal ta i bruk et skjema. Dette skjemaet kan gi forslag til nøkkelord, basert på det kontrollerte vokabularet. Eventuelt kan andre emneordslister eller tesauruser som Humord, Tekord eller ACM index terms også brukes, som et hjelpemiddel for forskere til å finne standardiserte termer i deres fagområde, noe tidligere studier har vist at forfattere har sett nytteverdien av. Et slikt skjema kan bidra til at termene formaliseres, og dermed samler relaterte emner.

Forskning som er emnebeskrevet med de samme nøkkelordene, vil naturlig nok bli samlet ved et søk med bruk av nøkkelord. Det bør også vurderes om relatert forskning skal kobles sammen ved bruk av ulike eksisterende løsninger for lenket data. Søking i fulltekst i forskningspublikasjoner er vanskelig å gjennomføre om man ikke har et system som tilrettelegger for søking i flere kilder samtidig, slik som Oria. Emneord på ulike språk bør oversettes, slik at forskningen er like tilgjengelig på norske og engelske versjoner av en nettside.

For å skape flere innganger, kan nøkkelordene brukes i ulike sammenhenger. De kan organiseres slik at de lager en formalisert emneoversikt over forskningen, de kan brukes på forskernes ansattssider som en oversikt over hva hver enkelt ansatt forsker på, eller for å emnebeskrive publikasjonene. Ulike visuelle modeller som viser relasjoner mellom nøkkelordene kan gi nye perspektiv på de ulike fagområdene, og dermed gi rom for nye oppdagelser.

En naturlig utvikling kan etter hvert være å koordinere nøkkelord fra forskningspublikasjoner med emneordene i bibliotekskatalogen. Eventuelt kan også opprettelsen av et institusjonelt arkiv vurderes. Westerdals Oslo ACT skal nok en gang fusjoneres, -denne gang med Høgskolen Kristiania. Nye vurderinger må derfor gjøres når det er avklart hvordan

fagstrukturen på avdelingene vil være organisert, og hvordan de nye de nettsidene vil være bygget opp.

4.1 Studiens relevans

Forskernes vitenskapelige publikasjoner kan være svært relevante, både for egne og andres studenter og faglige ansatte, men en bedret forskningsformidling ved høyskolen kan ha flere positive ringvirkninger. Mediene kan vise større interesse og henvende seg oftere til høyskolen, for eksempel i tilfeller der man ønsker synspunkt fra forskere ved aktuelle samfunnsspørsmål. Dette kan øke markedsverdien og vekke interesse hos potensielle søkere, som igjen kan føre til høyere søkertall. Et omdømme som en høyskole med stort fokus på, og med høy kvalitet på forskningen, kan tiltrekke seg flere forskere –i tillegg til studenter, og ikke minst vil det være positivt for forskernes egne karrierer.

5 Muligheter for videre forskning

Integrering med bibliotekskatalogen

WOACT har et godt fungerende bibliotek, men katalogen ikke er mye brukt i litteratursøk. Både studenter og ansatte foretrekker å spørre bibliotekarene, eller lete direkte i hyllene.

Flere bibliotek har tatt i bruk sosiale bokmerkingsverktøy som *LibraryThing*⁴² eller *CiteULike*⁴³ og lenker disse ressursene til egne sider. Erfaring viser derimot at slike systemer er vanskelig å vedlikeholde og skape engasjement rundt, sammenlignet med systemer som er utviklet spesifikt til egne forhold. En alternativ løsning kan være å utforske mulighetene som ligger i kohas taggesky-funksjon, og kombinere dette med nøkkelord fra publikasjonene eller forskningssidene på fag- eller ansattssidene. Dette forutsetter dog at funksjonen der man kan søke i de ulike tidsskriftene fra grensesnittet til koha, blir utviklet. Man bør samtidig tenke på at WOACT ønsker å formidle egen forskning. Et verktøy som lenker ut av nettsidene til andre ressurser vil kanskje ikke være en optimal løsning.

Likevel, -den viktigste oppgaven til et bibliotek er å gi best mulig tilgang til flest mulig brukere. Inntil koha får på plass en lenketjener til tidsskriftabonnementene, kunne et

⁴² <https://www.librarything.com/>

⁴³ <http://www.citeulike.org/>

institusjonelt arkiv gi en samlet oversikt over publikasjonene ved WOACT, og gi enda en inngang for søk og gjenfinning.

I det videre arbeidet med nøkkelordene, kunne det også vært nyttig å undersøke om emneinngangene økte interessen for forskningen, for eksempel ved å bruke statistikk over nedlastninger og sidebesøk på nettsidene.

In conclusion, the integration of expert and author generated descriptive metadata can advance and improve the quality of metadata for web content, which in turn could provide useful data for intelligent web agents, ultimately supporting the development of the Semantic Web.

(Robertson & Greenberg, 2002, s. 51)

6 Litteraturliste

- Almo, M. (2011). *Humord: Indekseringsregler*. Hentet 30. mars 2017 fra http://bibsys.no/files/pdf/handbok/humord_indekseringsregler_handbok.pdf
- Arch, X. (2007). Creating the academic library folksonomy: Put social tagging to work at your institution. *College & research libraries news*, 68(2), 80-81.
- Bailin, S. C. & Truskowski, W. (2002). Ontology negotiation between intelligent information agents. *The Knowledge Engineering Review*, 17(1), 7-19. doi:10.1017/S0269888902000292
- Beall, J. (2008). The Weaknesses of Full-Text Searching. *The Journal of Academic Librarianship*, 34(5), 438-444. doi:<https://doi.org/10.1016/j.acalib.2008.06.007>
- BibSonomy. (udatert). *Use the tag system*. Hentet 2. mai 2017 fra https://www.bibsonomy.org/help_en/Tags
- Bizer, C., Heath, T. & Berners-Lee, T. (2009). Linked data-the story so far. I A. Sheth (Red.), *Semantic services, interoperability and web applications: emerging concepts* (s. 205-227). USA: IGI Global.
- Broughton, V. (2004). *Essential classification*. London: Facet.
- Broughton, V. (2006). *Essential thesaurus construction*. London: Facet Publishing.
- Buckland, M. K. (1999). *Vocabulary as a central concept in library and information science*. Hentet 18. april 2017 fra <http://people.ischool.berkeley.edu/~buckland/colisvoc.htm>
- Bürger, T., Zaihrayeu, I., Andrews, P., Babenko, D., Pane, J. & Popov, B. (2010). *Report on the state-of-the-art and requirements for annotation representation models* (DISI-10-007). Hentet fra <http://eprints.biblio.unitn.it/1808/>
- DBpedia. (2017a). *About*. Hentet 2. mai 2017 fra <http://wiki.dbpedia.org/about>
- DBpedia. (2017b). *Facts & Figures*. Hentet 14. mai 2017 fra <http://wiki.dbpedia.org/about/facts-figures>
- de Keyser, P. (2012). *Indexing: from thesauri to the semantic web*. Oxford: Chandos.
- Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N. & Taibi, D. (2012). Linked education: Interlinking educational resources and the Web of data. I *Proceedings of the 27th Annual ACM Symposium on Applied Computing, Trento, Italy* (s. 366-371). New York: ACM.
- Giasson, F. (2006). *Ping the semantic web.com: share your RDF data with the world!* Hentet 2. mai 2017 fra <http://pingthesemanticweb.com/>
- Gil- Leiva, I. & Alonso- Arroyo, A. (2007). Keywords given by authors of scientific articles in database descriptors. *Journal of the American Society for Information Science and Technology*, 58(8), 1175-1187. doi:10.1002/asi.20595

- Glushko, R. J. (Red.). (2013). *The Discipline of organizing*. Cambridge, Mass: MIT Press.
- Golder, S. A. & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2), 198-208. doi:10.1177/0165551506062337
- Golub, K., Lykke, M. & Tudhope, D. (2014). Enhancing social tagging with automated keywords from the Dewey Decimal Classification. *Journal of Documentation*, 70(5), 801-828. doi:10.1108/JD-05-2013-0056
- Golub, K., Moon, J., Tudhope, D. & Nielsen, M. L. (2009). Enhancing social tagging with a knowledge organization system. I *IFLA 2009 Satellite Meeting "Emerging trends in technology: libraries between Web 2.0, semantic web and search technology"*, August 19-20, 2009, Florence, Italy. Florence: IFLA.
- Göttingen State and University Library. (2017). *CACAO - Project details (Cross-language Access to Catalogues And On-line libraries)*. Hentet 2. mai 2017 fra <https://www.sub.uni-goettingen.de/en/projects-research/project-details/projekt/cacao/>
- Gross, T. & Taylor, A. G. (2005). What have we got to lose? The effect of controlled vocabulary on keyword searching results. *College & Research Libraries*, 66(3), 212-230.
- Gross, T., Taylor, A. G. & Joudrey, D. N. (2015). Still a lot to lose: The role of controlled vocabulary in keyword searching. *Cataloging & Classification Quarterly*, 53(1), 1-39. doi:10.1080/01639374.2014.917447
- Guy, M. & Tonkin, E. (2006). Tidying up tags. *D-lib Magazine*, 12(1), 1082-9873. Hentet fra <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january06/guy/01guy.html>
- Hammond, T., Hannay, T., Lund, B. & Scott, J. (2005). Social bookmarking tools (I): A general review. *D-Lib Magazine*, 11(4). doi:10.1045
- Huang, A. W.-C. & Chuang, T.-R. (2009). Social tagging, online communication, and Peircean semiotics: a conceptual framework. *Journal of Information Science*, 35(3), 340-357. doi:10.1177/0165551508099606
- International Organization for Standardization. (2011). ISO 25964-1:2011 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval (s. 152). Sveits: ISO copyright office.
- International Organization for Standardization. (2013). ISO 25964-2:2013 Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies (s. 99). Sveits: ISO copyright office.
- Jung, H., Lee, S., Kim, P., Lee, M.-K., You, B.-J. & Sung, W.-K. (2012). *Use case: OntoFrame 2008 - A semantic portal service of academic research information*. Hentet 2. mai 2017 fra <https://www.w3.org/2001/sw/sweo/public/UseCases/OntoFrame/>
- Kakali, C. & Papatheodorou, C. (2010). Exploitation of folksonomies in subject analysis. *Library & Information Science Research*, 32(3), 192-202. doi:<https://doi.org/10.1016/j.lisr.2010.04.001>

- Kipp, M. E. I. (2005). Complementary or discrete contexts in online indexing: A comparison of user, creator, and intermediary keywords. *Canadian Journal of Information and Library Science*, 29(4), 419-436.
- Kipp, M. E. I. (2007). Tagging practices on research oriented social bookmarking sites. Paper presentert på *Proceedings of the 35th conference of the Canadian Association for Information Science, Montreal, QC, May 10-12, 2007*, Montreal: University of Alberta. Hentet fra <https://journals.library.ualberta.ca/ojs.cais-acsi.ca/index.php/cais-asci/article/view/223/185>
- Kipp, M. E. I. (2011a). Tagging of biomedical articles on CiteULike: A comparison of user, author and professional indexing. *Knowledge Organization*, 38(3), 245-261.
- Kipp, M. E. I. (2011b). User, author and professional indexing in context: An exploration of tagging practices on CiteULike. *Canadian Journal of Information and Library Science*, 35(1), 17-48.
- Kirchhoff, L. (2010). Applying social network analysis to information retrieval on the world wide web: A case study of academic publication space. *University of St. Gallen, Business Dissertations*, 1-241. Hentet fra <https://login.ezproxy.hioa.no/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=lxh&AN=82753382&site=ehost-live>
- Kolbitsch, J. (2007). *WordFlickr: A solution to the vocabulary problem in social tagging systems*. Paper presentert på: Proceedings of I-MEDIA'07 and I-SEMANTICS'07. Graz, Østerrike. Hentet fra <https://www.econstor.eu/obitstream/10419/44447/1/641113358.pdf> - page=85
- Lancaster, F. W. (2003). *Indexing and abstracting in theory and practice* (3. utg.). London: Facet Publishing.
- Lu, K. & Kipp, M. E. I. (2014). Understanding the retrieval effectiveness of collaborative tags and author keywords in different retrieval environments: An experimental study on medical collections. *Journal of the Association for Information Science and Technology*, 65(3), 483-500. doi:10.1002/asi.22985
- Marlow, C., Naaman, M., Boyd, D. & Davis, M. (2006). HT06, tagging paper, taxonomy, Flickr, academic article, to read. *Proceedings of the seventeenth conference on hypertext and hypermedia, August 22-25, 2006, Odense, Denmark* (s. 31-40). Odense: ACM.
- Marshall, L. (2003). Specific and Generic Subject Headings: Increasing Subject Access to Library Materials. *Cataloging & Classification Quarterly*, 36(2), 59-87. doi:10.1300/J104v36n02_07
- Mathes, A. (2004). *Folksonomies: Cooperative classification and communication through shared metadata*. Hentet 2. mai 2017 fra <http://www.adammathes.com/academic/computer-mediated-communication/folksonomies.html>
- Matthews, B., Jones, C., Puzoń, B., Moon, J., Tudhope, D., Golub, K. & Nielsen, M. L. (2010). An evaluation of enhancing social tagging with a knowledge organization

- system. *Aslib Proceedings: New Information Perspectives*, 62(4/5), 447-465.
doi:10.1108/00012531011074690
- McGuinness, D. L. (udatert). *Ontologies and Online Commerce*. Hentet 12. april 2015 fra <http://www-ksl.stanford.edu/people/dlm/papers/ontologies-and-online-commerce-camera-ready.htm>
- MetaMorphosis+. (udatert). *About MetaMorphosis+*. Hentet 20. april 2017 fra http://metamorphosis.med.duth.gr/mod/content_item/about.php
- Miao, Q., Meng, Y., Fang, L., Nishino, F. & Igata, N. (2015). Link scientific publications using linked data. I *Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), Februar 7-9, 2015, Anaheim, CA, USA* (s. 268-271). Anaheim: IEEE.
- Miksa, F., Qin, J., Greenberg, J., O' Neill, E. T. & Pasch, G. (2002). Subject metadata from the other side: Sponsored by SIG CR. *Proceedings of the American Society for Information Science and Technology* (s. 527-528). Hoboken.
- National Information Standards Organization. (2005 (R2010)). *Z39.19 Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Hentet fra http://www.niso.org/apps/group_public/project/details.php?project_id=46
- Noy, N. F. & McGuinness, D. L. (2001). *Ontology development 101: A guide to creating your first ontology*. Hentet fra <http://www-ksl.stanford.edu/people/dlm/papers/ontology-tutorial-noy-mcguinness.pdf>
- Papaleo, L. (2014). Introduction to XML and its applications. I M.-A. Sicilia (Red.), *Handbook of metadata, semantics and ontologies* (s. 109-139). Hackensack, N.J: World Scientific.
- Pera, M. S., Lund, W. & Ng, Y.-K. (2009). A sophisticated library search strategy using folksonomies and similarity matching. *Journal of the American Society for Information Science and Technology*, 60(7), 1392-1406. . doi:doi:10.1002/asi.21072
- Peters, I. (2009). *Folksonomies: Indexing and retrieval in Web 2.0*. Berlin: De Gruyter.
- Princeton University. (2015). *WordNet: A lexical database for English*. Hentet 26. april 2017 fra <http://wordnet.princeton.edu>
- Robertson, W. D. & Greenberg, J. (2002). Semantic web construction: An inquiry of authors' views on collaborative metadata generation. I *Proceedings of the International conference on Dublin Core and metadata for e-Communities, October 13-17, 2002, Florence, Italy* (s. 45-52). Florence: Firenze University Press.
- Rosenfeld, L., Morville, P. & Arango, J. (2015). *Information architecture: For the web and beyond* (4. utg.). Sebastopol, Calif: O'Reilly Media.
- Smith, G. (2008). *Tagging : people-powered metadata for the social web*. Berkeley, Calif: New Riders.

- Spiteri, L., F. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. *Webology*, 4(2), 41. Hentet fra <http://www.webology.org/2007/v4n2/a41.html>
- Stanford Center for Biomedical Informatics Research. (2015). *Protégé*. Hentet 17. april 2015 fra <http://protege.stanford.edu/>
- Suonuuti, H. (2015). *Termlosen: Kort innføring i begrepsanalyse og terminologiarbeid*. Oslo: Språkrådet.
- The CLEF Initiative. (udatert). *CLEF Initiative*. Hentet 2. mai 2017 fra <http://www.clef-initiative.eu/>
- Thomson Reuters. (2015). *Web of Science Core Collection Help: Keywords*. Hentet 2. mars 2017 fra http://images.webofknowledge.com/WOKRS523R4/help/WOS/hp_full_record.html-dsy1028-TRS_keywords_plus
- Universitetsbiblioteket i Oslo. (2016). *Realfagstermer*. Hentet 16. mai 2017 fra <https://www.ub.uio.no/bruk/sok-i-hele-biblioteket/beskrivelser-av-termer/real-fagstermer.html>
- Universitetsbiblioteket i Oslo. (udatert). *Informasjon*. Hentet 16. mai 2017 fra <http://data.ub.uio.no/skosmos/nb/about>
- Van Damme, C., Hepp, M. & Siorpaes, K. (2007). Folksonology: An integrated approach for turning folksonomies into ontologies. I *International Workshop on Bridging the Gap between Semantic Web and Web 2.0 (SemNet 2007), at the 4th European Semantic Web Conference, Innsbruck, Austria* (s. 1-14). Innsbruck: Citeseer.
- Vander Wal, T. (2007). *Folksonomy*. Hentet 10. mars 2017 fra <http://vanderwal.net/folksonomy.html>
- Voorbij, H. J. (1998). Title keywords and subject descriptors: a comparison of subject search entries of books in the humanities and social sciences. *Journal of Documentation*, 54(4), 466-476. doi:doi:10.1108/EUM0000000007178
- Voß, J. (2007). Tagging, Folksonomy & Co - Renaissance of Manual Indexing? Hentet fra <https://arxiv.org/abs/cs/0701072v2>
- Westerdals Oslo ACT. (udatert). *Hei!* Hentet 8. november 2016 fra <https://www.westerdals.no/artikkel/hei/>
- World Wide Web Consortium. (2009a). *W3C Semantic Web Frequently Asked Questions*. Hentet 18. april 2017 fra <https://www.w3.org/2001/sw/SW-FAQ>
- World Wide Web Consortium. (2009b). *W3C Semantic Web Frequently Asked Questions*. Hentet 18. april 2017 fra <https://www.w3.org/2001/sw/SW-FAQ>
- World Wide Web Consortium. (2009c). *W3C Semantic Web Frequently Asked Questions: 2.6. ... tagging, folksonomies*. Hentet 18. april 2017 fra <https://www.w3.org/2001/sw/SW-FAQ-rehtag>

World Wide Web Consortium. (2009d). *W3C Semantic Web Frequently Asked Questions: 2.6. ... tagging, folksonomies*. Hentet 18. april 2017 fra [https://www.w3.org/2001/sw/SW-FAQ - reltag](https://www.w3.org/2001/sw/SW-FAQ-rehtag)

World Wide Web Consortium. (2016). *Semantic Web Development Tools*. Hentet 2. mai 2017 fra <https://www.w3.org/2001/sw/wiki/Tools>

6.1 Personlig kommunikasjon

Anders Olof Larsson, førsteamanuensis, Westerdals Oslo ACT (samtale, 28. mai 2017)

Vedlegg:

Vedlegg 1: Nøkkelord fra forskergruppen fACT

Vedlegg 2: Nøkkelord fra forskningsgruppen Information Systems

Vedlegg 3: LOD-sky

Vedlegg 4: Ordsky

Vedlegg 5: DocuBurst

Vedlegg 6: OntoFrame

Vedlegg 7: Samtykkeskjema

7 Vedlegg

7.1 Vedlegg 1 Nøkkelord fra forskergruppen FACT

Forsker A, førsteamanuensis, avdeling for ledelse - Strategi og PR:

Journalistikk, Politisk Kommunikasjon, Sosiale medier, Strategisk kommunikasjon

Forsker B, programansvarlig og lektor, avdeling for ledelse - Prosjektledelse kunst og kreativ næring:

Kreative næringer, Design Thinking, Strategi, Kreativitet, Marketing, Reklame

Forsker C, førsteamanuensis, avdeling for kommunikasjon - Grafisk design:

visuell kultur, fashion, medieestetikk, digitale medier, merkevarebygging

Forsker D, professor, avdeling for film, tv og spill - Spilldesign:

Dataspill, mediebrukere, digitale medier, mediedebatter, onlinekultur

Forsker E, programansvarlig og førsteamanuensis, avdeling for kommunikasjon - Tekst og skribent:

Skrivekunst, Flerspråklighet, Inbetweenness, Pedagogikk, Kreativitet

Forsker F, førsteamanuensis, avdeling for kommunikasjon - Digital markedsføring:

Psykologi, multisensorisk persepsjon, temporal persepsjon, multimedia, menneske-maskin interaksjon

Forsker G, programansvarlig, avdeling for kommunikasjon - Tidligere studieprogram: Experience and Event Design:

Kulturanalys, event, (musikk)festivaler, ungdomskultur, populærmusikk

Forsker H, førsteamanuensis, avdeling for ledelse - Prosjektledelse kunst og kreativ næring:

media and television studies, humour and comedy studies, media and ethics, media and nostalgia, arts and society

Forsker I, professor II, avdeling for ledelse - Prosjektledelse kunst og kreativ næring:

media, mediepolitikk, innovasjon, ledelse, beslutninger

Forsker J, førsteamanuensis, avdeling for ledelse - Strategi og PR:

medier, motoffentligheter, strategisk kommunikasjon, kjønn og seksualitet, sosiologi

Forsker K, programansvarlig og førstelektor, avdeling for kommunikasjon - Grafisk design:

grafisk design, spekulativt kritisk design, bærekraftig design

7.2 Vedlegg 2 Nøkkelord fra forskningsgruppen Information Systems

Fagdidaktikk informatikk

Digital innovation

Large-scale IT-projects

Digital platforms and ecosystems

IS and infrastructures

Social issues of IT

Agile software development

Human-Computer Interaction

Universell utforming av IT

Social media

Health information management

Social cities

Big data

ERP systems

Digital marketing

Behavioural economics

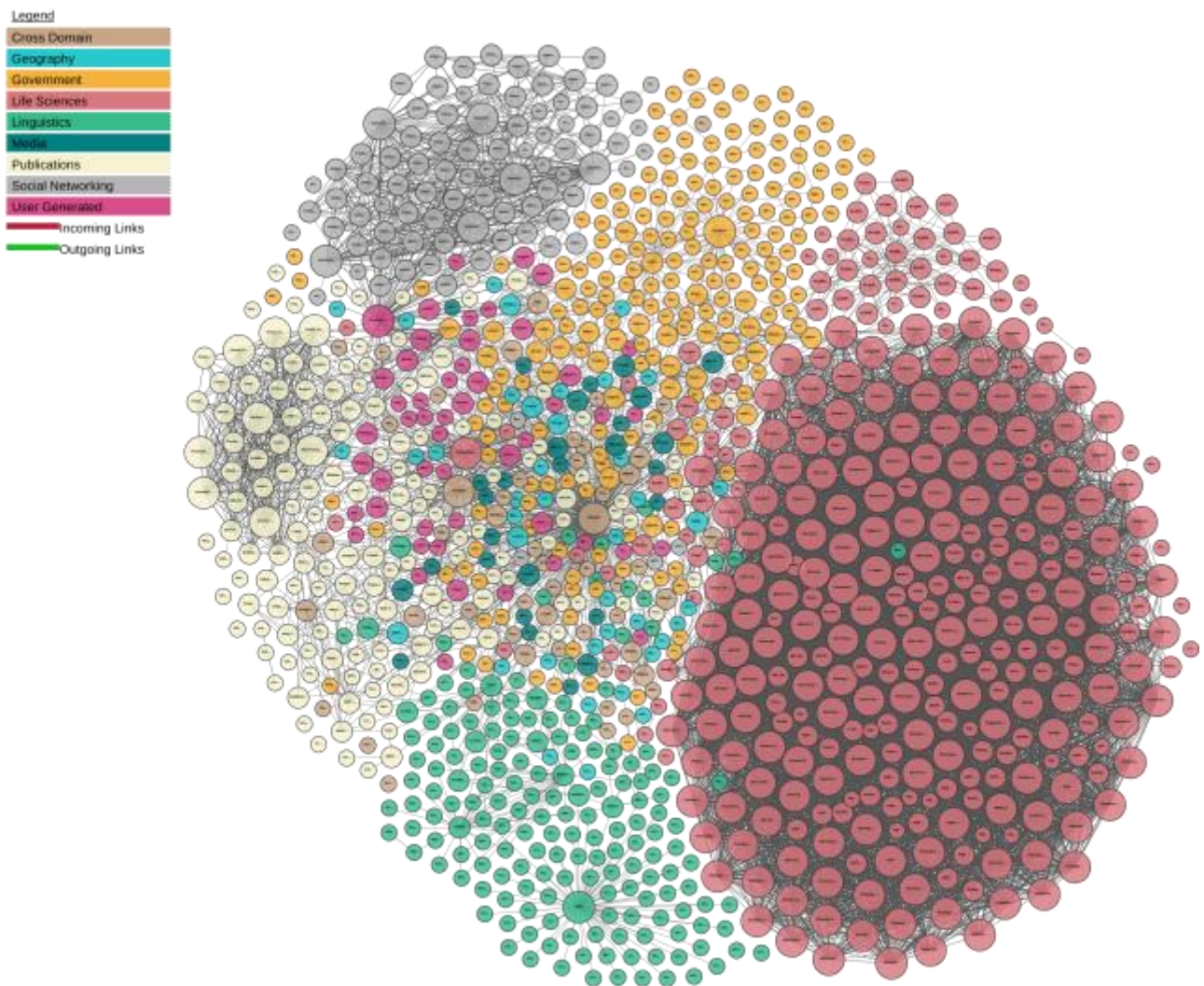
Business intelligence processes

Research methods in IS

eGovernment

Interaction design

7.3 Vedlegg 3 LOD-sky

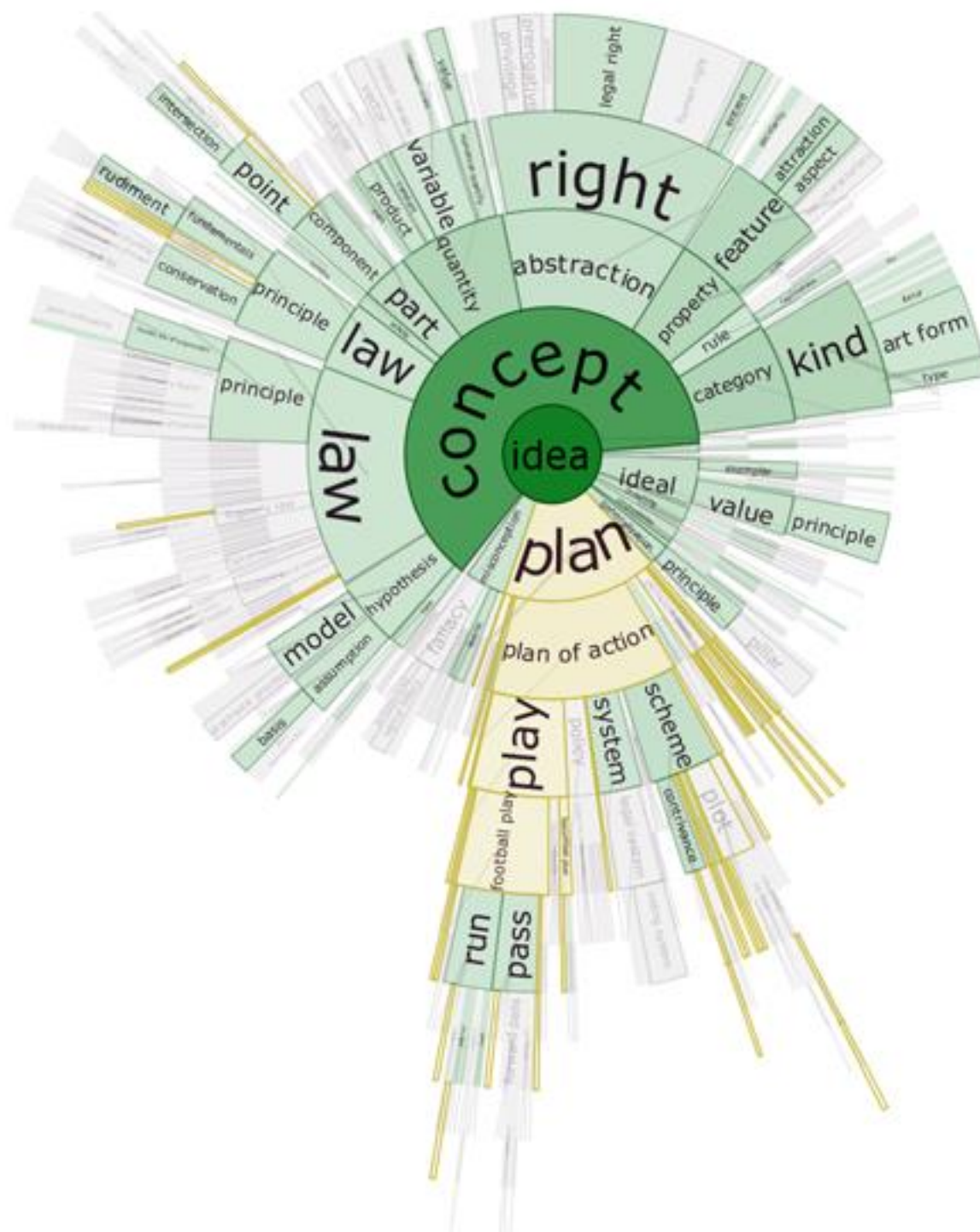


Linking Open Data cloud diagram 2017

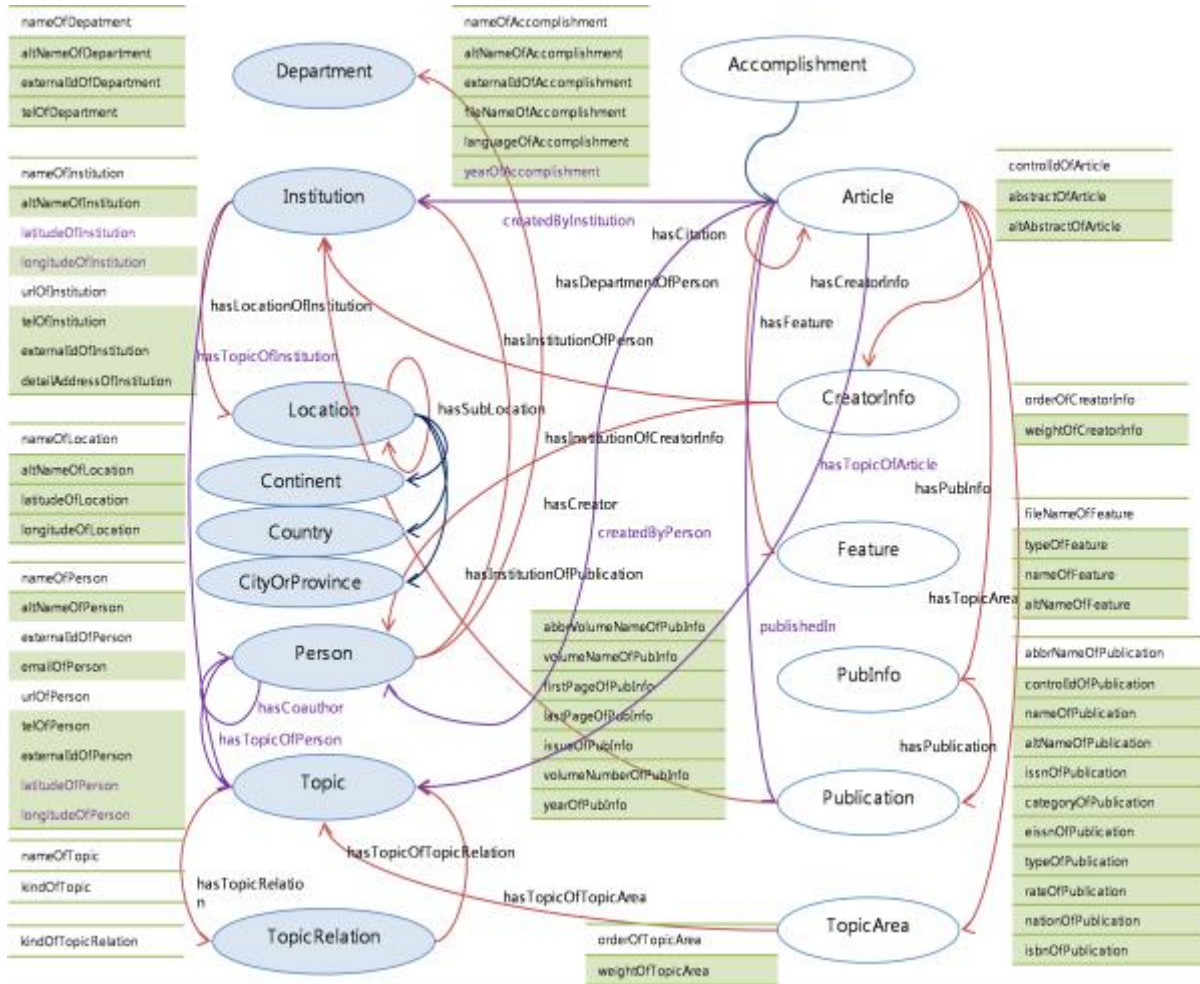
Av Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch og Richard Cyganiak

<http://lod-cloud.net/>

7.5 Vedlegg 5 DocuBurst



7.6 Vedlegg 6 OntoFrame



© Copyright 2009, Korea Institute of Science and Technology Information (KISTI)

7.7 Vedlegg 7 Samtykkeskjema

Forespørsel om deltakelse i forskningsprosjektet

«Emneindeksering og gjenfinning av tverrfaglig forskning -muligheter og utfordringer».

Bakgrunn og formål

Hvordan lenke forskning sammen fra ulike høyskoleavdelinger, og som er på flere ulike språk, slik at forskningen synliggjøres bedre og blir lettere tilgjengelig for studenter og andre interesserte?

Prosjektet er en bacheloroppgave ved Høyskolen i Oslo og Akershus, Fakultet for samfunnsvitenskap, Institutt for arkiv-, bibliotek- og informasjonsfag, fordypning i Kunnskapsorganisasjon og gjenfinning. Oppgaven er mitt eget initiativ, men med bakgrunn i et reelt prosjekt ved Westerdals Oslo ACT der formålet er å formidle forskningen ved institusjonen bedre. Oppgaven vil være utformet som en litteraturbasert studie.

Et fåtall ansatte ved WOACT er valgt ut til å svare på spørsmål som måtte oppstå underveis i oppgaven. De ansatte er valgt ut fra deres posisjon som initiativtager til formidlingsprosjektet, som medlemmer av forskningsgruppen fACT, og som bibliotekar.

Hva innebærer deltakelse i studien?

Deltagelse i studien består av å dele tags/stikkord som skal brukes til å emnebeskrive forskning, og å svare på spørsmål, hovedsakelig via e-post. Spørsmålene vil for det meste omhandle hvor og hvordan man har tenkt å publisere forskningen.

Hva skjer med informasjonen om deg?

Alle personopplysninger lagres på en privat PC eller skytjeneste, begge med passordbeskyttelse. Evt. bruk av navn vil kun bli delt med veileder og prosjektdeltagerne. Det er likevel mulig at man kan identifiseres i oppgaven, på bakgrunn av kombinasjoner av opplysninger om institusjon, avdeling, stilling og kjønn.

Prosjektet avsluttes 19. mai 2017. Dokumenter i skytjenester vil slettes, andre elektroniske dokumenter (som for eksempel den ferdige bacheloroppgaven) vil lagres på ekstern disk og oppbevares i privat bolig med boligalarm. E-post kan om deltagerne ønsker det, slettes.

Frivillig deltakelse

Det er frivillig å delta i studien, og du kan når som helst trekke ditt samtykke uten å oppgi noen grunn. Dersom du trekker deg, vil alle opplysninger om deg bli anonymisert.

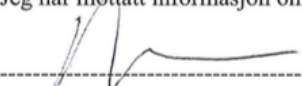
Dersom du har spørsmål til studien, ta kontakt med.

Kontaktinformasjon til veileder er Tine Lodberg Frost; tine.frost@hioa.no, telefon (kontor) 67 23 80 39.

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Samtykke til deltakelse i studien

Jeg har mottatt informasjon om studien, og er villig til å delta

 Anders Olof Larsson 15. mai 2017

(Signert av prosjektdeltaker, dato)