

Teknisk løsning

Mappingverktøyet ccmapper utvikles av prosjektets tekniske personale i samarbeid med det tyske firmaet Pansoft som står bak utviklingen av WebDewey-verktøyet. Mappingverktøyet vil dermed enkelt kunne brukes av andre mappingmiljøer.

Programvaren skal håndtere vedlikehold av mappinger ved endringer i kilde- og målvokabularene. Vi forutsetter åpen kildekode. Ferdige mappinger publiseres fortløpende som RDF/SKOS og MARC 21 på <http://data.ub.uio.no/>.

Prosjektstøtte

Mappingprosjektet er finansiert fra Nasjonalbiblioteket innen satsningsområdet Nasjonalbibliotekets digitale tjenester som grunnlag for nye tilbud. Prosjektet drives av Universitetsbiblioteket ved UiO i samarbeid med redaksjonen for norsk WebDewey.

Les mer om prosjektet her:

ub.uio.no/om/prosjekter/

Kontaktinformasjon:

Unni Knutsen, HumSam-biblioteket,
Universitetet i Oslo
unni.knutsen@ub.uio.no

Ill.: Colourbox.no

Mapping

mot norsk WebDewey
Kunnskapsorganisering i ulendt terreng



Hvordan oppstod mappingprosjektet?

I 2014 fikk Universitetsbiblioteket ved Universitetet i Oslo (UiO) midler fra Nasjonalbiblioteket til å utvikle metodikk for mapping av emnevokabularer mot Deweys desimalklassifikasjon, som en selvstendig delaktivitet i prosjektet På vei mot en generell norsk tesaurus. Erfaringene vi fikk i dette, og også tidligere mappingprosjekter, førte til at Universitetsbiblioteket ved UiO i 2015 mottok prosjektmidler til å utføre selve mappingen. Prosjektet er toårig.

Hva er mapping?

ISO-standard 25964-2 *Information and documentation: Thesauri and interoperability with other vocabularies* (2013), definerer mapping som en aktivitet hvor det etableres relasjoner mellom begreper i ett vokabular og begreper i et annet. Det vokabularet som danner utgangspunktet for mappingen kalles i ISO-standarden for kildevokabular, mens vokabularet vi mapper mot kalles for målvokabular.

Det er tre hovedtyper mappingrelasjoner (ekvivalens, hierarkisk og assosiativ relasjon) som i vårt prosjekt er realisert ved hjelp av fire ulike relasjonstyper:

- eksakt ekvivalens (kodet =EQ)
- tilnærmet ekvivalens (kodet ~EQ)
- overordnet mapping (kodet BM)
- relatert mapping (kodet RM)

Hvilke vokabularer er involvert?

Vi har to kildevokabularer som mappes mot samme målvokabular, norsk WebDewey:

- Humord, en tesaurus som hovedsakelig dekker humaniora og samfunnsvitenskapene
- Realfagstermer, et emnevokabular (på bokmål og nynorsk) for naturvitenskap, informatikk og matematikk

Hvorfor mapper vi?

Formålet med prosjektet Mapping mot norsk WebDewey er å etablere interoperabilitet mellom vokabularer på tvers av språk og bibliografiske baser. Dette skal sikres i form av en felles verbal inngang til de involverte vokabularene via norsk WebDewey.

De mappede vokabularene skal legge grunnlaget for utbedring av sluttbrukertjenester i form av søk på tvers av samlinger og språk, samt muligheter for orientering i emnebaserte metadata før søk.

Vokabularene vil føre til en berikelse av norsk WebDewey på ulike måter:

- Mappingprosjektet supplerer norsk WebDewey med nye begreper
- Gjennom mappingen tilfører vi synonymer til begreper som allerede finnes
- Mappingarbeidet medfører bygging av nye numre med klassebetegnelser og registertermer
- Mappingen viser begrepets relasjon til klassen, i motsetning til en registerinnførsel, som ikke viser forholdet mellom oppslagstermen og klassen
- Indeksere vil få økt støtte for valg av klassenummer og emneord
- Gjennom analyser av mappingrelasjoner og tilmappede begreper fra kildevokabularene kan klassifikasjonssystemet forbedres og videreutvikles

Hvordan mapper vi?

Sentralt i prosjektet står tanken om mapping som en datastøttet intellektuell prosess. Dataverktøyet gir automatiske forslag til deweyklasser som det kan være aktuelt å mappe kildebegrepene til.

Selve mappingen er en tredelt prosess:

1. Først må betydningen av begrepet (ikke termen som representerer det) avklares.
2. Deretter må vi identifisere hvilket/hvilke klassenumre i WebDewey som representerer dette begrepet. Her er de automatisk genererte forslagene til mappinger basert på tekstlighet et godt hjelpemiddel.
3. Til sist angir vi hvilken relasjonstype som best beskriver graden av samsvar mellom begrepene i kilde- og målvokabular.

Ved prosjektets oppstart fantes det ingen internasjonale retningslinjer for mapping av emnevokabularer til/fra Dewey, selv ikke i ISO-standard 25964-2. Et første viktig skritt i mappingarbeidet var å jobbe aktivt innen EDUG (European DDC Users Group) hvor det finnes et mappingmiljø. Resultatet av denne prosessen, dokumentet *EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)*, ble publisert høsten 2015 (<http://edug.pansoft.de/tiki-index.php?page=EDUG+workshops>).

Relasjonstyper

Vi bruker fire ulike relasjonstyper for å angi graden av samsvar mellom begrepene som mappes

Kildevokabularer

Humord og Realfagstermer

Målvokabular

Norsk WebDewey

Eksakt ekvivalens

Begrepet tilsvarer hele deweyklassen

Transcendentalfilosofi

141.3 Transcendentalisme

Tilnærmet ekvivalens

Begrepet har stort overlapp med deweyklassen

Symmetrigrupper

512.2 Grupper og gruppeteori

Overordnet mapping

Begrepet er snevrere enn deweyklassen

Leksikalsk syntaks

415 Standardspråkets grammatikk

Relatert mapping

Begrepet er assosiert med deweyklassen

Boformer

643.1 Boliger

For di våre vokabularer er inndelt etter emne og målvokabularet WebDewey er inndelt etter fag, vil mange kildebegreper få flere mappinger mot WebDewey, for eksempel:

Industri

338 Produksjon

Industri

322.3 Næringsliv og industri

Industri

343.07 Regulering av økonomisk virksomhet

Industri

333.7965 Energi til industribruk

Acooli språk

496.5 Nilo-saharaspråk

Acooli språk

T6--965 Nilo-saharaspråk

Tistedalsmål

439.8279482321 Dialekter--Halden (Østfold)

Vi mapper både til hovedtabellnumre og hjelpetabellnumre:

Vi bygger numre i WebDewey når det er nødvendig for å uttrykke vårt kildebegrep: