

Mapping for sluttbrukertjenester

Sluttrapport



Skrevet av Dan Michael O. Heggø, Karoline Hoff, Unni Knutsen, Heidi Sjursen Konestabo, Viola Kuldvere og Vibeke Stockinger Lundetræ.

Universitetsbiblioteket i Oslo

Oslo, 28. juni 2019

Innhold

Begrepsforklaringer	1
Innledning og historikk	3
Utvikling av metodikk for mapping	3
Mappingaktiviteter	4
Publisering av data	4
Utvikling av mappingverktøy	5
Prosjektet Mapping for sluttbrukertjenester (2017–2019)	6
Hovedaktivitet 1: Utforske det kunnskapsorganisatoriske potensialet i mapper fra et sluttbrukerperspektiv	7
Litteraturstudien	7
Observasjon av studenter	10
Søkelogganalyse	12
Hovedaktivitet 2: Utnyttelse av mapper i sluttbrukerverktøy	18
Utvidelse av søk fra mapper (Mappingutvidelse)	18
Muligheter for mappingutvidelse i Oria	18
Utvidelse under indeksering	18
Utvidelse under søking	19
Testverktøy	19
Funn fra testverktøyet	23
Implementering av mappingutvidelse i Scriptotekets Emnesøk	25
Videre arbeid	26
Hovedaktivitet 3: Ferdigstille mappingen	28
Felles for begge vokabularene	28
Humor	31
Realfagstermer	33
Relevans for norsk WebDewey	34
Personale og økonomi	38
Formidlingsaktiviteter	39
Avslutning og konklusjoner	41
Litteraturliste	43
Vedlegg 1 – Søkeoppgaver til observasjonsundersøkelsen	45
Vedlegg 2 – Økonomirapport	46

Begrepsforklaringer

Deweys desimalklassifisering (DDK/Dewey): Universelt klassifiseringssystem som brukes av bibliotek over hele verden. Eies og utvikles av OCLC. Publisert på nett under navnet WebDewey, som er en løpende oppdatert versjon av 23. utgave. For den fullstendige oversettelsen til norsk bokmål av Nasjonalbiblioteket, se **Norsk WebDewey**.

ccmapper: Web-basert verktøy for opprettelse og vedlikehold av mappinger mot WebDewey, utviklet på oppdrag fra mappingprosjektet.

Discovery-system: Brukes om søkesystemer for bibliotekressurser rettet mot sluttbrukere, som fokuserer på enkelt søk med relevanssortering og fasettavgrensning («start bredt, snevre inn») og som ofte inkluderer materiale utover den tradisjonelle bibliotekskatalogen, f.eks. en sentralindeks for artikkelmateriale.

European DDC Users Group (EDUG): Samarbeid mellom europeiske brukerinstusjoner og OCLC for å fremme utviklingen av Deweys desimalklassifisering.

Humord: Emneordsvokabular som dekker humaniora og samfunnsvitenskap med tilgrensende fagområder. Vokabularet er strukturert som en tesaurus med hierarkiske og assosiative relasjoner, og inneholder rundt 18 600 begreper med termer på bokmål og engelsk, og rundt 9 800 henvisningstermer/synonymer per språk. Utvikles gjennom et samarbeid mellom universitetsbibliotekene i Oslo, Bergen og Tromsø, med UiO HumSam-biblioteket som koordinator og sekretariat. Vokabularet brukes også av Senter for studier av Holocaust og livssynsminoriteter samt Nobelinstituttets bibliotek.

ISO-25964: ISO-standard i to deler som omhandler tesauruskonstruksjon (del 1) og mapping mot andre vokabularer (del 2).

Kildevokabular: Et vokabular som utgjør startpunktet når man søker et korresponderende begrep i et annet vokabular (målvokabularet). I dette prosjektet er Humord og Realfagstermer kildevokabularer.

Mapping: Aktivitet hvor det etableres relasjoner mellom begreper i to ulike kontrollerte vokabularer. Man mapper fra et kildevokabular og finner korresponderende begreper i et målvokabular. Mapping brukes også til å benevne resultatet av en mappingaktivitet, dvs. relasjonen mellom et begrep i ett vokabular og ett eller flere begreper i et annet vokabular.

Mappingrelasjoner: ISO-standarden for mapping (ISO 25964-2) opererer med tre mappingtyper: ekvivalens, hierarkisk og assosiativ, som gir opphav til fem typer mappingrelasjoner: =EQ (eksakt ekvivalens), ~EQ (tilnærmet ekvivalens), BM (overordnet), NM (underordnet) og RM (relatert/assosiativ). De samme relasjonene gjenfinnes i SKOS-standardens skos:exactMatch, skos:closeMatch, skos:broadMatch, skos:narrowMatch og skos:relatedMatch. I dette prosjektet bruker vi ikke relasjonen NM. Dette er nedfelt i *EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)*.

Målvokabular: Et vokabular der man søker etter korresponderende begreper til begreper fra kildevokabularet. I inneværende prosjekt er målvokabularet norsk WebDewey.

Norsk WebDewey: Fullstendig oversettelse av Deweys desimalklassifisering til norsk bokmål. Inneholder rundt 44 000 klasser og blir løpende oppdatert. Inneholder også en betydelig mengde bygde numre som ikke finnes i den engelske utgaven av WebDewey.

Oria: En felles portal til det samlede materialet som finnes ved de fleste norske fag- og forskningsbibliotek, supplert med en mengde elektronisk materiale fra åpne kilder.

Overlappsvokabularet: Sett av termer som finnes både i ett av kildevokabularene (som foretrukket eller alternativ term i Humord/Realfagstermer) og i målvokabularet (som klassebetegnelse eller registerterm i norsk WebDewey).

Primo: Et discovery-system utviklet av Ex Libris i 2007. En sentral tjeneste er Primo Central Index (PCI) som aggregerer og gjør søkbart fulltekstressurser fra kommersielle kanaler og dokumenter som det gis åpen tilgang til. Unit (Direktoratet for IKT og fellestjenester i høyere utdanning og forskning) tilpasser Primo under merkenavnet Oria.

Realfagstermer: Emneordsvokabular som i hovedsak dekker naturvitenskap, matematikk og informatikk. Vokabularet inneholder rundt 11 600 begreper med termer på bokmål, nynorsk og engelsk, assosiative relasjoner, og rundt 5000 henvisningstermer per språk. Utvikles gjennom et samarbeid mellom realfagsbibliotekene ved Universitetet i Oslo og Universitetet i Bergen.

SKOS (Simple Knowledge Organization System): standard for å representere tesauri, klassifikasjonsskjemaer og andre typer kontrollerte vokabularer på den semantiske veven.

Vokabular/kontrollert vokabular: Sett av termer, emneord eller koder som representerer begreper og relasjonene mellom dem, og som kan bli brukt i informasjonsgjenfinning.

Innledning og historikk

I 2014–2015 deltok Universitetsbiblioteket i Oslo (UBO) sammen med Nasjonalbiblioteket (NB) i *Tesaurus forprosjekt*, som skulle gi beslutningsgrunnlag for en eventuell etablering av en norsk generell tesaurus med utgangspunkt i tesaurusen Humord. Arbeidet resulterte i rapporten *En norsk generell tesaurus?*, som de to institusjonene fikk til behandling i 2015.¹

I forbindelse med prosjektet, mottok UBO også midler til å utrede metodikk for mapping mot Dewey. Arbeidet bygde dels videre på metoder som var utviklet i tidligere prosjekter i Realfagsbiblioteket, og resulterte i en egen delrapport *Metodikk for mapping av Humord mot WebDewey* (2015).²

Vesentlig i dette prosjektet var å utforske ISO-25964 del 2 (2013), en standard som har dannet det teoretiske og metodiske fundamentet for alt mappingarbeid i senere prosjekter. I forbindelse med dette valgte vi en overordnet mappingmodell og besluttet hvilke mappingrelasjoner vi ville ta i bruk i vårt mappingarbeid (=EQ, ~EQ, BM, RM, se *Begrepsforklaringer* for forklaring). Mappingrelasjonene er kompatible med SKOS/RDF (exactMatch, closeMatch, broadMatch, relatedMatch) og derfor godt egnet for publisering på den semantiske weben, noe som hele tiden har vært et mål.

For videre beskrivelse av modellvalg og metodikk, vises det til prosjektrapporten.²

I forlengelse av metodikkprosjektet fikk UBO i 2015 prosjektmidler fra NB til prosjektet *Mapping mot norsk WebDewey*. Prosjektet var et toårig prosjekt med midler for årene 2015 og 2016. Siden det var midler igjen ved årsskiftet 2016/2017, ble prosjektet forlenget inn i 2017. Sluttrapporten³ er datert 6. juni 2017.

Hovedmålet for dette prosjektet var å mappe vokabularene Humord og Realfagstermer mot norsk WebDewey. Prosjektet hadde videre to delmål:

1. Prosjektet skulle bidra til videreutvikling av metodikk for mapping mot Deweys desimalklassifikasjon.
2. Mappingarbeidet skulle utføres som en datastøttet intellektuell prosess og resultere i utvikling av programvare for mapping av vokabularer, inkludert vedlikehold av mappinger i forbindelse med endringer i de involverte vokabularene.

Utvikling av metodikk for mapping

ISO-25964 del 2 beskriver mapping ut fra et generelt og teoretisk perspektiv, men går ikke inn på bestemte vokabularer. Vi så tidlig et behov for å utarbeide praktiske retningslinjer for mapping mot Deweys desimalklassifikasjon, både fordi ISO-standardene i liten grad fokuserer på klassifikasjonsskjemaer og fordi det er en del særegenheter ved Dewey.

¹ <https://perma.cc/J84V-7BSK>

² <https://perma.cc/EMN9-SH7P>

³ <https://perma.cc/Z7YV-JPF7>

Vi jobbet tett med den europeiske Dewey-brukergruppen EDUG fordi det her finnes miljøer som har lang erfaring med både mapping og Dewey. Under EDUGs årlige møte i 2015, arrangerte mappingprosjektet en workshop om implikasjonene ved å mappe emnevokabularer til Deweys desimalklassifisering. Vi anså dette som et nødvendig skritt på veien mot å utvikle en metodikk for mapping i vårt prosjekt. Andre bibliotek har riktignok mappet emnevokabularer til Deweys desimalklassifisering, men ingen andre hadde på dette tidspunkt mappet i tråd med anbefalingene i ISO-25964-2.

Basert på innspillene fra workshopen ble det satt ned en redaksjonsgruppe for å utvikle et sett med internasjonale anbefalinger om mapping til Dewey. Nasjonalbiblioteket (ved Elise Conradi) og Universitetsbiblioteket (ved Grete Seland og Unni Knutsen) deltok i redaksjonsgruppen, sammen med representanter fra Tyskland, Sverige og utviklerne av Dewey (OCLC). *EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)* ble publisert høsten 2015 og gjennomgikk en mindre revisjon i juni 2016 etter EDUGs årlige møte dette året.

En av anbefalingene i nevnte dokument, er å utarbeide et styringsdokument for prosjektet. Dette ble utarbeidet av UBO i etterkant og publisert på prosjektets nettside i en norsk⁴ og en engelsk⁵ utgave.

Mappingaktiviteter

I påvente av at mappinganbefalingene skulle ferdigstilles, og fordi nye problemstillinger stadig dukket opp og måtte avklares, omtalte vi i starten mappingaktivitetene våre som testmapping. Testmappingen utgjorde ca. 340 begreper. Disse ble på det tidspunktet vi sendte inn vår midtveisrapport til NB (oktober 2015), omgjort til ferdigmappede begreper basert på anbefalingene fra EDUG.

På samme tid startet vi med å mappe det såkalte overlappsvokabularet (se begrepsforklaring) mellom våre kildevokabularer (Realfagstermer og Humord) og norsk WebDewey. Overlappsvokabularet utgjør ca. 30 % av hvert vokabular. Per ultimo mai 2017, da sluttrapporten fra prosjektet ble sendt til NB, var totalt 16 359 kildebegreper mappet. Dette utgjorde totalt 42 482 mappinger med et snitt på 2,6 mappinger per begrep. Per denne dato var 66 % av Humord-begrepene og 29 % av Realfagstermene mappet.

Publisering av data

I løpet av det første prosjektet hadde vi fått etablert en nattlig eksport fra mappingverktøyet vårt til norsk WebDewey, slik at alle ferdigmappede og godkjente begreper ble synlige og søkbare i norsk WebDewey så fort som mulig.

På dette tidspunktet hadde vi også fått på plass fortløpende publisering av kildevokabularene (Humord og Realfagstermer) og mappingene som åpne, lenkede data basert på SKOS/RDF på

⁴ <https://perma.cc/F6Y9-VML7>

⁵ <https://perma.cc/2FKB-PXWV>

<https://data.ub.uio.no/>, slik at mappingene kan brukes av andre utviklere og sees i sammenheng med andre vokabularer.

Utvikling av mappingverktøy

Vi har hele tiden forutsatt at mapping mellom Humord, Realfagstermer og norsk WebDewey vil medføre en kombinasjon av maskinelt og intellektuelt arbeid.

I Realfagsbibliotekets prosjekt *Felles terminologi for klassifisering med Dewey* (2014)⁶ ble det utviklet et mappingverktøy kalt μ mapper.⁷ I prosjektet *Mapping mot norsk WebDewey* bygde vi videre på dette arbeidet, med mål om å utvikle en applikasjon som kunne generere bedre forslag og som kunne støtte i håndtering av endringer i vokabularene, både på kilde- og målsiden. Vi kalte dette nye mappingverktøyet ccmapper (concept context mapper)⁸ og utarbeidet skisser for brukergrensesnittet.

Under EDUG-møtet i Napoli i 2015 ble vi kontaktet av Pansoft, firmaet som utvikler WebDewey, for å diskutere om vi kunne samarbeide om å utvikle et mappingverktøy. Pansofts forslag brakte det innenfor rekkevidde å få produsert et komplett mappingverktøy, og ikke kun en prototype som vi hadde sett for oss og fått innvilget midler til. Etter samtaler med NB, fikk vi tillatelse til å inngå samarbeid med Pansoft om å utvikle et mappingverktøy basert på metodikk og interaksjonsdesign fra mappingprosjektet.

UBO og Pansoft inngikk en kontrakt ved årsskiftet 2015/2016. Vi jobbet tett med Pansoft med regelmessige videokonferanser og utstrakt testing av funksjonalitet. Dette var en ressurskrevende arbeidsmetode både for Pansoft og prosjektet, men resulterte i et sluttprodukt som er blitt meget bra og dekker våre behov godt. Vi fikk dessuten opparbeidet oss ny IKT-kompetanse ved å delta på disse betingelsene. Mappingprosjektet har fått full tilgang til kildekode til verktøyet og har stått for oversettelsen av verktøyet til norsk bokmål.

I prosjektet *Mapping mot norsk WebDewey* brukte vi μ mapper i den første fasen, mens ccmapper ble tatt i bruk fra januar 2017. Endringshåndteringsmodulen var da fortsatt ikke på plass, men denne ble tatt i bruk fra medio mai 2017.

ccmapper har gitt oss verdifull støtte til et tidvis krevende intellektuelt arbeid. Nå som mappingen er ferdig og vi er over i en driftsfase, er det helt avgjørende å ha endringshåndteringsstøtte som gir oss oversikt over endringer i Dewey (både nye numre og andre slags endringer) som kan påvirke mappingene som allerede er gjort, men også endringer i kildevokabularene (Realfagstermer og Humord) som kan føre til endringer i enkeltmappingen.

Verktøyet har også vist seg å være interessant for flere aktører. Verbundzentrale des GBV i Göttingen har også tatt i bruk ccmapper, og OCLC i USA har vist interesse.

⁶ <http://urn.nb.no/URN:NBN:no-44610>

⁷ <https://lambda.biblionaut.net/>

⁸ <http://ccmapper.no.pansoft.de/> (krever innlogging)

Prosjektet Mapping for sluttbrukertjenester (2017–2019)

I februar 2017 ble det sendt en ny søknad til Nasjonalbiblioteket. Dette prosjektet fikk navnet *Mapping for sluttbrukertjenester*. Som tittelen antyder, var hovedformålet for prosjektet å utnytte mappingene ved emnesøk i sluttbrukerverktøy. Prosjektets intensjon var å bidra til at sluttbrukere kan få hjelp til å utforske emnebaserte informasjonsbehov på en intuitiv måte. Gjennom utprøvinger og prototyping har vi ønsket å sikre at sluttbrukere får bedre og flere søkeinn ganger til emnedata. Videre beskrev vi en situasjon hvor mapping til WebDewey gir mulighet for å utvikle flerspråklige emnesøk i sluttbrukersystemene. En forutsetning for at prosjektet skulle lykkes, var at emnevokabularene Humord og Realfagstermer ble mappet ferdig.

Dette prosjektet fikk tre definerte delmål:

1. Utforske det kunnskapsorganisatoriske potensialet i mappinger fra et sluttbrukerperspektiv.
2. Utvikle grensesnitt for sluttbrukerverktøy som utnytter mappinger i emnesøk.
3. Ferdigstille mappingen av vokabularene Humord og Realfagstermer mot norsk WebDewey.

En midtveisrapport ble sendt til NB i februar 2018.⁹

Den foreliggende sluttrapport gir et bilde av arbeidet i prosjektet, våre funn og resultater. Rapportens hoveddel er organisert etter de tre delmålene i prosjektet. Relevante forhold som har beskrevet i tidligere rapporter har blitt trukket inn, men vi henviser til disse for å få et komplett bilde av mappingprosjektet.

⁹ <https://perma.cc/GN3M-4BFL>

Hovedaktivitet 1: Utforske det kunnskapsorganisatoriske potensialet i mapper fra et sluttbrukerperspektiv

Litteraturstudien

Før vi begynte med våre egne forsøk, så vi det som nødvendig å undersøke hva eksisterende forskning viste når det gjelder studenters søkeadferd ved emnebaserte informasjonsbehov. Vi gjennomførte derfor en litteraturstudie for å få et overblikk på den nåværende situasjonen og svare på spørsmålet «hvordan utfører studenter emnesøk i et discovery-system som Oria?»

For å svare på dette søkte vi å identifisere dokumenter som handlet om:

- Hvordan studenter søker og i hvilke kilder.
- Hvordan studenter søker i discovery-baserte systemer.
- Hvordan studenter utfører emnesøk.

Vi ønsket i hovedsak studier der laveregradsstudenter var en del av, eller hele, gruppen som ble undersøkt, da vi er av den oppfatning at om vi kan forbedre søkeopplevelsen for disse, vil det også ha positive ringvirkninger for andre sluttbrukere, som masterstudenter og ansatte.

Vi valgte en rekke søketermer og utførte søkene våre i det vi anser som noen av de mest relevante artikkeldatabasene innen bibliotekfag. Studien ble begrenset til bøker og fagfellevurderte artikler tilgjengelig gjennom Universitetsbibliotekets abonnementer på artikkeldatabaser og samlinger. På grunn av den raske utviklingen i søkeverktøyteknologi og informasjonsadferd valgte vi å avgrense til dokumenter utgitt etter 2000, de fleste etter 2010. Søkene ble utført i perioden 21. juni – 4. august 2017, og det er kun dokumenter på engelsk som har blitt benyttet. Studien siktet altså ikke på å være en uttømmende litteraturstudie, men bør likevel kunne sies å utgjøre et representativt utvalg av relevant, nyere forskning på temaet.

Det ble benyttet kombinasjoner av søketermer som information searching, topical search, subject headings, searching, exploratory search, web searching, internet searching, discovery systems, primo, thesaurus, undergraduate students og undergraduates. Vi søkte i UiOs instans av Oria, og artikkeldatabasene LISA og LISTA.

Det er gjort en rekke studier av studenters søkeatferd i discovery-systemer, og en stor andel av disse er systemorienterte undersøkelser og evalueringer av *systemet* framfor undersøkelser av studentenes *atferd*, med et formål om å enten velge en systemleverandør eller å forbedre søkesystemet.

I Oria, og flere andre discovery-baserte søkesystemer, består grensesnittet for enkle søk av søkeboks og muligheter for å filtrere og kutte ned trefflistene gjennom valg av fasetter i trefflisten etter at søket er utført (post-filtrering).

I en brukerundersøkelse av EBSCO Discovery Service hadde studentene problemer med å utnytte filtreringsfunksjonene, hvis de i det hele tatt benyttet dem (Cassidy, Jones, McMain,

Shen & Vieira, 2014). I andre undersøkelser har disse mulighetene både blitt trukket fram som nyttige og noe studentene bruker (Kliwer, Monroe-Gulick, Gamble & Radio, 2016; Mahoney & Leach-Murray, 2012). Mahoney og Leach-Murray trekker fram at det er bedre å filtrere en treffliste etter et bredt søk framfor å formulere og fylle ut et avansert søk først. «[...] we had the impression that our main users were gravitating to a more Google-like search experience without having to select search limits before performing a search» (s. 332).

Siden discovery-baserte søketjenester søker i en rekke kilder samtidig, vil de nødvendigvis også gi langt flere treff og lengre trefflistene enn et søk i en tradisjonell bibliotek katalog. Mahoney og Leach-Murray så at studentene hadde problemer med å skille mellom artikkeltittel og tidsskrifttittel i en referanse, og mener at dette kan bety at de også vil ha problemer med å navigere i en treffliste der dokumenttyper som tidsskriftartikler og bøker er blandet. Studenter har også vist en tendens til å ikke være villige til å se over flere enn en håndfull av de første treffene i en lang liste før de heller utfører et nytt søk (Kliwer et al., 2016), og stoler på at relevanssorteringen i trefflistene plasserer «det beste» treffet øverst (Asher, Duke & Wilson, 2013).

Flere undersøkelser trekker fram studentenes manglende forståelse av hvilke kilder det søkes i, og hvordan søket og rangeringen av trefflistene fungerer (Asher et al., 2013; Dalal, Kimura & Hofmann, 2015). Dette er i tråd med undersøkelser av unge menneskers informasjonskompetanse, der de viser en tendens til å bruke kort tid på å vurdere trefflistene og velge kilder som var «gode nok».

En vanlig antakelse om de som er oppvokst med internett, er at de navigerer i søk og informasjonskilder raskere og lettere enn eldre generasjoner. Denne antagelsen er tilbakevist flere ganger de siste årene (David, Ian, David & Peter, 2011; Kirschner & De Bruyckere, 2017; Rowlands et al., 2008). Undersøkelser av unge menneskers informasjonskompetanse og –atferd viser derimot at de er raskere til å godta et svar som «godt nok» (David et al., 2011; Porter, 2011), og at de ikke håndterer multitasking bedre enn andre (David et al., 2011; Kirschner & De Bruyckere, 2017).

«Digital natives» har heller ikke vist seg å være flinkere til å søke enn andre (Rowlands et al., 2008). De bruker gjerne Google eller Wikipedia som utgangspunkt for søk og bruker nærmest utelukkende naturlig språk som søkeord (Porter, 2011).

I intervjuer med bibliotekarer, studenter og uteksaminerte studenter fant Osborne og Cox (2015) at de fleste de intervjuet benyttet flere søkeverktøy for hvert søk. «(...) indicating that neither the internet nor the traditional library catalogue could fully meet their information retrieval needs.» (s. 41). Dette passer med funnene fra Kliwer, Monroe-Gulick, Gamble og Radio (2016), som observerte at studentene vekslet mellom å søke i Primo og andre søketjenester som Google under litteratursøk.

I litteraturen opptrådte emneorienterte søk ofte som en del av større studier som fokuserte på søke- og informasjonssøkeferd hos studenter i en mer helhetlig kontekst. De studiene vi fant beskrev hvordan utilstrekkelig vokabular og kunnskap om fagbegreper kan være en utfordring for lavere grads-studenter. Studenter begynner gjerne et søk bredt for å orientere seg i fagområdet (Kliwer et al., 2016). Når disse søkene gir feil eller for få treff, har flere sett den samme tendensen til at studentene heller enn å bytte søkeord, legger til flere ord og på den

måten formulerer mer presise søk (Cassidy et al., 2014; Dalal et al., 2015; Dempsey & Valenti, 2016). Unge mennesker benytter seg ofte av et naturlig språk. Ordene de velger ligner ofte emneord eller nøkkelord, men de har ofte problemer med å velge de «riktige» ordene, altså de som samsvarer med kontrollerte emneord og andre søkeinnganger (Dalal et al., 2015; Dempsey & Valenti, 2016; Drabentstott, 2003; Kliwer et al., 2016). I en analyse av studenters søkehistorikk fra Summon fra Meadow og Meadow (2012) fant de at flertallet av søkene var emnesøk, og at søkene var av god kvalitet. «These results could signify that the single search box model of discovery is sufficient for most students. However, there are some who do not understand the most effective ways to search and others who do not even understand the meaning of searching library resources» (s. 171).

En mulighet til å navigere i emneordssystemer kan være en hjelp for studenter som sliter med å formulere informasjonsbehovet sitt gjennom gode søkeord. Bauer og Peterson-Hart gjennomførte en brukertesting av et grensesnitt som lot studentene navigerte via fasetter i LCSH (Library of Congress Subject Headings). De fant at fasettene var populære hos brukerne, men at navigering via fasettene ikke ble brukt nevneverdig. «[...] simply exposing existing subject headings in faceted display will not be enough to significantly increase use of subject headings» (2012, s. 355).

Shiri og Revie (2006) gjennomførte en undersøkelse av et «thesaurus enhanced search system on the web». Her var informantene akademisk ansatte og høyere grads-studenter. Med utgangspunkt i en tesaurus fikk informantene forslag til hvordan de kunne utvide og forbedre søkene sine. De fant bl.a. at søkeforslagene gjorde informantene, og da spesielt studentene, mer oppmerksomme på alternative søketermer. Brukerne mente forslagene fra tesaurusen var nyttige.

Antell og Huang (2008) diskuterer også utfordringene studenter har knyttet til å identifisere emneord og hvordan det påvirker søkeprosessen. «[...] even when students do locate correct subject terms, it is in a hap hazard way – they stumble upon them rather than employing a search strategy to locate them. The subject terms that they locate are rarely the most appropriate ones for their search needs. But perhaps most disturbing is the fact that, during the observation interviews, so many students simply gave up when the OPAC search did not meet their needs. They assumed that the library – with its millions of volumes – did not have any relevant materials for their needs.» De fant at studentene jevnt over hadde problemer med å utnytte emneordene og utføre gode søk. Hjelpetiltak som foreslås er bl.a. å tilby hjelp «just in time», altså rett etter et (mislykket) søk. De foreslår tagging som et tillegg til emneord og trekker fram navigering via fasetter i trefflisten som en nyttig måte å vise frem emneordene på.

Med tanke på vårt prosjekt kan de viktigste funnene oppsummeres slik: «Google-generasjonen» er ikke raskere enn eldre generasjoner når de søker eller navigerer i søkesystemer (David et al., 2011; Kirschner & De Bruyckere, 2017; Rowlands et al., 2008), de er heller ikke bedre til å multitask (David et al., 2011; Kirschner & De Bruyckere, 2017), men de har en tendens til å akseptere trefflistene de får (David et al., 2011; Porter, 2011). Studenter bruker både bibliotekenes discovery-systemer og andre internett-ressurser, noe som indikerer at verken internett eller discovery-systemene oppfyller informasjonsbehovene deres fullstendig (Osborne & Cox, 2015; Kliwer et al., 2016). De bruker naturlig språk, men har problemer med å finne søketermer som korrelerer med systemets emneord (Dalal et al., 2015; Dempsey & Valenti, 2016; Drabentstott, 2003; Kliwer et al., 2016), og de har en tendens til å søke bredt (Kliwer et

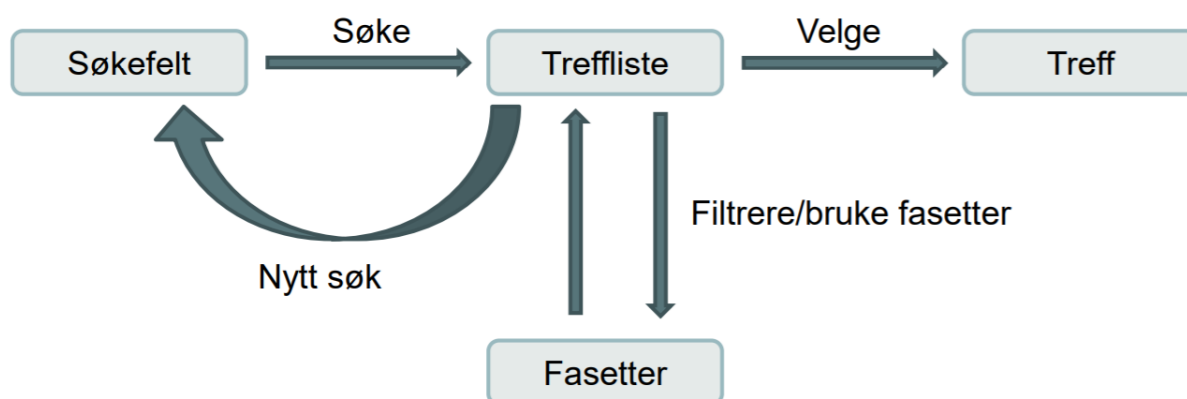
al., 2016). Studenter har dessuten vanskelig for å bruke systemenes avgrensingsmuligheter i søk (Cassidy et al., 2014), og mangler forståelse for hva bibliotek katalogen inneholder og hvordan relevansrangeringen i trefflisten fungerer (Asher et al., 2013; Dalal et al., 2015).

I litteraturen vi fant var studiene som regel enten systemorienterte, og fokuset lå på generell søkeadferd i bestemte sluttbrukersystemer, noe som ofte utelot emne-aspektet, eller brukerorientert, hvor en del av systemegenskapene vi ønsket, som at vi står overfor et discovery-system, uteble. Vi ønsket å få vite noe om kombinasjonen, studenters emnesøk i discovery-systemer, og dette er mye av grunnen til at vi gikk videre med en egen brukertesting.

Observasjon av studenter

For å kunne utnytte mappingene på en hensiktsmessig måte for å forbedre emnesøk i sluttbrukertjenestene våre, ville vi først kartlegge hvordan studenter bruker tilbudet som finnes i Oria i dag. I tråd med kunnskapen vi fikk gjennom litteraturstudien, utformet vi en rekke oppgaver til en brukertest hvor vi kunne observere hvordan studenter går til verks i Oria for å løse emnebaserte informasjonsbehov. Der en søkelogganalyse ville kunne fortelle oss noe om hva studentene søker på, kunne en brukerobservasjon fortelle oss noe om hvorfor og hva de tenker mens de søker.

Vi så for oss en ideell brukerreise gjennom Oria som i figur 1 under: brukeren søker først i grensesnittets søkefelt, og om trefflisten ikke er tilfredsstillende velger vedkommende enten å filtrere resultatene ved hjelp av fasetter eller å gjøre et nytt søk. Dette steget gjentas til brukeren får et tilfredsstillende resultat, og klikker seg videre inn på treffet fra trefflisten.



Figur 1: Brukerreise gjennom Oria for å løse et informasjonsbehov.

Det vi ønsket å fange opp med vår undersøkelse var hvor i prosessen studentene eventuelt føler at de møter på vanskeligheter, hvilke ord de bruker og hvor de finner dem, om de bruker filtreringsmulighetene, og om vi kunne identifisere steder mappingene kan være til hjelp.

Vi forfattet fire oppgaver (se Vedlegg 1) av relativt allmenn karakter, som bevisst var konstruert på en måte som ville å gi studentene i studien vanskeligheter med det vi, basert på erfaring med egne brukere, samt det vi lærte av litteraturstudien, mistenkte at studenter kan støte på i emnebasert informasjonssøking. Vi ønsket for eksempel å se når de valgte å søke på norsk, og når de søkte på engelsk, hvordan de valgte seg ord å søke på, og hvordan de håndterte

trefflister som ikke stod til forventningene. Testbrukerne fikk instruksjoner om å gå frem som de normalt ville gjort for å løse de fiktive oppgavene, og at de kunne søke hvor de ville, men at de måtte ende opp med relevante dokumenter i Oria til slutt.

Fem studenter fra forskjellige studieprogram ble rekruttert, så flere fagområder fra både HumSam og Realfag var representert. Datainnsamlingen foregikk ved at en fra prosjektet satt i rommet med testbrukeren og noterte underveis. Testbrukerne ble bedt om å tenke høyt mens de løste oppgavene, i tillegg ble det tatt opptak av lyd og skjerm. Etter endt datainnsamling transkriberte vi opptakene, og fargekodet søkeord, søkeforslag og negative/positive kommentarer fra brukerne. Deretter analyserte vi notatene og transkriberingene.

Resultatene våre stemte i stor grad overens med resultatene vi fant i litteraturstudien. Vi så først og fremst veldig liten variasjon i hvilke ord studentene valgte å søke på. Nesten alle søketermene var hentet direkte fra oppgaveteksten.

«Jeg har problemer med å finne, på en måte, noen knagger jeg kan feste søkene mine på.»

Studentene uttrykte aldri at det lå en spesifikk plan eller noen bevisst tanke bak valget av søketermer. De reflekterte i veldig liten grad over søk eller trefflistene i sin helhet, de snakket heller om bestemte treff og hvorvidt de var relevante eller ikke. Om de ikke fant noen relevante treff, skrev de inn et nytt søk med nært identiske søkeord. De brukte få termer i hvert søk, og det var lite hierarkisk navigering. Ofte inkluderte de preposisjoner og konjunksjoner som «i», «på» og «blant». I tillegg brukte de ofte geografiske termer som avgrensing.

Eksempel på hvordan studenten brukte veldig få termer, og søkte veldig bredt:

Student: Vet egentlig ikke helt hva annet jeg skulle søkt på, enn «hval i Norge», hvaler, det blir...

Studenten går tilbake til fanen med Google.

Studenten googler «fakta om hval».

Mangelen på refleksjon rundt søkingen kan muligens forklares ved at det er enklere å kommentere og si noe om det man faktisk ser på skjermen enn hva man har i hodet, men det kan også bety at studentene ikke hadde et bevisst forhold til søkeordene de valgte.

En annen tendens som kan tyde på at brukerne er vant med søkesystemer som Google var at de viste en ekstrem tillit til systemet, og virket overbevist om at dersom trefflistene ikke stod til forventningene, så var det fordi det ikke fantes dokumenter. De sa seg fort fornøyde med det de fant, og så sjeldent eller aldri på treff etter de to første sidene.

«Jeg føler ikke at det er noe altså, føler ikke det er noe særlig mer jeg kunne lagra.»

Studentene brukte få synonymer, og om de følte de behøvde mer informasjon om et tema så valgte de å søke etter emnet i Google. Om de fikk null treff søkte de ofte mye bredere, og om det var få treff undersøkte de ofte postene nøye, uansett hvor irrelevante treffene var. Ingen av studentene kommenterte i utgangspunktet støy eller ekstremt lange trefflister, og de gav ikke inntrykk av å ha bevisste tanker rundt hvordan de kunne innsnevre søkene utover å bruke en

annen kombinasjon av de samme ordene. De valgte heller å lete nøye gjennom trefflistene de fikk, fremfor å bearbeide selve søket.

Eksempel på hvordan studenten gikk fra null treff til et mye bredere søk:

Studenten gjør et enkelt søk på «insektsforskere» i Oria, ingen treff.

Student: Nei.

Prosjektmedarbeider: Hva tenker du nå?

Student: Kanskje på... insekt. Insekter?

Studenten gjør et enkelt søk på «insekter» i Oria, 3625 treff.

Studenten scroller i trefflisten.

Filtrering ved hjelp av fasetter ble brukt i veldig liten grad, og vi noterte oss at de mest brukte fasettene var de som var plassert øverst på siden. Det mest brukte filteret var å avgrense til fagfelleverderte artikler, men vi fikk inntrykk av at studentene ikke var klar over at de dermed ekskluderte relevante bøker fra trefflisten.

«Også er det litt sånn slitsomt når det er så sinnssykt mange treff, så er det egentlig veldig mange måter å avgrense på, men det glemmer jeg.»

Vi valgte å ikke systematisk vurdere hvorvidt dokumentene de valgte var «riktige», men vi så mange etter en ganske strevsom søkeprosess satt igjen med dokumenter som enten var alt for generelle eller alt for spesifikke.

Et av hovedfunnene våre må kunne sies å være at det var vanskelig for studentene i undersøkelsen å komme på synonymer og alternative søketermer. Det er vanskelig å komme på søkeord man ikke ser, og det er en grunnleggende forventning i at søketjenester skal være google-aktige. De har høy tillit til systemet, og stoler på at det presenterer det mest relevante først, så det er tilsynelatende ikke noen praktisk forskjell i om trefflisten inneholder 100 eller 100 000 dokumenter: de ser kun på de to første sidene med treff. Tilliten skinner også igjennom ved at dersom de fikk få treff, så antok de at det var fordi det ikke fantes dokumenter om emnet, ikke fordi de burde søkt på en annen term.

Testbruker: «Et spørsmål om systemet deres: er det sånn at det mest essensielle er på toppen, selvfølgelig. Altså, i form av det man søker på, så blir det mindre og mindre?»

[...]

Prosjektmedarbeider: «Det er intensjonen, i alle fall»

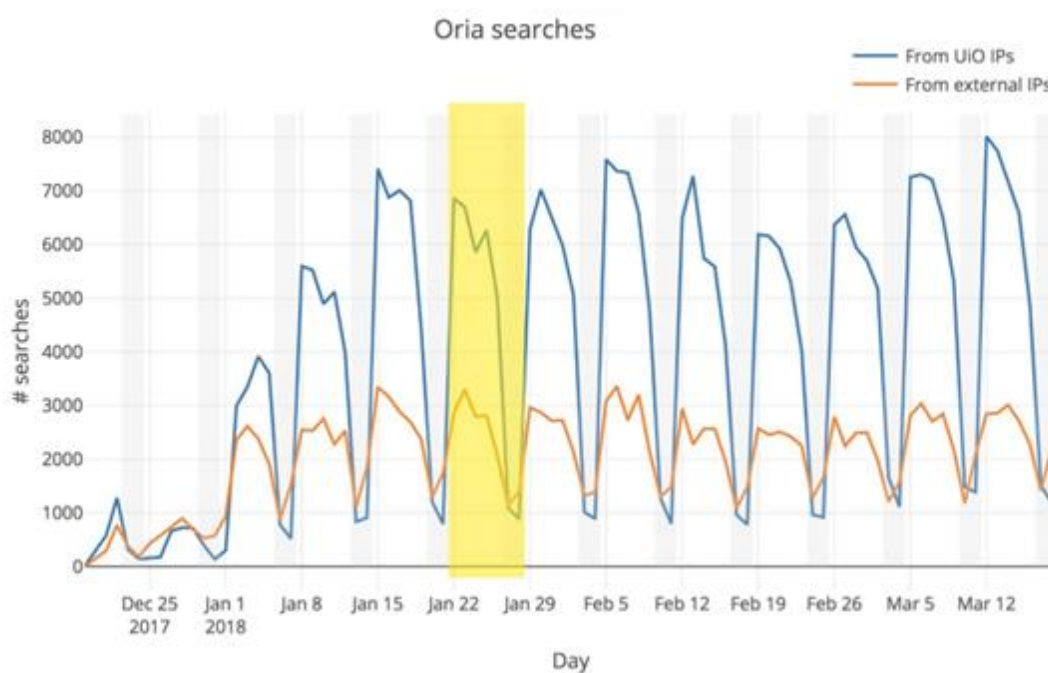
Søkelogganalyse

For å få et mer representativt innblikk i hvordan brukerne våre søker i Oria, besluttet vi å logge alle søkesesjoner over en viss periode. Vi vurderte hvordan vi eventuelt kunne begrense utvalget til kun å omfatte studenter, men endte med en løsning som logget alle søk i Oria, inklusiv søk av egne ansatte.

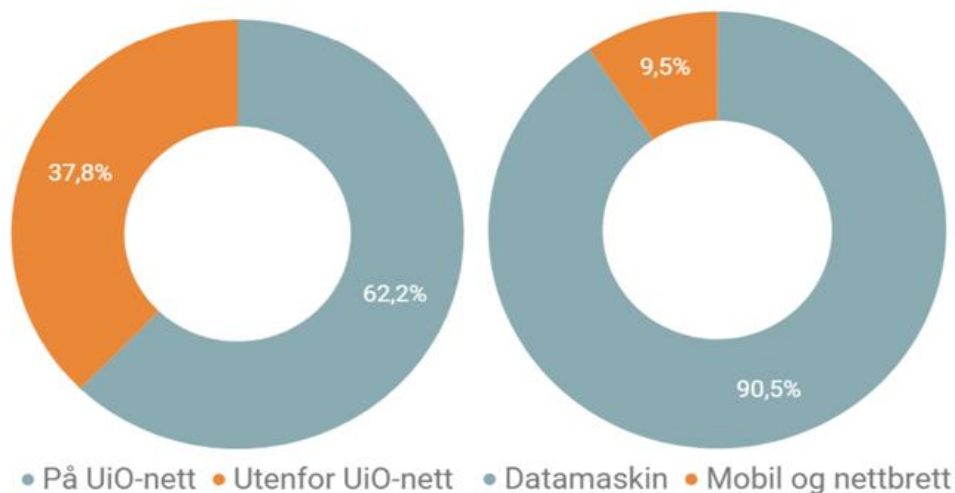
I dette utvalget, bestemte vi oss for en tilfeldig uke (uke 4, 2018). Vi samlet inn en mengde informasjon om disse sesjonene, for eksempel hvor lenge de varte, om søkene ble skrevet eller

limt inn, hvor mange treff søkene ga, om brukerne benyttet seg av avgrensing ved hjelp av fasetter, om de søkte i alle ressurser i Oria eller begrenset det til BIBSYS-konsortiets eller eget biblioteks dokumentbeholdning, om de brukte avansert søk, så på mer enn første side i en treffliste, eller tittet nærmere på noen av treffene.

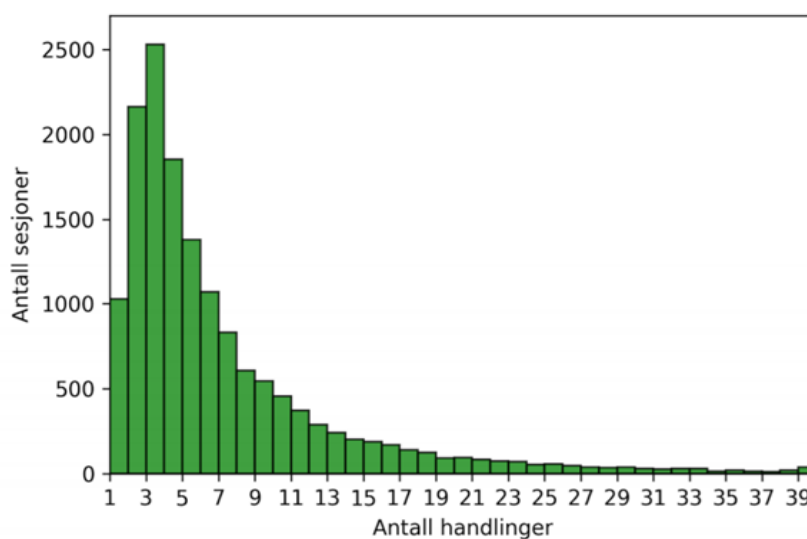
Som det framgår av figur 2 under var dette en ganske gjennomsnittlig uke, både når det gjaldt antall søk og hvorvidt søkene kom fra interne eller eksterne IP-adresser. En overvekt av brukere søkte fra UiO-nettet, men det var også en betydelig andel søk utenfra (figur 3, venstre del). Andelen mobilbrukere var lavere enn for en gjennomsnittlig nettside (figur 3, høyre del). Alt i alt inneholdt utvalget vårt ca. 15 000 sesjoner med ca. 49 000 handlinger (datasettet kan deles på oppfordring). Handlinger er interaksjoner med nettstedet som søk, avgrensninger, åpning av poster med mer. Som vi ser i figur 4 er mange sesjoner korte, noe som kan skyldes både at mange bruker Oria for å slå opp et kjent dokument og at mange gir opp. Samtidig ser vi at det er en lang hale av sesjoner med en betydelig mengde aktivitet.



Figur 2: Antall søk per dag i Oria-instansen til UBO fra UiO-IP-adresser og eksterne. Det gule feltet viser testperioden vår, som fremstår som en ganske gjennomsnittlig uke.



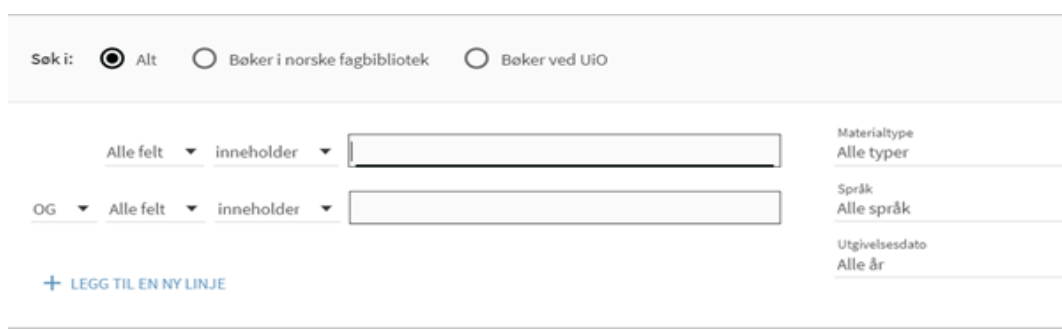
Figur 3: Brukersegmenter for testtuken vår.



Figur 4: Antall handlinger per sesjon.



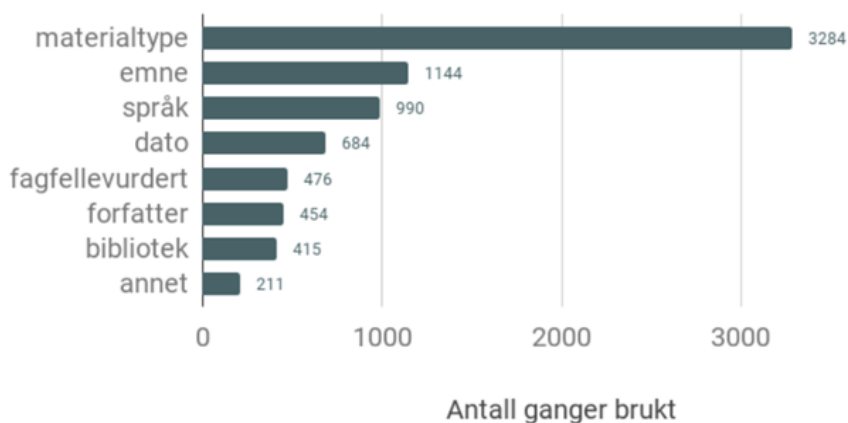
Figur 5: Enkelt søk i Oria.



Figur 6: Avansert søk i Oria.

97 % av sesjonene inneholdt minst ett enkelt søk (figur 5), mens 8 % av seksjonene inneholdt minst ett avansert søk (figur 6).

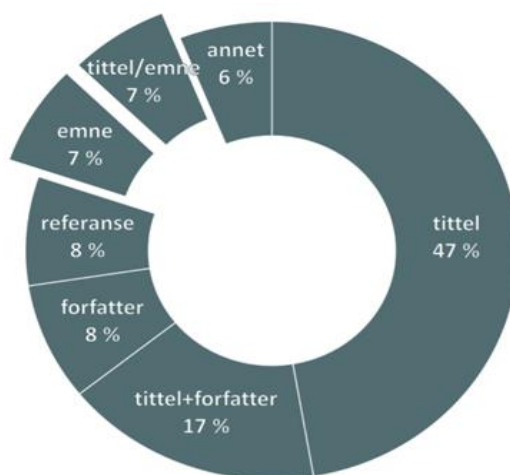
Standard oppsett for UBO i Oria er søk i alt materiale, altså både artikler og annet fra Primo Central Index (sentralindeksen), UBOs samlinger og BIBSYS-konsortiets samlinger. Treffmengden blir derfor for svært stor, og når relevanssorteringen ikke fungerer optimalt, vil det være formålstjenlig med avgrensing av søket ved hjelp av fasetter. Kun 12 % av sesjonene inneholdt imidlertid fasettavgrensinger. Når fasettering brukes, filtreres det for det meste på materialtype, emne og språk, se figur 7.



Figur 7: Bruk av fasetter.

For å kunne gå dypere inn i materien, hentet vi ut 2 000 tilfeldig valgte sesjoner fra hele datasettet, som ble lagt i et regneark for manuell analyse.

Siden én og samme sesjon kunne inneholde flere forskjellige typer informasjonssøk, f.eks. et tittelsøk fulgt av et emnesøk, delte vi dem opp i *søkeoppgaver*, der vi definerte en søkeoppgave som et sett av handlinger som så ut til å dekke et logisk adskilt informasjonsbehov. De 2000 *søkesesjonene* endte opp som 3900 *søkeoppgaver*, som alle ble kategorisert. En del søk ble gjenskapt i Oria for å kunne forstå dem bedre.



Figur 8: Fordeling av søkeoppgaver basert på manuell kategorisering.

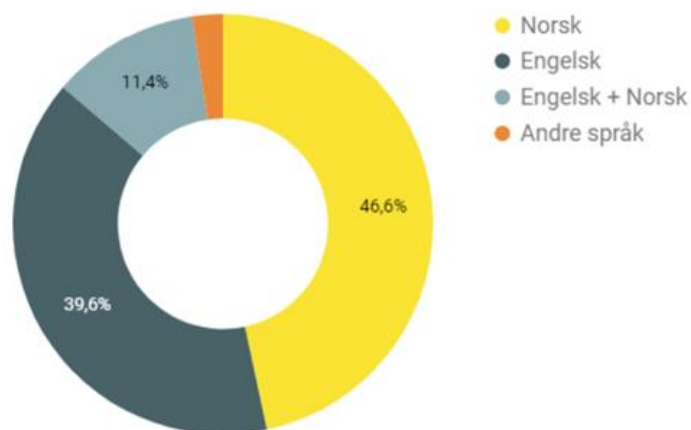
Figur 8 viser resultatet av kategoriseringen. Nær halvparten av søkene (47 %) var søk på tittel, fulgt av kombinasjon av forfatternavn og ord fra tittel (17 %) og forfatternavn (8 %). I kategorien referanse (8 %) har brukeren i de fleste tilfellene limt, noen ganger skrevet, inn en referanse, ofte til pensumlitteratur. Alle disse søkene kan karakteriseres som *verifiseringssøk* i og med at de er søk etter en kjent entitet. I den engelskspråklige litteraturen innen informasjonssøkeatferd beskrives dette gjerne som «known item search». Uansett betegnelse, er dette søket kommet i stand fordi brukeren vet at det finnes en gitt tittel, forfatter eller utgivelse og dermed søker etter denne.

I utvalget har vi også 14 % som vi kan betegne som emnemessige søk. Siden de fleste brukerne ikke søker spesifikt i emnefeltene, men i den generelle søkeboksen, har vi ikke kunnet fastslå 100 % sikkert om dette er søk etter et bestemt emne. I halvparten av tilfellene har det fremkommet ganske tydelig fra sesjonen som helhet at søkeren har vært på jakt etter stoff om et gitt emne. I korte sesjoner er dette vanskeligere, og i rundt halvparten av tilfellene har vi valgt å bruke kategoriseringen «tittel eller emne». F.eks. når en bruker søker etter et generelt emne som «medical genetics» og åpner en av mange bøker med dette emnet som tittel. Eller når en bruker søker etter «klinisk psykiatri» og ikke går videre.

At noen søk er mislykkede er lett å se, men å fastslå hvor mange som er vellykkede (i den betydning at informasjonsbehovet ble dekket), er ikke mulig uten å spørre brukerne. Vi definerte allikevel et kunstig suksesskriterium basert på hvorvidt et søk førte til åpning av en *tilsynelatende* relevant post eller nettressurs eller ei. *Tilsynelatende* er understreket fordi det selvfølgelig i mange tilfeller ikke var enkelt for oss å avgjøre relevansen for den faktiske brukeren, men vi gjorde allikevel et forsøk, uten å bruke altfor mye tid på dette. «Known item»-søk tilfredsstilte oftere dette suksesskriteriet (75 %) enn emnesøk (63 %), noe som kan tyde på at brukere oftere lykkes med førstnevnte, men disse tallene må tas med en klype salt.

Siden vi i mappingprosjektet har arbeidet med mapping av emneord, var det naturlig å gå litt dypere inn på søkeoppgavene som ble kategorisert som emnesøk, noe som utgjorde ca. 280 søkeoppgaver.

Det første vi gjorde var å kategorisere disse etter søkespråk, se figur 9. Siden én søkeoppgave kan bestå av flere etterfølgende søk, så vi at noen forsøkte å søke på flere språk, men svært få. I nesten halvparten av søkene (47 %) ble det kun brukt norske søketermer, mens nær 40 % kun brukte engelske. Bare 11 % søker både på norsk og engelsk. Siden vår Oria-instans inkluderer dokumenter på mange ulike språk, innebærer denne søkestrategien at brukerne lett kan gå glipp av mange potensielt relevante dokumenter.



Figur 9: Fordeling av søkespråk i de 280 emnesøkeoppgavene, manuelt klassifisert.

På tidspunktet da vi gjorde denne analysen var Realfagstermer allerede et flerspråklig vokabular; fullstendig oversatt til nynorsk, mer enn 90 % oversatt til engelsk, og med et lite antall vitenskapelige termer på latin, mens Humord kun inneholdt termer på bokmål. Vi foretok derfor et lite eksperiment ved at vi undersøkte om vi ville få bedre gjenfinning på emnesøkene i undersøkelsen dersom vi forutsatte at også Humord var oversatt til engelsk. Vi fant at 75 % av søketermene matchet Humord-vokabularet godt.

At det ville være et stort potensiale for bedre gjenfinning dersom Humord ble oversatt til engelsk, fikk oss til å rette en henvendelse til Nasjonalbiblioteket om å omdisponere noen av ressursene våre til engelsk oversettelse av vokabularet. Dette ga NB tillatelse til. Vi gjør rede for status for denne aktiviteten senere i rapporten.

I tillegg til norske og engelske ressurser, finnes det også ressurser på mange andre språk i Oria. Vår mapping til Deweys desimalklassifisering vil kunne føre til at disse ressursene gjenfinnes ved at søk mot ett vokabular også fører til søk mot tilknyttede Dewey-numre. Dewey-numrene vil kunne fungere som nav ut til andre ressurser med samme klassenumre. Disse ressursene kan enten ha emnebeskrivelser på andre språk eller mangle emneord i det hele. Potensialet som mapping har for bedre trefflistene i Oria vil bli diskutert senere i rapporten.

Når vi sammenlikner med observasjonsundersøkelsen, ser vi mange likheter i denne undersøkelsen, men også én vesentlig forskjell: brukerne i denne undersøkelsen bruker langt mer spesifikke søketermer enn studentene i observasjonsundersøkelsen.

Hovedaktivitet 2: Utnyttelse av mappinger i sluttbrukerverktøy

En viktig komponent i dette prosjektet har vært å utforske muligheter for utnyttelse av mappingene ikke bare i WebDewey, men også direkte i sluttbrukersøk. Sluttbrukersøk foregår i vårt tilfelle gjennom Oria, som gir søk både mot bibliotekskatalogen og mot artikler og annet materiale fra en ekstern indeks (Primo Central Index).

Utvidelse av søk fra mappinger (Mappingutvidelse)

Utnyttelsen vi primært har sett på er *søkeutvidelse*. Et eksempel på søkeutvidelse: Hvis brukeren søker etter «Gresk-romersk mytologi», kan søket utvides til å også inkludere Dewey-nummeret 292.13.

Arbeidet med utnyttelse av mappinger har tatt utgangspunkt i resultatene fra søkelogganalysen (Hovedaktivitet 1). Fordi bruken av *avansert søk* var liten (9 % av søk), har vi begynt med å se på mulighetene for å utnytte mappingene i *enkelt søk*.

Søkeutvidelse i enkelt søk byr imidlertid på en rekke utfordringer. Uten entitetsgjenkjenning, vet vi for eksempel ikke om brukeren er ute etter et emnesøk eller å finne igjen en kjent entitet (*known item search*), og en søkeutvidelse vil generere støy i sistnevnte tilfelle. Siden vi også har sett i søkelogganalysen at en klar overvekt av søkene er nettopp søk etter kjent entitet, er dette en risiko vi ikke kan ta lett på.

Selv med manuell kategorisering, var det 7 % av søkeoppgavene som vi ikke klarte å bestemme om var emne eller tittelsøk, så å implementere dette i et system er ikke trivielt, og en utvidelse må nok kunne aktiveres (*opt-in*) eller deaktiveres (*opt-out*) av brukeren.

Muligheter for mappingutvidelse i Oria

Siden Oria ikke er et system vi drifter selv, har vi diskutert muligheter for søkeutvidelse først med Unit 10. april 2018, og deretter med Ex Libris 21. august 2018. To ulike strategier er beskrevet under.

Utvidelse under indeksering

I Primo kan det settes opp såkalte normaliseringsregler som kjører når postene importeres fra kildesystemet (f.eks. Alma) og indekseres. For autoriserte emneord fra kontrollerte vokabular som Humord og Realfagstermer er det allerede satt opp regler som utvider postene med ikke-foretrukne termer (synonymer, oversettelser, m.m.), slik at disse også blir søkbare via emneindeksen.

Det skal være mulig å sette opp en tilsvarende utvidelse for tilmappede begreper. Hvis f.eks. en importert e-bok-post kommer med et Dewey-nummer, kan denne utvides med termer fra Realfagstermer og Humord.

Hovedulempen med en slik løsning er at utvidelsen ikke kan velges bort under søking, så det er viktig at utvidelsen er nokså presis. For eksakte mappings (=EQ) kan det allikevel være en løsning verdt å vurdere.

Utvidelse under søking

Primo tilbyr også en funksjonalitet kalt «Search Expansion With Controlled Vocabulary», som utvider søket med synonymer fra kontrollerte vokabularer som LCSH (Library of Congress Subject Headings) og MeSH (Medical Subject Headings), men som også gir brukeren muligheten til å velge bort utvidelsen (*opt out*).

En slik utvidelse er forbundet med mindre risiko, fordi brukeren selv kan velge bort utvidelser som genererer støy.

En utfordring ved utvidelse med et klassifikasjonssystem som Dewey er å presentere Dewey-klassen på en forståelig måte. En beskjed av typen «Søket inkluderer også Dewey-nummeret 709.04, trykk [her](#) for å kun søke etter «Moderne kunst» vil ikke gi mye mening for den vanlige bruker. Inkludering av klassebetegnelse hjelper heller ikke alltid: «Søket inkluderer også 709.4 1900-tallet, 1900-1999» i dette tilfellet. Generelt sett er det utfordrende å presentere en Dewey-klasse for brukeren på en forståelig måte, så det kan være det er bedre med en helt generell tekst («Søket ditt ble utvidet med ett eller flere klassifikasjonsnumre. [Vis detaljer] eller [deaktiver utvidelsen]»).

Etter at prosjektet er slutt kommer vi til å jobbe videre mot Unit og Ex Libris for å prøve å få testet en slik løsning.

Testverktøy

Parallelt med at vi har jobbet mot Unit/Ex Libris for å utforske muligheter for å implementere en søkeutvidelse i Oria, har vi laget et testverktøy for å få et bedre bilde av konsekvensene av en slik utvidelse.

Hvis en bruker søker etter et emne som finnes i et av våre kontrollerte vokabular (Humord og/eller Realfagstermer), og emnet er mappet til eksakt like (=EQ) Dewey-klasser, kanskje også tilnærmet like (~EQ) Dewey-klasser, har antakelsen vår vært at vi vil få en bedre treffliste (flere relevante treff høyt oppe) ved å utvide søket til å også inkludere Dewey-numrene til de tilmappede klassene. Antakelsen er basert på at Dewey-numre skal kunne gi en nokså presis emnemessig beskrivelse av et dokument, men forutsetter at brukeren faktisk er ute etter et emnesøk, ikke etter en kjent entitet (*known item search*). Hvis en bruker eksempelvis er ute etter boka med tittelen «Generell relativitetsteori» vil det kunne være støyende om dokumenter om emnet *Generell relativitetsteori* havner foran den søkte tittelen i trefflista.

Siden Dewey-numre som oftest er tilordnet bøker, ikke artikler, er det også rimelig å anta at en mappingutvidelse med Dewey-numre vil løfte frem bøker framfor artikler. Noe som kan være bra eller ikke bra, avhengig av hva brukeren er ute etter. Mer generelt er utfordringen i Oria at systemet inneholder poster fra mange ulike kilder med veldig ulik metadatakvalitet og -omfang, og at forbedring av en type søk fort kan gå på bekostning av en annen type søk.

For å få et visst inntrykk av effekten av mappingutvidelse, gjør testverktøyet to parallelle søk mot Oria-API-et; ett uten mappingutvidelse og ett med mappingutvidelse, og sammenligner resultatlistene visuelt. Siden vi har sett i brukerundersøkelsen at få brukere ser gjennom mange treff, har vi valgt å bare sammenligne de første 20 resultatene. Hvis et dokument havner lenger ned i resultatlisten er det i praksis «utenfor rekkevidde» for brukeren – det er usannsynlig at brukeren vil se det, selv om det teknisk sett er der.

I eksempelet i figur 10 er det søkt på «Feminisme» i alle felt, noe som gir 14 630 treff uten mappingutvidelse (venstre kolonne) og 16 606 treff med (høyre kolonne). Figur 11 viser det samme søket mot emnefeltet i Oria, ikke alle felt. Det gir 7 747 treff uten mappingutvidelse (venstre kolonne) og 9 716 med (høyre kolonne). Vi kan se at trefflistene er ulike i forhold til sortering. Vi ser i dette tilfellet at den prosentvise forskjellen i trefflisten på søk med og uten mappinger, øker når vi kun søker i emnefeltet. Stikkprøver tyder på at dette er gjennomgående.

Search field: Any field | Query: Feminisme | Instance: UIO | Limit: 20

Without mappings

any,contains,Feminisme

Isr10,exact,305.42

Isr10,begins_with,305.420*

Isr10,exact,320.5622

Isr10,begins_with,320.56220*

14 630 hits after 0.646 seconds [SHOW IN ORIA](#)

book 160 s. (Print book@UBO) 1 → 1
Hva er feminisme / Universitetsforl. cop. 2009
 Cathrine Holst (1974-)\$SQCathrine Holst (1974-)
 Subjects: **Feminism** **Feminisme** **Feministisk teori**
 Kvinnebevegelse Kvinnefrigjøring Likestilling
 Dewey: 305.42
[More editions](#)

book 217 s. 2 → X
Feminisme / Aarhus Universitetsforlag 2012
 Subjects: **Feminisme** **Feministisk litteraturteori**
 Dewey: 082 305.42 801.95 801.95082

book 64 s. ill. 3 → X
Feminisme / Libretto forl. cop. 2004
 Kaye Stearman\$SQKaye Stearman
 Subjects:
 Dewey: 305.42

article 4 → X
Féminisme /
 s.n.
 Subjects: **Language**
 Dewey:

article 5 → X
Féminisme /
 s.n.
 Subjects: **Language**
 Dewey:

article 6 → X
Féminisme /
 Subjects:
 Dewey:

book 442 s. ill. (Print book@UBO) 7 → X

With mappings

any,contains,Feminisme

Isr10,exact,305.42

Isr10,begins_with,305.420*

Isr10,exact,320.5622

Isr10,begins_with,320.56220*

16 606 hits after 0.874 seconds [SHOW IN ORIA](#)

book 160 s. (Print book@UBO) 1 → 1
Hva er feminisme / Universitetsforl. cop. 2009
 Cathrine Holst (1974-)\$SQCathrine Holst (1974-)
 Subjects: **Feminism** **Feminisme** **Feministisk teori**
 Kvinnebevegelse Kvinnefrigjøring Likestilling
 Dewey: 305.42
[More editions](#)

book 268 s. ill. 25 cm (Print book@UBO) 10 → 2
**Sex, vold og feminisme : hvordan voldtekt og porno
 ble politisk på 1970-tallet** / Cappelen Damm
 akademisk 2018
 Trine Rogg Korsvik (1972-) (forfatter)\$SQTrine Rogg
 Korsvik (1972-)
 Subjects: **Feminisme** **Frankrike** **Kvinnebevegelser**
 Norge Pornografi Prostitusjon
 Seksuelle overgrep Voldtekt
 Dewey: 305.42
[More editions](#)

book 250 s. (Print book@UBO) X → 3
**Under det rosa täcket : om kvinnlighetens vara och
 feministiska strategier** / Wahlström & Widstrand
 1996
 Nina Björk (1967-)\$SQNina Björk (1967-)
 Subjects: **Feminisme**
 Dewey: 305.42
[More editions](#)

book 231 s. (Print book@UBO) X → 4
Feminist studies, critical studies / Macmillan Press
 1988
 Subjects:
 Dewey: 305.42
[More editions](#)

book XIX, 715 s. (Print book@UBO) X → 5

Feminisme (humord)

Type:
Mappings
 closeMatch Kvinnesak (Realfagstermer)
 exactMatch 305.42
 Kvinners sosiale rolle og status (DDC23)
 exactMatch 320.5622
 Feminisme (DDC23)
[» Skosmos](#)

Feministisk filosofi (realfagstermer)

Type:
Mappings
 exactMatch 396 Feminisme (Tekord)
 exactMatch 305.42
 Kvinners sosiale rolle og status (DDC23)
 exactMatch 320.5622
 Feminisme (DDC23)
 closeMatch ()
[» Skosmos](#)

Figure 10: Enkelt søk etter «Feminisme» i testverktøyet. Helt til høyre side vises treff i Humord og Realfagstermer og =EQ-mappingene deres til Dewey-numrene 305.42 og 320.5622. I venstre kolonne indikerer rød bakgrunn dokumenter som forsvinner ut (av de første 20 treffene) når mappingutvidelsen skrur på. I høyre kolonne indikerer grønn bakgrunn nye dokumenter som kommer inn blant de første 20 treffene når utvidelsen skrur på. Grå bakgrunn indikerer dokumenter som finnes blant de første 20 treffene både med og uten mappingutvidelse. Øverst til høyre i hvert treff vises det akkurat hvilken posisjon et dokument går fra og til i trefflista når mappingutvidelsen skrur på.

Without mappings	With mappings
<input checked="" type="checkbox"/> sub,contains,Feminisme	<input checked="" type="checkbox"/> sub,contains,Feminisme
<input type="checkbox"/> Isr10,exact,305.42	<input checked="" type="checkbox"/> Isr10,exact,305.42
<input type="checkbox"/> Isr10,begins_with,305.420*	<input checked="" type="checkbox"/> Isr10,begins_with,305.420*
<input type="checkbox"/> Isr10,exact,320.5622	<input checked="" type="checkbox"/> Isr10,exact,320.5622
<input type="checkbox"/> Isr10,begins_with,320.56220*	<input checked="" type="checkbox"/> Isr10,begins_with,320.56220*
7 439 hits after 0.03 seconds (fra cache)	9 420 hits after 0.047 seconds (fra cache)
SHOW IN ORIA	SHOW IN ORIA
<p>book 231 s. (Print book@UBO) 1 → 1</p> <p>Feminist studies, critical studies / Macmillan Press 1988</p> <p>Subjects:</p> <p>Dewey: 305.42</p> <p>More editions</p>	<p>book 231 s. (Print book@UBO) 1 → 1</p> <p>Feminist studies, critical studies / Macmillan Press 1988</p> <p>Subjects:</p> <p>Dewey: 305.42</p> <p>More editions</p>
<p>book 83 s. 22 cm (Print book@UBO) 2 → X</p> <p>Farvel alle hemmeligheter / Aschehoug 2018</p> <p>Helene Guåker (1983-) (forfatter)\$QHelene Guåker (1983-)</p> <p>Subjects: Feminisme</p> <p>Dewey: 839.82 839.8218</p> <p>More editions</p>	<p>book (Print book@UBO) 15 → 2</p> <p>The dialectic of sex : the case for feminist revolution / Jonathan Cape 1971</p> <p>Shulamith Firestone\$QShulamith Firestone</p> <p>Subjects: Feminisme Kjønnsroller</p> <p>Dewey: 305.42</p> <p>More editions</p>
<p>book LII, 139 s. (Print book@UBO) 3 → 12</p> <p>Kvinneundertrykkelsen / De norske bokklubbene 2004</p> <p>John Stuart Mill (1806-1873)\$QJohn Stuart Mill (1806-1873)</p> <p>Subjects: Feminisme</p> <p>Dewey: 192 305.42</p> <p>More editions</p>	<p>book ix, 252 s. (Print book@UBO) 18 → 3</p> <p>What is feminism? / Basil Blackwell 1986</p> <p>Subjects:</p> <p>Dewey: 305.42</p> <p>More editions</p>
<p>article S. [1] (Print book@UBO) 4 → X</p> <p>Kvindespørgsmaalet og Spørgvognskuskene / Aftenposten A/S 1883</p> <p>Camilla Collett (1813-1895)\$QCamilla Collett (1813-1895)</p> <p>Subjects: Feminisme</p> <p>Dewey:</p>	<p>book 250 s. (Print book@UBO) X → 4</p> <p>Under det rosa täcket : om kvinnlighetens vara och feministiska strategier / Wahlström & Widstrand 1996</p> <p>Nina Björk (1967-)\$QNina Björk (1967-)</p> <p>Subjects: Feminisme</p> <p>Dewey: 305.42</p> <p>More editions</p>
<p>book 267 s. fig. (Print book@UBO) 5 → 13</p> <p>Makt & kön : tretton bidrag till feministisk kunskap / Brutus Östlings Bokförlag 1997</p> <p>Subjects:</p> <p>Dewey: 305.42</p>	<p>book XVII, 171 s. fig. (Print book@UBO) X → 5</p> <p>What is feminism? : an introduction to feminist theory / Sage 1999</p> <p>Chris Beasley\$QChris Beasley</p> <p>Subjects:</p> <p>Dewey: 305.42</p> <p>More editions</p>
<p>article S. [225]-228 (Print book@UBO) 6 → X</p>	<p>book VII, 280 s. (Print book@UBO) X → 6</p>

Feminisme (humord) ^

Type:

Mappings

closeMatch [Kvinnesak \(Realfagstermer\)](#)

exactMatch [305.42](#)

Kvinnens sosiale rolle og status (DDC23)

exactMatch [320.5622](#)

Feminisme (DDC23)

[» Skosmos](#)

Feministisk filosofi (realfagstermer) ^

Type:

Mappings

exactMatch [396 Feminisme \(Tekord\)](#)

exactMatch [305.42](#)

Kvinnens sosiale rolle og status (DDC23)

exactMatch [320.5622](#)

Feminisme (DDC23)

closeMatch ()

[» Skosmos](#)

Figur 11: Det samme søket mot emne-feltet (avansert søk), ikke enkelt søk mot alle felt.

Funn fra testverktøyet

Fra søkelogganalysen (Hovedaktivitet 1) hentet vi ut alle søkene klassifisert som emnesøk, og testet de 30 første emneordene som har =EQ og/eller ~EQ-mappinger. Vi har så sammenlignet treffmengdene og relevansen i trefflistene med og uten utvidelse med tilmappede Dewey-numre.

Med mappingutvidelse får vi naturligvis flere treff enn uten, men i søketilfellene vi har sett på har forskjellen egentlig vært overraskende liten. En viktig grunn til dette er antakelig at vi allerede har oversatt brorparten av kildevokabularene til engelsk, noe som fører til at vi allerede ved enkelt søk får treff i den engelske litteraturen ved søk på norsk.

Siden treffmengdene i Oria er så store, og vi gjennom våre undersøkelser har sett at studentene sjelden ser på mer enn rundt to sider av trefflisten, anså vi det som hensiktsmessig å heller vurdere kvaliteten til de 20 første treffene, enn kvantiteten. Vi hadde en antakelse om at i de tilfellene hvor treffmengden ikke økte i særlig grad, ville søketermene med mappinger endre relevansrangeringen og resultere i mer relevante trefflister. Innholdet i trefflisten vil altså i stor grad være den samme, men rekkefølgen vil endres, og relevansen bør bli bedre når man utvider med Dewey-nummer.

For et generelt emneord som *Spania*, der det er rimelig å anta at brukeren er ute etter historisk eller geografisk litteratur om Spania, vurderer vi at kvaliteten på de to første sidene av trefflistene blir vesentlig bedre med mappingutvidelse. Trefflisten uten mappinger inkluderer for eksempel CD-plater med lydspor kalt «Spania», samt en mengde kokebøker med spanske retter og noe juridisk litteratur. Av de 20 første treffene uten mappinger, antar vi at 10 ikke er relevante for brukeren. Med mappingutvidelsen anser vi 18-20 av de 20 første treffene som relevante.

Ved et søk på termen «Miljørett» uten mappingutvidelse, er 7 av de 10 første treffene to-siders artikler fra Nordisk domssamling. 10 av de 20 første treffene er slike artikler, som antakelig er for spesialiserte for noen som kun søker etter «Miljørett». Når vi skrur på mappingutvidelse, forsvinner disse lenger ned i trefflisten, til fordel for dokumenter som vi tror er bedre egnet for en innføring i dette emnet.

Generelt synes vi å se at mappingutvidelse med =EQ-er og ~EQ-er fører til flere relevante treff. Spesielt for generelle emner får vi bedre trefflister når vi inkluderer mappinger.

Men: jo snevrere emnet er, jo mindre sannsynlig er det at vi har en =EQ eller ~EQ til emnet. Vi tok ikke høyde for å utvide med de svakere relasjonene, BM og RM, i denne omgangen fordi vi i forkant av utviklingen av testverktøyet manuelt utforsket en rekke søk, og oppdaget at utvidelser med Dewey-numre mappet med RM og BM i mange tilfeller resulterte i støy. Dette gjaldt særlig i tilfeller der et emneord hadde fått mange mappinger, både =EQ, ~EQ, BM og kanskje til og med RM. I tilfeller hvor det finnes litteratur på et emne, vil ikke disse svakere relasjonene hjelpe brukeren i særlig grad.

For eksempel så vi at å utvide med BM og RM i et søk på «Nordøstpassasjen» resulterte i en treffliste med mye støy fordi vi har litteraturbelegg for termen. Samtidig ser vi for oss at det kan være mulig å heller utnytte disse relasjonene i et system som kan plukke opp at det er snakk om et veldig lite, eller ikke-eksisterende, litteraturbelegg. Termer som får en svakere mapping, særlig RM, er ofte vanskelige å plassere i WebDewey, og ofte vanskeligere å finne relevant

litteratur om. Her ser vi for oss at mappingene til termer som *kun* har BM og RM kan utnyttes i form av for eksempel valgfrie søkeutvidelser, eller som tips til videre søking. Vi anser det derfor som hensiktsmessig å utrede dette videre.

For de relativt enkle søkene vi har sett på, har vi altså sett en positiv effekt på relevansrangeringen i trefflistene i Oria når vi utvider søkene med =EQ og ~EQ-mappinger. Dette har nok litt med at Oria i utgangspunktet har en ganske dårlig relevansrangering for enkle søk.

En av utfordringene med å jobbe mot Oria er at vi ikke helt kan si *hvordan* relevansen beregnes, eller hva som gjør at ett treff rangeres høyere enn et annet. Det vi antar, etter å ha utført en rekke søk, er at treff på tittel vektet høyt i Oria. Magre metadata som kanskje kun har tittel og forfatter, kommer høyt opp i trefflistene, og vi antar at det er fordi det ikke er så mye annet å hente informasjon fra. Søketermen utgjør på denne måten en vesentlig del av postens metadata. Det er uheldig at dårlig metadatakvalitet kanskje gjør at treff som ikke nødvendigvis er relevante, vektet såpass høyt som de gjør.

For eksempel vil vi ved et søk etter «Tempus», altså grammatisk tid, få over 33 000 treff. Om vi så legger til det engelske «Tense», øker treffmengden til over 360 000. Tross tendensen til mindre støy i søk med mappingutvidelser, får vi som det 8. treffet det svenske tidsskriftet Tempus, som tar for seg verdensnyheter, i trefflisten med mappinger. Treff 12 er en artikkel fra et medisinsk tidsskrift. Felles for disse to treffene er magre metadata, stort sett kun tittel. Her er det lite gunstig at disse postene vektet så høyt i en treffliste som inkluderer søk etter Dewey-numre, når totalmengden treff er så stor.

Testverktøyet gir oss mulighet til å eksperimentere med både fritekstsøk og emnesøk. Etter hva vi har lært av tidligere forskning og egne undersøkelser, utfører studenter søk etter emne i liten grad. Når vi for forsøkets skyld avgrensner søket til emnefeltet, får vi en vesentlig mindre treffmengde, og vi kan da være helt sikre på at utvidelsen gir relevante treff. Utfordringen ligger i å vite at studenten gjør et emnesøk, noe vi ikke kan være helt sikre på når de velger å søke i scopet «alt». Det vi potensielt kan tilby, er å gi dem et valg når trefflisten presenteres: «Det ser ut som du søker etter et emne, ønsker du å inkludere klassenumre i søket ditt?». På denne måten kan vi hjelpe dem å spisse eget søk. Testverktøyet gir mulighet til å undersøke dette videre.

Vi ser at brukerne har mange skrivefeil når de søker, og at de sjelden søker på den formen for emneordet som vi bruker. Det bør derfor legges ressurser i forbedring av støtte for korrektur av skrivefeil og stemming av søkeord i Oria.

Vi kan konkludere med at å utvide søkestrenger i Oria med tilmappede Dewey-numre med relasjonene =EQ og ~EQ i søk på generelle emner har en positiv effekt på relevansrangeringen av treffene. Det er utfordrende å jobbe med et system hvor vi ikke kan forstå relevansrangeringen som ligger bak, men i den grad vi kan se en generell tendens, ser vi at trefflistene på ingen måte blir dårligere, og vi kan anta at trefflistene nå inneholder litteratur som ellers ikke ville blitt fanget opp.

Implementering av mappingutvidelse i Scriptotekets Emnesøk

Utover utnyttelse av mappingene i norsk WebDewey og i Oria, har vi også sett på utnyttelse i Scriptotekets Emnesøk¹⁰ (Heggø, Traavik & Lundevall, 2016), som er en egenutviklet webapplikasjon som demonstrerer noen muligheter for emnesøk som vi per i dag ikke har i Oria, men med data fra Oria.

Siden vi drifter denne applikasjonen selv, kan vi enkelt eksperimentere med den, så vi har implementert en enkel mappingutvidelse á la utvidelsen i testverktøyet. Hvis brukeren søker etter et emne som har en eller flere mappings med relasjonen =EQ, gis brukeren mulighet til å inkludere disse i søket, se figur 12. Resultatet blir et utvidet søk som inkluderer flere e-bøker vi ikke selv har klassifisert, se figur 13. Bytter en til «Bredt søk» inkluderes også mappings med relasjonen ~EQ.

The screenshot shows the search interface for 'Optoelektronikk'. At the top, there is a search bar with 'Optoelektronikk' entered and a dropdown menu set to 'Begynner med'. To the right of the search bar are buttons for 'Bredt', 'Snevert', and 'Alle bibliotek'. Below the search bar, a red oval highlights a suggestion: 'Vil du ha flere treff? Prøv kontrollert søkeutvidelse.' The main search results list several books related to optoelectronics, including 'Molecular beam epitaxy : materials and applications for electronics and optoelectronics' by Asahi, H., 'Graphene optoelectronics : synthesis, characterization, properties, and applications' by Yusoff, Abdul Rashid bin M., 'Electrical properties of materials' by Solymar, L., and 'Cambridge illustrated handbook of optoelectronics and photonics'. On the right side, there is a sidebar titled 'Optoelektronikk' which provides additional information, including synonyms in English and Norwegian, and related concepts in other vocabularies.

Figur 12: Emnesøk foreslår mappingutvidelse.

¹⁰ <https://app.uio.no/ub/emnesok/>

Begynner med ▾ Optoelektronikk

Bredt Snevert Alle bibliotek ▾

⚠ Søket ble utvidet med ett eller flere tilknyttede begreper («621.381045» i Norsk WebDewey, «Optoelektronikk» i Tekord). [Skru av søkeutvidelse.](#)

Handbook of organic materials for electronic and photonic devices (Woodhead Publishing, an imprint of Elsevier 2019)
Ostroverkhova, Oksana, editor.
E-bok (ingen tilgang)
621.381045

Molecular beam epitaxy : materials and applications for electronics and optoelectronics (Wiley, First edition. 2019)
Asahi, H.
Trykt utgave ved UIO.
Molekylærstråleepitaksi Optoelektronikk Spintronikk
621.38152 530.4275

Electrical and Optoelectronic Properties of the Nanodevices Composed of Two-Dimensional Materials : Graphene and Molybdenum (IV) Disulfide (Springer Singapore : Imprint: Springer 2018)
Liu, Cheng-Hua, author.
E-book
621.381045

Advances in imaging and electron physics. : Volume 201 (Academic Press 2017)
Hawkes, Peter W., author.

Optoelektronikk
Brukt for: Optronikk
English: Optoelectronics • Optronics • Optoelectronics
Nynorsk: Optoelektronikk • Optronikk
Se også: Fotonikk
Underordnet: Optoelektronikk : Håndbøker
Optoelektronikk : Lærebøker
Meld fra om problem

Optoelektronikk, også kalt optronikk er en sammenslåing av begrepene optikk og elektronikk og er relatert til en sensor som arbeider både med fotoner og elektroner. Et synonymt begrep er elektrooptiske system (EOS).

Bekreftet betydning [Les mer på Wikipedia \(nb\)](#) [Wikidata](#)

Related concepts in other vocabularies:

- Equivalent concept 621.381045 Optoelektronikk i Norsk WebDewey
- Equivalent concept 621.383 Optoelektronikk i Tekord [Vis i Skosmos](#)

Figur 13: Emnesøk med mappingutvidelse aktivert.

Videre arbeid

I alle former for søkeutvidelse vil det være til stor hjelp for systemet å kunne vite noe om hva slags type søk brukeren gjør. I enkelt søk hadde det vært interessant å utforske flere muligheter for å fange opp om brukeren er ute etter et bestemt dokument, et emne eller noe annet. Mulige hjelpemidler kan være:

- Autofullføring mot en liste av titler, forfattere og emneord. I de tilfellene der brukeren velger et forslag fra listen, får vi vite om brukeren er ute etter en kjent entitet eller et utforskende emnesøk (som kan utvides med mappinger).
- Entitetsgjenkjenning. Etter å ha konsultert folk med kompetanse på området, har vi fått forståelse for at dette er langt fra enkelt, men hvis en kan få det til å virke bra, vil det være et svært nyttig hjelpemiddel for å avgjøre når søkeutvidelse kan brukes og ikke. Det vil også kunne føre til langt høyere presisjon i søket. Når en bruker spør oss i skranken om en «innføringsbok i Java» skjønner vi mennesker med en gang at personen ikke er ute etter en bok om øya, vi kjenner igjen entiteten. Og skulle vi ikke skjønne det kan vi spørre: Mener du øya eller programmeringsspråket, og deretter avgrense søket. For å kunne gjøre entitetsgjenkjenning i et søkesystem, trengs det entiteter. Vi har tro på at kontrollerte og rike entiteter i form av emneord med synonymer og Dewey-mappinger vil kunne være nyttige for et entitetsgjenkjenningssystem, utforsking av dette vil kunne være et fremtidig prosjekt.
- Smart veiviser. I søkeundersøkelsen vår så vi lite av de google-lignende søkene («Hvorfor kommer det en pipelyd fra kneet mitt?», «Hva er tyngdeakselerasjonen på månen?» o.l.). Til forskjell fra generelle søkemotorer som Google har vi en mer begrenset mengde typer dokumenter, og en større mengde kontrollerte metadata om hvert dokument. Det kan være dette kunne vært utnyttet i en veiviser som kunne stilt spørsmål litt i samme gaten som spørsmål vi ville spurt brukerne i en fysisk skrankesituasjon.

I nær fremtid må vi nok leve med at vi i en overvekt av søkene *ikke* vet hva slags type søk brukeren gjør. Det sikreste vil da være å begrense mappingutvidelse til de tilfellene der vi vet at brukeren er ute etter emnesøk, det vil si når brukeren bruker en av emneindeksene i avansert søk eller bruker en av emnefasettene, men resultatene fra testverktøyet tyder på at vi heller ikke skal utelukke utvidelse i enkelt søk. Etter dette prosjektet er slutt kommer vi til å fortsette å arbeide mot systemleverandørene for en implementasjon av mappingutvidelse i Primo.

Hovedaktivitet 3: Ferdigstille mappingen

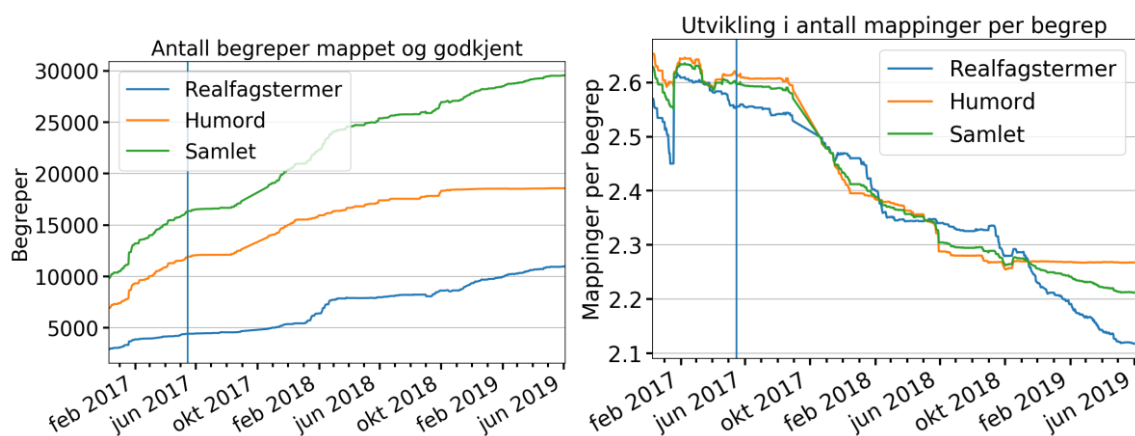
Felles for begge vokabularene

Et viktig delmål i prosjektet har vært å ferdigstille mappingen av begge vokabularene. I arbeidet med dette har vi benyttet mappingmetodikken og -verktøyet ccmapper som ble utviklet i prosjektet *Mapping mot norsk WebDewey (2015–2017)*, som igjen bygget på erfaringer fra to tidligere prosjekter ved UiO Realfagsbiblioteket.

Ved oppstarten av prosjektet *Mapping for sluttbrukertjenester* i mai 2017, var 66 % av Humord og 29 % av Realfagstermer mappet, til sammen rundt 42 000 mappinger fordelt på rundt 16 000 begreper (2,6 mappinger per begrep).

Ved avslutningen av prosjektet i juni 2019 hadde vi produsert 65 417 mappinger fordelt på 29 575 begreper (2,2 mappinger per begrep). Humord ble ferdig mappet senhøstes 2018, mens Realfagstermer ble ferdig ved prosjektets avslutning i juni 2019. Progresjonen for begge vokabular vises i figur 14 (venstre del). Et lite knippe emneord har vi valgt å ikke mappe fordi de kun gir mening i kombinasjon med andre emneord, mer om det under.

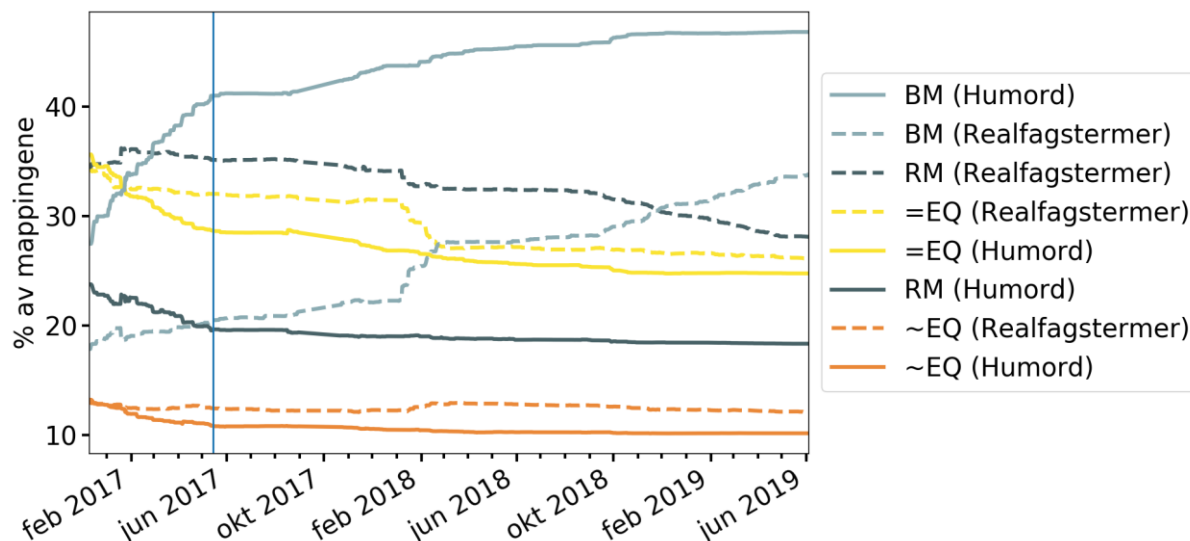
Som vi ser i figur 14 (høyre del) har antall mappinger per begrep falt mot slutten av prosjektet for begge vokabular. Dette skyldes sannsynligvis at vi har spart de vanskeligste og mest spesielle emneordene til slutt. For en del termer har det vært utfordrende å i det hele tatt finne en plassering i norsk WebDewey.



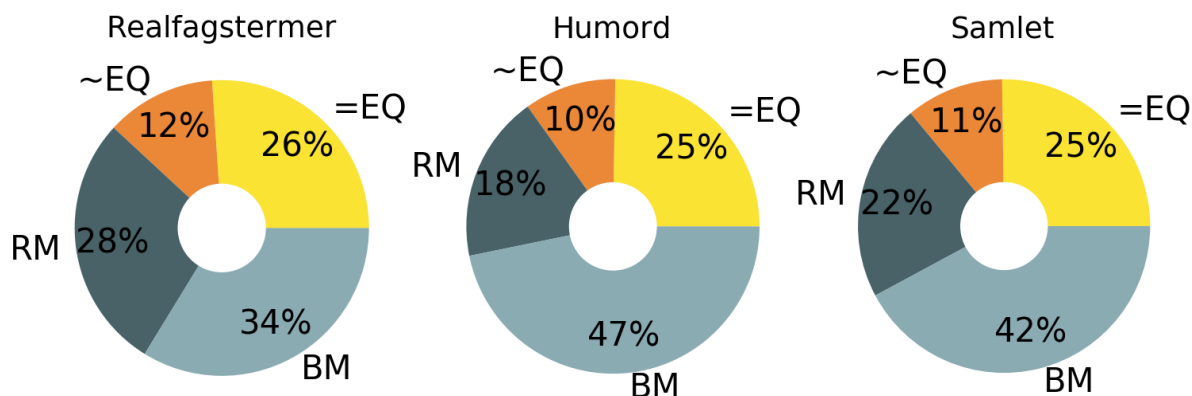
Figur 14: Progresjon i mappinger (venstre) og utvikling i antall mappinger per begrep (høyre). Den vertikale linjen viser oppstarten av prosjekt *Mapping for sluttbrukertjenester*.

Fordelingen av de fire ulike mappingrelasjonene har også endret seg gjennom prosjektet (figur 15). Figur 16 viser fordelingen ved prosjektslutt. Som vi ser har vi endt opp med 25 % =EQ, 11 % ~EQ, 42 % BM og 22 % RM.

Andelen =EQ og ~EQ er nokså lik i Humord og Realfagstermer, mens andelen BM er høyere i Humord og andelen RM er høyere i Realfagstermer.



Figur 15: Andel for hver av de fire mappingrelasjonene, fordelt på kildevokabular, fra vi begynte å bruke ccmapper i januar 2017 og frem til prosjektets slutt. Den vertikale linjen viser oppstarten av prosjekt *Mapping for sluttbrukertjenester*.



Figur 16: Fordeling av mappingrelasjoner ved prosjektslutt (juni 2019).

Hvis vi ser på hvor stor andel av begrepene som *kun* har RM-mappinger, er det snakk om 460 begreper i Realfagstermer (4,1 %) og 617 i Humord (3,4 %). Ser vi på begreper som *kun* har BM-mappinger, er det snakk om 3936 i Realfagstermer (35 %) og 8322 i Humord (46 %).

I rapporten *På vei mot en generell norsk tesaurus?* (Gulbransen, Heggø, Knutsen & Seland, 2015) hadde vi en antakelse (side 16-17) om at mappingrelasjonen =EQ nærmest ville være uaktuell:

«Ifølge ISO-standardene vil ekvivalensrelasjonen være den mest brukte i mappingarbeid. Dette vil antakeligvis gjelde i større grad ved mapping mellom to tesauri – ikke mellom en tesaurus og et klassifikasjonsskjema. I vårt prosjekt ser det foreløpig ut til at exactMatch sjelden er aktuelt, med unntak av eksempler som topptermen 'samfunnsvitenskap' i Humord som kan mappes mot klasse

300 i Dewey. Klassebetegnelsene representerer ofte en klynge av begreper, og dermed vil en exactMatch være uaktuell. Derimot står vi ofte overfor problemstillingen om man skal betegne en relasjon som closeMatch eller broadMatch.»

Antakelsen om at =EQ (exactMatch) ville være nærmest uaktuell viste seg å ikke stemme, da så mange som 25 % av alle mappingene endte opp som =EQ. Årsaken til dette ligger nok i retningslinjene for bruk av =EQ som ble definert i *EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)*. Av begrepene med =EQ-mappinger, er det også interessant at så mye som 40 % har fått mer enn én =EQ-mapping. Kildebegrepene med flest antall =EQ-mappinger er vist i tabell 1. At så mange begreper har fått mer enn én mapping skyldes nok delvis at Dewey ikke er polyhierarkisk, men også delvis at =EQ-relasjonen vår, som følger de nevnte retningslinjene, ikke er veldig streng. Med en streng =EQ-relasjon (transitiv og symmetrisk), ville en jo fra $A =EQ B$ og $A =EQ C$ kunne slutte at $B =EQ C$, noe en ikke generelt kan fra våre mappinger. I bruksområder der dette vil skape problemer, kan imidlertid =EQ-mappingene «nedgraderes» til ~EQ.

Kildebegrep	Antall =EQ-mappinger
Nikkel (REAL012167)	10
Platina (REAL012193)	9
Metaller (REAL002143, HUME08374)	7
Barnelitteratur (REAL030071, HUME07037)	6
Plutonium (REAL012351)	6
Betong (REAL011065, HUME16843)	6
Genetikk (REAL013424)	6

Tabell 1: Kildebegreper med mer enn 5 =EQ-mappinger

Størstedelen av mappingene er BM (42 %), noe som viser at vi har en stor andel begreper i kildevokabularene våre som er snevrere enn klassene i norsk WebDewey.

I prosjektets første fase valgte vi å mappe mot alle faglige plasseringer i WebDewey der begrepet ble funnet. Etterhvert har vi valgt å bruke litt mer skjønn når det gjelder RM-mappinger – blir assosiasjonen for fjern vil mappingene lett kunne oppfattes som støy. Et eksempel er emneordet *Miljø*, som ble mappet tidlig i prosjektperioden (høsten 2015) og som fikk 6 mappinger: 1 =EQ, og 5 RM. I dag ville vi muligens droppet RM til f.eks. 201.77, som ligger under «Religioners holdninger til sosiale spørsmål», i hvert fall for Realfagstermer.

I en del tilfeller har det vært vanskelig å finne rett plassering for et begrep i Dewey. I mange tilfeller skyldes dette at vokabularene våre er mer spesifikke enn Deweys desimalklassifikasjon, og omfatter fagområder som ikke er godt dekket i Dewey-systemet. Den hierarkiske inndelingen av Dewey har også i noen tilfeller vært annerledes enn den etablerte faglige inndelingen, og andre fagområder har manglet helt. For eksempel finnes det ikke noen god plassering for begreper som har med jordas indre prosesser å gjøre, kun klassenumre for jordas indre struktur

og egenskaper. Generelt er Deweys desimalklassifisering mer utbygget når det gjelder strukturgeologi enn prosessgeologi. Et annet eksempel er *Mikrometeorologi*, som har her-note i klassenummer 551.66 Mikroklimatologi. Ifølge geologene på Realfagsbiblioteket, er ikke mikroklimatologi og mikrometeorologi det samme; begrepet hører riktigere hjemme i 551.5 Meteorologi. Vi har her valgt å mappe til begge numre.

Når det gjelder vitenskapelig plassering av biologiske taxa, er dette et område som kan endre seg mye over tid, og inndelingen av Dewey-systemet vil nødvendigvis ligge litt etter oppdatert taksonomi. Der vi ser at vitenskapelig konsensus avviker fra Deweys inndeling, har vi valgt å mappe etter Deweys desimalklassifisering der det finnes tydelig henvisning til taxa: *Akasier* mappes til 583.633 Mimosoideae på grunn av her-note under denne, selv om akasier tilhører familien Fabaceae og dermed burde mappes til 583.63.

Som en konsekvens av mappingarbeidet er store deler av både Humord og Realfagstermer gjennomgått med tanke på rydding og retting i begreper og termer. Begrepenes innhold med synonymer og henvisninger, samt relasjoner mellom begreper, Humords hierarki og Realfagstermers strenger er revidert spesielt der det har vært nødvendig og nyttig for mappingarbeidet, først og fremst for å kunne finne korrekt plassering av et begrep i Dewey-systemet.

Humord

Humord-tesaurusen ble ferdig mappet i siste halvdel av oktober 2018. Da bestod vokabularet av 18 440 begreper. Pr. 15/6-19 består vokabularet av 18 620 termer, altså en økning på 180 begreper. Vi vet at bruttotilveksten har vært på omlag 250 begreper, så vi har i samme periode slått sammen eller slettet rundt 70 begreper. Dette kommer som en følge av at vi enda en gang går gjennom vokabularet fordi det oversettes til engelsk.

Humord er strukturert i faglige hierarkier. I tillegg inneholder tesaurusen geografiske navn, navn på perioder og historiske begivenheter samt navn på skikkelser fra litteratur, mytologi og religion. Hierarkiet «Generelt» består av 450 termer. Noen av disse har vist seg vanskelig å mappe, nettopp fordi de er så generelle i sin natur; innholdsbeskrivende emneord av allmenn karakter, eventuelt generelle emneord slik de er beskrevet i Hjortsæter (2009). De passer over alt i WebDewey. Noen av disse har vi valgt å ikke mappe, og dermed ansett dem for å være «umappelige». Det vil ikke gi noen nytte for sluttbruker å finne plassering i mange Dewey-klasser. Eksempler på begreper vi har ansett som umappelige er: Aktualitet, Deltakelse, Fleksibilitet, Privilegier, Analyse.

Hierarkiet med formtermer; termer over bibliografisk form og manifestasjonsform, er kun mappet dersom de har funnet en naturlig plass i Hjelpetabell 1.

Vi ser at for deler av vokabularet, får vi en opphopning av mappinger. Et eksempel på dette er Dewey-nummer 415 med klassebetegnelse Standardspråkets grammatikk. Her har vi mappet hele 137 Humord, hvorav 127 med BM-relasjon. Dette kan tyde på at Dewey er for lite utbygd, men kanskje også på at Humord er svært godt utbygd på dette området. Vi har gjort utviklerne av Dewey-systemet, OCLC, oppmerksomme på dette med tanke på en mulig utvidelse av tabellen her. Tabell 2 viser Dewey-klassene med flest BM-mappinger.

Dewey-klasse	Antall BM-mappinger
415 Standardspråkets grammatikk	127
T4--5 Standardspråkets grammatikk	116
495.4 Tibetanske og beslektede tibeto-burmesiske språk	65
414 Standardspråkets fonologi og fonetikk	63
T6--954 Tibeto-burmesiske språk	57
292.13 Mytologi--gresk religion	54
T4--15 Fonologi, fonetikk, rettskrivning	50
415.6 Verb	44
T4--55 Substantiver, pronomener, adjektiver, artikler	42
T4--56 Verb	42
948.21 Oslo (Norge : fylke)-historie	42
914.821 Oslo (Norge : fylke)--geografi	42
T2--4821 Oslo fylke	42

Tabell 2: Dewey-klassene med flest BM-mappinger fra Humord.

Humord har et godt utbygd hierarki for geografiske navn og historiske stedsnavn med til sammen 4 000 emneord. Andelen av disse med =EQ er 37,5 %, altså høyere enn i vokabularet som helhet, men det er også en stor andel BM (57,5 %), som vi ser et eksempel på nederst i tabell 2: Humord inneholder 42 stedsnavn i Oslo.

I tillegg til at vi har mappet numre fra hovedtabellen, har vi også mappet til hjelpetabellnumre. Et eksempel på dette er humordtermen Leksika som er mappet til H1-03 Ordbøker, leksika, konkordanser. Tilsvarende er Rogaland mappet til Rogaland fylke i H2-4834. Vi støtte imidlertid på et problem som var knyttet til å bygge hjelpetabellnumre fordi det ikke var mulig å få disse godkjent til nasjonalt bruk. Det betyr at vi ikke får godkjent H2-36834 (Rogaland i oldtiden) per i dag siden dette bygde hjelpetabellnummeret er sammensatt av H2-368 og numrene som følger etter H2-48. Vi har likevel bygget hjelpetabellnumre, men unnlatt å eksportere dem til norsk WebDewey. I påvente av at OCLC forhåpentligvis vil tillate bygde hjelpetabellnumre, finnes disse numrene kun i prosjektets WebDewey-instans. Etter prosjektmedarbeidernes oppfatning er det viktig å kunne benytte ferdig bygde og godkjente hjelpetabellnumre ved nummerbygging, slik at man ikke må starte fra begynnelsen hver gang man skal bygge klassenumre med flere ledd. Vi har tatt dette opp med OCLC og avventer svar. Av samme grunn besluttet vi å mappe til H3 fordi bygde numre innen litteratur og kunst vil være svært nyttige for klassifikatorer, potensielt også i sluttbrukersystemer.

Realfagstermer

Realfagstermer ble ferdig mappet i mai 2019. Vokabularet besto da av 11 200 begreper, etter å ha blitt kontinuerlig oppdatert og ryddet i gjennom hele prosjektperioden. Nedenfor belyses noen utfordringer og problemstillinger knyttet til ferdigstillingen av mappingen av Realfagstermene.

Etter at overlappsvokabularet var mappet, ble det mer komplisert å finne fram til gode plasseringer for termene i Dewey. Dette skyldes en høy grad av fagspesifisitet i Realfagstermer, kombinert med at vokabularet ikke er inndelt hierarkisk. Den faglige betydningen av enkelttermer kan være vanskelig å forstå med mindre man selv er fagperson, som for eksempel for *Sløyfekvantegravitasjon*, *Fikspunkter* eller *Væskeinneslutninger*. Det ble derfor nødvendig å utnytte fagkompetansen til Realfagsbibliotekets fagansvarlige for å slutføre mappingen. Fagansvarlige innen kjemi, informatikk, fysikk, farmasi, matematikk, biologi, astrofysikk og geologi fikk lister med umappede emneord, som de skulle plassere innenfor ett eller flere steder i Dewey-hierarkiet. De kunne foreslå spesifikke Dewey-numre, eller i svært vanskelige tilfeller kun peke til riktig sted i hierarkiet. Dette var til stor hjelp for mapperne, som på grunnlag av forslagene i de fleste tilfeller kunne komme fram til plassering og mappingrelasjon.

Som for Humord, har det også i Realfagstermer utkrystallisert seg områder i Dewey som har fått svært mange BM-mappinger. Tabell 3 viser klassene som har fått flest BM-mappinger fra Realfagstermer, om enn ikke like mange.

Dewey-klasse	Antall BM-mappinger
005.133 Bestemte programmeringsspråk	103
661.895 Organometalliske forbindelser	39
005.1 Programmering	32
004.62 Grensesnittprotokoller og kommunikasjonsprotokoller	30
004.6 Grensesnitt og kommunikasjon	23
523.112 Galakser	22
541.39 Kjemiske reaksjoner	22
005.131 Symbolsk logikk (matematisk logikk)	20
547.7 Makromolekyler og lignende forbindelser	20
516.35 Algebraisk geometri	18
552.5 Sedimentære bergarter	18
512.2 Grupper og gruppeteori	18
579.6 Hattsoyper	18
511.3 Matematisk logikk (symbolsk logikk)	18

Tabell 3: Dewey-klassene med flest BM-mappinger fra Realfagstermer.

I tillegg til de frittstående emneordene, har Realfagstermer nær 15 000 emneord i streng, de fleste automatisk generert i forbindelse med opprettelsen av vokabularet i 2011. Over 14 000 av disse er det vi kaller *virtuelle strenger* – de er delt opp i komponenter på de bibliografiske postene, og fungerer i praksis som forslag til sammensatte søk. Vi har ikke sett behovet for å mappe disse så lenge komponentene mappes. Av de 887 ekte strengene, som alle er av typen emne-emne, har vi mappet 467, f.eks. *Atlanterhavskveite--Embryologi*, som er mappet til det bygde nummeret 571.8617695 Embryologi--kveiter. Antallet mappede strenger er for lavt til å kunne vurdere om prekoordinerte strenger er lettere å mappe enn de enkelte emneord, slik vi spekulerte i et tidligere prosjekt (Gulbrandsen et al., 2015, s. 12). Årsaken til at vi ikke har mappet ferdig alle de 887 strengene er at de har en usikker fremtid i vokabularet.

Realfagstermer har i likhet med Humord hatt en utfordring med enkelte emneord som er såpass generelle at det gir lite mening å mappe dem inn i faglige kontekster. Eksempler på slike er *Interaksjoner*, *Struktur*, *Forekomster*, *Stabilitet*. Et knippe av disse ble betraktet som umappelige, mens andre fikk et avklart faglig omfang gjennom tilføyelse av kvalifikator, definisjon eller omarbeiding til streng.

Relevans for norsk WebDewey

Mappingprosjektet har tilført norsk WebDewey:



- 5604 bygde numre, noe som pr. 25. april 2019 utgjorde 26 % av alle bygde numre som var sendt inn til norsk WebDewey.
- Over 52 000 søkbare termer fra Humord (foretrukne og alternative termer på bokmål og engelsk) knyttet til over 18 600 begreper.
- Over 51 000 søkbare termer fra Realfagstermer (foretrukne og alternative termer på bokmål, nynorsk og engelsk) knyttet til over 11 600 begreper.

For omtrent 70 % av termene i vokabularene fantes det ikke ord-treff i WebDewey. Vi har dermed økt ordtilfanget vesentlig, som gjør at emnesøket i norsk WebDewey er betraktelig forbedret. Andelen bygde numre gjør også WebDewey mer relevant for norske forhold, fordi numrene gjenspeiler fagområdene som finnes i norske fag- og forskningsbibliotek. Emner med norsk relevans er tilført, og vi har også mange særnorske/nordiske begreper som ikke fantes i WebDewey fra før av. Eksempler: *Tvebandstriking*, *Ættestup*, *Bygdebøker*, *Rosemaling*, *Amerikabrev*, *Fossegrimer*, *Tingsteder*.

Ved søk etter begreper i norsk WebDewey ved hjelp av bla-funksjonen, har både Humord og Realfagstermer fått egne søkeinn ganger i form av nedtrekksmenyer, se figur 17. Det kan også søkes særskilt i vokabularene i avansert søk, figur 18. De to neste figurene viser eksempler på hvordan Humord og Realfagstermer vises i norsk WebDewey når en ser på en bestemt Dewey-klasse. Figur 19 viser et eksempel der Humord er godt utbygd og beriker WebDewey med snevrere emneord, mens figur 20 viser et eksempel der Realfagstermer gjør det samme.

Søk Bla i

4 treff Treff per side: Vis side: / 1

Treffliste		
1.	551.792	Pleistocen (istiden)
2.	 560.1792	Istiden--paleontologi, . . .
3.	 561.1192	Istiden--paleobotanikk, . . .
4.	791.4372	Enkeltfilmer

Figur 17: Egne søkeinn ganger i norsk WebDewey for Humord og Realfagstermer.

WebDewey-søk

Søk etter i

i

i

i

i

Behold søk

Figur 18: Søk i vokabularene fra avansert søk i norsk WebDewey.

H2--482331 Fredrikstad + ▾

- [H2-0](#) ▾ Hjelpetabell 2. Geografiske områder, historiske perioder, biografier
- [H2-3-H2-9](#) ▾ Bestemte verdensdeler, stater, lokalområder; himmellegemer utenfor jorda
- [H2-4-H2-9](#) ▾ Nyere tids stater og landområder; himmellegemer utenfor jorda
- [H2-4](#) ▾ Europa
- [H2-48](#) ▾ Skandinavia og Finland
- [H2-482-H2-484](#) ▾ Inndeling av Norge
- [H2-482](#) ▾ Østlandet
- [H2-4823](#) ▾ Østfold fylke
- [H2-48233](#) ▾ Fredrikstad-Sarpsborg-distriktet

H2--482331 Fredrikstad

Bygg nummer + ▾

[H2--482331](#) Fredrikstad START

Noter + ▾

Inkluderer: Borge, Kråkerøy, Onsøy, Rolvsøy

Kommentarer + ▾

Historikk + ▾

Bokbasen emner ✕ + ▾

Registertermer ✕ + ▾

- [Borge \(Østfold\)](#)
- [Fredrikstad \(Østfold\)](#)
- [Kråkerøy \(Østfold\)](#)
- [Onsøy \(Østfold\)](#)
- [Rolvsøy \(Østfold\)](#)

Realfagstermer ✕ + ▾

- [Fredrikstad](#) [bokmål] =EQ
- [Fredrikstad](#) [engelsk]

Humord ✕ + ▾

- [Fredrikstad](#) =EQ
- [Borge \(Østfold\)](#) BM
- [Glemmen](#) BM
- [Gressvik](#) BM
- [Hunn](#) BM
- [Kråkerøy](#) BM
- [Onsøy](#) BM
- [Rolvsøy](#) BM
- [Torsnes](#) BM

Figur 19: Eksempel der Humord er mer finindelt enn Dewey-klassen, mens Realfagstermer ikke er det. Merk at mappingrelasjonene vises i norsk WebDewey sammen med termene. Forklaringer på mappingrelasjonene vises når en holder markøren over begrepet.

681.2 Testeinstrumenter, måleinstrumenter, sensorer

- 600 ▾ Teknologi
- 680 ▾ Produksjon med bestemte bruksområder
- 681 ▾ Presisjonsinstrumenter og andre presisjonsapparater
- 681.2 Testeinstrumenter, måleinstrumenter, sensorer**
- 681.25 Optiske testeinstrumenter, måleinstrumenter, sensorer
- 681.28 Strømningsmålere

Bygg nummer + ▾

681.2 Testeinstrumenter, måleinstrumenter, sensorer START

Noter + ▾

Generell forinndeling kan brukes for hvert enkelt emne eller flere emner behandlet sammen

Inkluderer: Kalorimetre

Her: Testeinstrumenter, måleinstrumenter, sensorer for generell anvendelse innen vitenskap eller teknologi; testeinstrumenter, måleinstrumenter, sensorer for ikke-teknologisk anvendelse; instrumenter for måling av fysiske størrelser; elektriske og elektroniske instrumenter for måling av ikke-elektriske og ikke-elektroniske størrelser

Klassifiser testeinstrumenter, måleinstrumenter, sensorer for et bestemt felt innen vitenskap (unntatt instrumenter for måling av fysiske størrelser) i [681.75](#)

Klassifiser testeinstrumenter, måleinstrumenter, sensorer for en bestemt teknologisk anvendelse med nummeret innen produksjon, f.eks. luftfartøyinstrumenter [629.135](#), utstyr for medisinsk diagnostikk [681.761](#)

Instrumenter for måling av elektriske kvantiteter, se [621.37](#)

Instrumenter for testing og måling av elektroniske signaler, se [621.381548](#)

Instrumenter for måling av tid, se [681.11](#)

Registertermer ✕ + ▾

- [Detektorer \(sensorer\)--produksjonsteknologi](#)
- [Kalorimetre--produksjonsteknologi](#)
- [Måleinstrumenter--produksjonsteknologi](#)
- [Opptaksenheter--fysiske størrelser--produksjonsteknologi](#)
- [Pyrometre--produksjonsteknologi](#)
- [Sensorer--produksjonsteknologi](#)
- [Testeutstyr--produksjonsteknologi](#)

Realfagstermer ✕ + ▾

Sensorer [bokmål]	
Sensorar [nynorsk]	~EQ
Sensors [engelsk]	
Måleinstrumenter [bokmål]	RM
Måleinstrument [nynorsk]	
Halvlederdetektorer [bokmål]	
Halvleiar-detektorar [nynorsk]	BM
Semiconductor detectors [engelsk]	
Kalorimetre [bokmål]	
Kalorimeter [nynorsk]	BM
Calorimeters [engelsk]	
Kjemiske sensorer [bokmål]	
Kjemiske sensorar [nynorsk]	BM
Chemical sensors [engelsk]	
Langmuir-sonder [bokmål]	
Langmuir-sondar [nynorsk]	BM
Langmuir probes [engelsk]	
Mikrosensorer [bokmål]	
Mikrosensorar [nynorsk]	BM
Micro sensors [engelsk]	
Trykkmåling [bokmål]	
Trykkmåling [nynorsk]	BM
Pressure measurement [engelsk]	

Humord ✕ + ▾

- [Måleinstrumenter](#) RM

Figur 20: Eksempel der Realfagstermer er mer fininndelt enn Dewey-klassen, mens Humord ikke er det. Merk at termer vises på flere språk. Dette kommer også til å bli lagt til for Humord etterhvert.

Klassifikatorene ved UBO (og forhåpentligvis også i andre institusjoner) ser at mappingene er nyttige i sitt daglige klassifikasjonsarbeid. Det er enklere å finne korrekt plassering av et emne fordi ordtilfanget har blitt så mye større. Det økte antallet av bygde numre gjør at klassifikatorene slipper å bygge numre selv og kan være sikre på at numrene er riktig bygde.

Personale og økonomi

Følgende personer har vært knyttet til prosjektet i perioden juni 2017–mai 2019:

- Vibeke O. Gustavsen (januar–mai 2019 (50 %), eksterne midler)
- Karoline Hoff (januar 2018–desember 2018 (100%), januar–mai 2019 (40 %), eksterne midler)
- Berit Sonja Hougaard (til august 2017 (30 %))
- Dan Michael Heggø (fra september 2017 (30 %), eksterne midler)
- Unni Knutsen (30 %, prosjektleder)
- Heidi Sjursen Konestabo (10 %)
- Kristine Aalrust Kristoffersen (t.o.m. september 2018 (100 %), eksterne midler)
- Viola Kuldvere (10 %)
- Vibeke Stockinger Lundetræ (t.o.m. desember 2018 (30 %), januar–mai 2019 (50 %))
- Lars Erik Lørdahl (20 %)
- Hege Nenseth (januar-februar 2018, (100 %), eksterne midler)
- Grete Seland (t.o.m. januar 2018 (50 %), eksterne midler)

Grete Seland var formelt knyttet til prosjektet ut januar 2018, men forlot prosjektet i august 2017. Hege Nenseth fra Biblioteksentralen kom inn som erstatter for henne primo 2018.

I tillegg har fagansvarlige ved Realfagsbiblioteket bidratt betydelig i prosjektets siste fase.

På grunn av at vi mottok midler medio 2017, har vi, med tillatelse fra Nasjonalbiblioteket, vært i stand til å holde aktivitetene i gang ut mai 2019. Midlene fra NB har hovedsakelig gått til å lønne de eksternt engasjerte, det vil si Vibeke O. Gustavsen, Karoline Hoff, Dan Michael Heggø, Hege Nenseth, Grete Seland og Kristine Aalrust Kristoffersen.

Noen midler har gått til reisevirksomhet i forbindelse med presentasjoner på konferanser mv.

For fullstendig økonomirapport, se Vedlegg 2.

Formidlingsaktiviteter

Siden prosjektet *Mapping for sluttbrukertjenester* henger nøye sammen med de tidligere prosjektene *På vei mot en generell, norsk tesaurus? : delprosjekt tesaurusmapping* og *Mapping mot norsk WebDewey*, presenteres formidlingsaktivitetene fra disse prosjektene samlet under.

EDUG-presentasjoner:¹¹

- *Still crazy after all these years: DDC mapping in discovery systems* (Dan Michael O. Heggø, Karoline Hoff, Unni Knutsen & Vibeke Stockinger Lundetræ, Stockholm 2019).
- *Mapping for end users* (Karoline Hoff & Unni Knutsen, Bern 2018)
- *Mapping to Dewey: Experiences from the Oslo team* (Unni Knutsen, Paris 2017)
- *Mapping to WebDewey: Knowledge organization in a rough terrain* (Unni Knutsen & Dan Michael Heggø, Göttingen 2016)
- *EDUG's recommendations for mapping: Experiences from use and suggested revisions* (Grete Seland, Göttingen 2016)
- *Maintenance of mappings: Experiences from the Oslo team* (Vibeke Stockinger Lundetræ & Grete Seland, Göttingen 2016)
- *In search of a survival kit for mappers: Abstracting guidelines from complex mapping examples* (Grete Seland, Napoli 2015)
- *The mapping tool ccmapper (concept context mapper)* (Are Gulbrandsen & Dan Michael Heggø, Napoli 2015)

Formidling:¹²

- *Studenters søkeatferd i Oria* (Frokostseminar på Universitetsbiblioteket i Oslo 23. mai 2018. Karoline Hoff, Vibeke Stockinger Lundetræ & Unni Knutsen)
- *Mapping for sluttbrukertjenester* (Presentasjon på Biblioteksentralen 4. april 2018. Kristine Aalrust Kristoffersen)
- *Mapping for sluttbrukertjenester* (Presentasjon på BIBSYS-konferansen 21. mars 2018. Kristine Aalrust Kristoffersen)
- *Kunnskapsorganisering i ulendt terreng* (Pecha kucha på VIRAK 2017, Universitetsbiblioteket i Oslo 13. juni 2017. Kristine Aalrust Kristoffersen)
- *Mapping av emnevokabularene Humord og Realfagstermer til norsk WebDewey* (Presentasjon på VIRAK 2017 under workshop i regi av UBO/NB: «Hva er potensialet i fellesløsninger for emnedata?», Universitetsbiblioteket i Oslo 13. juni 2017. Grete Seland)
- *Mapping mot Dewey* (Presentasjon for ansatte ved UiT, Universitetsbiblioteket 15. mai 2017. Kristine Aalrust Kristoffersen)
- *Mapping mot Dewey* (Presentasjon for ansatte ved Nasjonalbiblioteket 28. februar 2017. Kristine Aalrust Kristoffersen)
- Populærvitenskapelig formidling på bloggen *Statistrikk.no*: «Masker, mapping og en treffende metafor» 20. februar 2017 og «Mer masker og mapping» 21. februar 2017. Grete Seland. Samme tema på *Statistrikketreff* 29. mars 2017 på Realfagsbiblioteket: <https://youtu.be/Jp6FKKAivL4?t=24m> (om mapping fra 24:00 til 32:20)
- *Hvorfor mapper vi?* (Forankring av mappingprosjektet. UiO/HumSam personalmøte, 12. desember 2016. Grete Seland)

¹¹ http://edug.pansoft.de/tiki-index.php?page=Annual_meetings

¹² <https://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-for-sluttbrukertjenester/>

- *Mapping mot Dewey* (Gjesteforelesning for bachelorstudenter i Bibliotek- og informasjonsvitenskap på Høgskolen i Oslo og Akershus. 6. oktober 2016. Kristine Aalrust Kristoffersen)
- *Norsk WebDewey – et verktøy i vekst* (Gjesteblogging på Nasjonalbibliotekets deweyblogg 23. august 2016. Grete Seland)
- *Principles underlying the EDUG recommendations for mapping involving Dewey.* (International Dewey Users Meeting, IFLA WLIC, 16. august 2016. Unni Knutsen)
- *Kunnskapsorganisering i ulendt terreng.* (Frokostseminar på Universitetsbiblioteket i Oslo 12. juni 2015. Grete Seland)
- *Dewey fra menneske til maskin.* (Kunnskapsorganisasjonsdagene 2015, 1. juni, Høgskolen i Oslo og Akershus. Elise Conradi & Unni Knutsen)
- *På randen av mapping* (Artikkel i Bibliotheca Nova 4, 2014, s. 36–46. Unni Knutsen & Are Gulbrandsen)

Trykksaker om mappingaktivitetene:

- Poster
- Brosjyre (finnes også i engelsk oversettelse)
- Flyer

Avslutning og konklusjoner

I søknaden som ble sendt til Nasjonalbiblioteket 16. februar 2017 oppga vi følgende mål:

Hovedmålet for prosjektet er å sørge for at mappede vokabularer kan utnyttes ved emnesøk i sluttbrukerverktøy. Det kunnskapsorganisasjonens potensialet i mappinger vil utforskes fra et sluttbrukerperspektiv. Dette vil legge grunnlag for utvikling av grensesnitt for sluttbrukerverktøy som utnytter mappinger i emnesøk. Dette forutsetter at Humord og Realfagstermer mappes ferdig mot norsk WebDewey.

Prosjektet vil bidra til at sluttbrukere skal få hjelp til å utforske emnebaserte informasjonsbehov på en intuitiv måte.

Gjennom utprøvinger og prototyping vil sluttbrukere få bedre og flere søkeinn ganger til emnedata. Mapping til norsk WebDewey gir mulighet for å utvikle flerspråklige emnesøk i sluttbrukersystemene.

Prosjektet vil føre til at Humord og Realfagstermer blir ferdig mappet til norsk WebDewey. Norsk WebDewey vil få økt funksjonalitet og relevans gjennom økt tilfang til termer og nye registerinnførsler via nummerbygging.

Indekserere vil få hjelp ved tilordning av emneord til en bestemt klasse i Dewey. Flere emnemessige innganger via nummerbygging og innlemmelse av vokabularene vil lette emneords- og klassifikasjonsarbeidet i alle typer bibliotek.

Mappingarbeidet vil ha stor overføringsverdi for andre som ønsker å mappe emnevokabularer til WebDewey ved hjelp av ISO 25964-2.

Mappingene blir tilgjengeliggjort som åpne, lenkede data i et standardisert format (SKOS/RDF) og kan brukes som rådata for tredjepart.

Det vil skje en stor grad av kunnskapsoverføring som konsekvens av prosjektet. Dette gjelder både mot miljøer som vi samarbeider med (ikke minst NB og EDUG) og internt i organisasjonen, og omfatter både kunnskapsorganisasjon og IKT-faglig kompetanse. Prosjektet innebærer også et kontinuerlig formidlingsarbeid i fagmiljøet generelt, knyttet til mappingmetodikk og mappinganvendelser.

Ved prosjektets slutt, mener vi at vi har innfridd alle målene. Begge vokabularer er mappet ferdige. Mappingene foreligger i norsk WebDewey og gir økt termtilfang og hjelp og støtte for de som skal klassifisere eller gi emneord. Mappingene våre er også fritt tilgjengelige for tredjepart. Vi har erfart at det er interesse for å laste ned data.

I løpet av inneværende prosjektperiode har Verbundzentrale des GBV i Göttingen tatt i bruk ccmapper-verktøyet i forbindelse med et prosjekt og tilpasset det sitt behov.

Vi har videre testet hvorvidt mapping mellom vokabularene og Dewey kan gi bedre søkeresultat for sluttbrukere. Vi konkluderer med at det virker lovende å utvide emnesøk til også å inkludere

Dewey-numre som anses som like eller tilnærmet like med termen fra kildevokabularet. Vi vil jobbe videre med Unit/Ex Libris for å få implementert funksjonalitet som understøtter dette ved emnesøk i Oria.

Vi har ervervet mye kunnskap og nye ferdigheter i prosjektet, og vil nå videreføre mapping som en del av de normale aktivitetene knyttet til emneordsarbeid. De kunnskapene og den erfaringen vi har fått, har vi også videreformidlet i stor grad. I etterkant av prosjektslutt, planlegger vi å publisere engelskspråklige artikler om brukerundersøkelsene og verdien av mapping i sluttbrukersystem. Vi er blitt sterkt oppfordret til dette, ikke minst gjennom EDUG-samarbeidet.

Etter prosjektets slutt etableres følgende aktiviteter som permanente oppgaver i biblioteket: oversettelse av termer til engelsk (begge vokabular), mapping av nye termer, oppdateringer av mappingene ved endringer i vokabularene (endringshåndtering), og oppfølging av ccmapper og integrasjonen med WebDewey.

Vi vil gjerne benytte anledningen til å takke Nasjonalbiblioteket for all støtte, både faglig og finansielt.

Litteraturliste

- Antell, K. & Huang, J. (2008). Subject Searching Success: Transaction Logs, Patron Perceptions, and Implications for Library Instruction. *Reference & User Services Quarterly*, 48(1), 68–76. <https://doi.org/10.5860/rusq.48n1.68>
- Asher, A. D., Duke, L. M. & Wilson, S. (2013). Paths of Discovery: Comparing the Search Effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources. *College & Research Libraries*, 74(5), 464–488. <https://doi.org/10.5860/crl-374>
- Bauer, K. & Peterson-Hart, A. (2012). Does faceted display in a library catalog increase use of subject headings? *Library Hi Tech*, 30(2), 347–358. <https://doi.org/10.1108/07378831211240003>
- Cassidy, E. D., Jones, G., McMain, L., Shen, L. & Vieira, S. (2014). Student Searching with EBSCO Discovery: A Usability Study. *Journal of Electronic Resources Librarianship*, 26(1), 17–35. <https://doi.org/10.1080/1941126X.2014.877331>
- Dalal, H. A., Kimura, A. K. & Hofmann, M. A. (2015). Searching in the Wild: Observing Information-Seeking Behavior in a Discovery Tool. *ACRL 2015 Proceedings – March 25–28, 2015, Portland, Oregon*, 668–675.
- David, N., Ian, R., David, C & Peter, W. (2011). Google Generation II: Web behaviour experiments with the BBC. *Aslib Proceedings: New Information Perspectives*, 63(1), 28–45. <https://doi.org/10.1108/00012531111103768>
- Dempsey, M. & Valenti, A. M. (2016). Student Use of Keywords and Limiters in Web-scale Discovery Searching. *The Journal of Academic Librarianship*, 42(3), 200–206. <https://doi.org/10.1016/j.acalib.2016.03.002>
- Drabenstott, K. M. (2003). Do nondomain experts enlist the strategies of domain experts? *Journal of the American Society for Information Science and Technology*, 54(9), 836.
- EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC) (2016). Hentet fra http://edug.pansoft.de/tiki-download_file.php?fileId=146, arkivert på <https://perma.cc/HEP4-MXFV>
- Gulbrandsen, A., Heggø, D. M. O., Knutsen, U., & Seland, G. (2015). *På vei mot en generell norsk tesaurus?: Delprosjekt Metodikk for mapping av Humord mot WebDewey: Rapport*. Hentet fra https://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-for-sluttbrukertjenester/delte-dokumenter/metodikk_mapping_sluttrapport.pdf, arkivert på <https://perma.cc/EMN9-SH7P>
- Heggø, D. M. O., Låberg, K. T., & Lundevall, M. (2016). Scriptotek/emnesøk. *Zenodo*. <https://doi.org/10.5281/zenodo.593528>
- Hjortsæter, E (2009). *Emneordskatalogisering: Innholdsanalyse, emnerepresentasjon og lagring*. Oslo: ABM-media.
- International Organization for Standardization. (2013). *Information and documentation: Thesauri and interoperability with other vocabularies: Part 2: Interoperability with other vocabularies. (ISO 25964-2: 2013)*. Geneve: International Organization for Standardization
- Kirschner, P. A. & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education*, 67, 135–142. <https://doi.org/10.1016/j.tate.2017.06.001>
- Kliwer, G., Monroe-Gulick, A., Gamble, S, & Radio, E. (2016). Using Primo for undergraduate research: A usability study. *Library Hi Tech*, 34(4), 566–584. <https://doi.org/10.1108/LHT-05-2016-0052>

- Kuldvere, V., Lundevall, M., Hegna, K., Konestabo, H. S., Låberg, K. T., Flatby, E. S. & Greenall, R. (2013). *Realfagstermer og TEKORD: RDF som plattform for sammenlikning og sammenføring av emnesystemer?: Rapport*. Hentet fra <https://www.ub.uio.no/om/prosjekter/avsluttet/realfagstermer-tekord/realfagstermer-og-tekord-rapport.pdf>, arkivert på <https://perma.cc/9QU6-PGTG>
- Kuldvere, V., Flatby, E.S., Heggø, D.M.O, Konestabo, H.S., Lundevall, M. & Låberg, K.T. (2014). *Felles terminologi for klassifikasjon med Dewey: Rapport*. Hentet fra <http://urn.nb.no/URN:NBN:no-44610>
- Mahoney, J. & Leach-Murray, S. (2012). Implementation of a Discovery Layer: The Franklin College Experience. *College & Undergraduate Libraries*, 19(2–4), 327–343. <https://doi.org/10.1080/10691316.2012.693435>
- Meadow, K. & Meadow, J. (2012). Search Query Quality and Web-Scale Discovery: A Qualitative and Quantitative Analysis. *College & Undergraduate Libraries*, 19(2–4), 163–175. <https://doi.org/10.1080/10691316.2012.693434>
- Osborne, H. M. & Cox, A. (2015). An investigation into the perceptions of academic librarians and students towards next-generation OPACs and their features. *Program*, 49(1), 23–45.
- Porter, B. (2011). Millennial Undergraduate Research Strategies in Web and Library Information Retrieval Systems. *Journal of Web Librarianship*, 5(4), 267–285. <https://doi.org/10.1080/19322909.2011.623538>
- Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., . . . Tenopir, C. (2008). The Google generation: The information behaviour of the researcher of the future. *Aslib Proceedings*, 60(4), 290–310. <https://doi.org/10.1108/00012530810887953>
- Shiri, A. & Revie, C. (2006). Query expansion behavior within a thesaurus-enhanced search environment: A user-centered evaluation. *Journal of the American Society for Information Science and Technology*, 57(4), 462–478. <https://doi.org/10.1002/asi.20319>

Vedlegg 1 – Søkeoppgaver til observasjonsundersøkelsen

1. Skriv en kort tekst om enten norske eller amerikanske insektforskere.
(Teste: Hvordan håndterer studentene tomme trefflister? Hvordan finner man synonymer og alternative søkeord?)
2. Hvalfangst har hatt stor betydning for den norske kystkulturen og var lenge en betydelig næringsvei. Fortell om hvalen som vilt dyr i Norge.
(Teste: Hvordan håndterer studentene støy? Avgrenser de trefflistene, og i så fall hvordan? Kan de avgrense på emne? Sted?)
3. Gjør rede for lesevaner blant tenåringer generelt, ikke avgrenset til norske forhold.
(Teste: Når søker de på norsk, og når søker de på engelsk? Velger de synonymer?)
4. Beskriv hvilke landpattedyr som lever på Svalbard i dag. Hvilke levekår har de, og hvordan har levekårene endret seg de siste femti år?
(Teste: Hvordan håndterer de vanskelige ord? Beveger de seg i emnehierarkiene? Hvor spesifikt søker de?)

Vedlegg 2 – Økonomirapport

Prosjekt 212066 Mapping for sluttbrukertjenester						
	Regnskap Hittil i Prosjektet	Regnskap Hittil i År	Årsprognose	Fremtidige års prognoser	Levetidsprognose	
Inntekter total	-4 761 232	-273 290	-248 765	0	-4 736 706	
Personal og indirekte kostnader total	4 499 363	679 024	591 911	0	4 412 251	
Drift total (inkludert investering og avskrivning)	403 717	4 202	30 000	0	429 515	
Overført fra ifjor	0	-268 087	-268 087	0	0	
Grand Total	141 848	141 848	105 059	0	105 059	
		Regnskap Hittil i Prosjektet	Regnskap Hittil i År	Årsprognose	Fremtidige års prognoser	Levetidsprognose
Inntekter total	Eksterne inntekter	-2 800 000	0	0	0	-2 800 000
Inntekter total	Egenandel	-1 961 232	-273 290	-248 765	0	-1 936 706
Personal og indirekte kostnader total	Lønnskostnader	2 155 347	382 474	407 164	0	2 180 037
Personal og indirekte kostnader total	Intern overhead	653 683	98 662	86 004	0	641 026
Personal og indirekte kostnader total	Eksternt frikjøp	229 271	19 739	0	0	209 532
Personal og indirekte kostnader total	Diverse lønn, art 5999	488	0	0	0	488
Personal og indirekte kostnader total	Internt frikjøp	1 842 915	181 258	98 743	0	1 760 400
Personal og indirekte kostnader total	Refusjon, NAV	-382 340	-3 108	0	0	-379 232
Drift total (inkludert investering og avskrivning)	Driftskostnader	316 962	0	30 000	0	346 962
Drift total (inkludert investering og avskrivning)	Reise	86 755	4 202	0	0	82 553
Overført fra ifjor	Overført fra ifjor	0	-268 087	-268 087	0	0
Grand Total	Total	141 848	141 848	105 059	0	105 059
	Egenandel:	41,2 %				
Dato: 20.06.2019 Ass.bibliotekdirektør Håvard Kolle Riis						