

Prosjektet SV-fag til Humord – Rapport

Målsetninger og rammer

- Målsetninger:
 1. Å integrere frie emneord for samfunnsvitenskapene i tesaurusen Humord og dermed gi HumSam-biblioteket et felles system for emneordsindeksering.
 2. Å gjøre det mulig å søke på felles emneord i hele bibliotekets bestand i Bibsys bibliotekbase.
- Utgangspunkt:
 - Frie emneord brukt på poster med klassifikasjon fra 300-gruppa i Dewey og avgrenset til bib.kode UHS.
 - Psykologi (Dewey 150) faller utenfor prosjektets rammer.
- Avslutning:

Prosjektet bør avsluttes med et konverteringsoppdrag hos BIBSYS hvor relevante poster i BIBSYS bibliotekbase tilføres nye Humord i marcfelt 698. (Dermed sikres samsøk på felles emneord også i eldre bestand, kfr Målsetninger, punkt 2.)
- Arbeidsgrunnlag: En råfil fra Bibsys på ca 270.000 excel-rader
- Tidsramme: Våren 2010 til sommeren 2011

Del 1: Manuell gjennomgang av råfilen

- Fraser ble identifisert og samlet.
- Formtermer, geografiske termer og perioder ble kodet og skilt ut.
- Trykkfeil og rot ble slettet.
- "Omtalte navn", som obligatorisk skal registreres i Bibsys marcfelt 610-640, ble slettet.
- Resultatfil: Ca 190.000 termer (hvorav 186.534 fagtermer)
- Anslått tidsbruk del 1: Ca 300 timer

Del 2: Etablering av fagfiler (til fordeling blant fagreferentene)

- Fagtermene (186.534) ble sortert etter Dewey-koder og deretter oppdelt i 14 fagfiler basert på fagreferentenes ansvarsområder.
- Hver fil ble sortert alfabetisk, emneordene påført frekvens og kodet med fagreferentens initialer. Dublette termer ble slettet.

- Prosjektstans:
7. februar ble prosjektet stoppet av ledelsen. Arbeidet ble tatt opp igjen 11. februar med følgende rammer:
 - Det skal ikke gjøres tilpasninger i hierarkiene i Humord.
 - Nye termer skal midlertidig innplassering (under en knuteterm), men slik at de kan brukes til indeksering og blir søkbare sammen med andre termer i Humord.
 - Det skal ikke arbeides med dette prosjektet etter 15. juli 2011.

Del 3: Gjennomgang av fagfiler

- I samråd med fagreferentene ble det besluttet å slette alle termer med bare en forekomst fra fagfilene.
- De ferdige filene ble fordelt blant fagreferentene for avkryssing av ønskede termer, påføring av kommentarer etc. Tidsfrist: 31. mars.
- Ved retur ble fagfilene - pluss filene med formtermer, geografiske termer og perioder - slått sammen til en alfabetisk fellesfil. Dubletter ble fjernet.
- Fellesfilen ble matchet mot Humord. Avkryssede (ønskede) termer uten Humord-match (altså ikke i Humord fra før) ble sortert ut i en endelig sluttfil med nye termforslag. Omfang: Ca 5.400 termer.

Del 4: Innplassering av nye termer

- Oppstart: 10. mai
- En knuteterm for midlertidig innplassering ble opprettet.
- Termtilfanget var preget av et mangfold som det har vært en stor utfordring å forholde seg til. I arbeidet ble det lagt vekt på å få laget "koplinger" mellom faglig beslektede termer og å få samlet synonymer og nærsynonymer. Det er derfor registrert langt flere se- og se-også-henvisninger enn det som er vanlig praksis. (Gjelder for eksempel mellom entall-/flertallsformer og rettskrivningsvarianter.)
- Det har vært et stort behov for å avklare ords betydning. Dette gjelder ikke bare fagtermer, men også generelle termer. Mange termer (også generelle termer) er derfor utstyrt med definisjoner.
- Kilder i arbeidet:
 - Bokmålsordboka, Store norske leksikon, Wikipedia og emneordslista ved Juridisk bibliotek.
 - Trefflister fra søk i Bibsys' bibliotekbase
 - Humsams Dewey-register (USVD) ble også brukt, men det viste seg at relativt få av de foreslåtte termene var registrert der.
 - På grunn av tidspress ble det liten tid til å bruke faglige oppslagsverk.
- 15. juli ble arbeidet avsluttet:
 - Hele filen med ca 5.400 termforslag var da gjennomgått.
 - Over 4000 nye termer var registrert i Humord (inkl. synonymer registrert som se-henvisninger).

- Av disse fikk 480 termer midlertidig innplassering. De øvrige ble innplassert i eksisterende faghierarkier.
- Anslått tidsbruk del 4: Ca 520 timer(?)

Problemer og mangler

- Stort tidspress og ferietid gjorde at det ble alt for liten tid til diskusjon av faglige problemer med fagreferentene. Arbeidet som er utført med å samle beslektede termer, vil likevel være til stor hjelp i etterarbeidet.
- Deler av hierarkiene i Humord er nå overfylt med termer. Mangler i den eksisterende strukturen er åpenbare.
- Tekniske begrensninger i Bibsys gjør at bare 150 av de midlertidig innplasserte termene blir vist på skjermen. (Kan løses manuelt, men krever arbeid.)
- Mange faglige problemer er oppdaget uten å være løst.
- Det ble ikke tid til korrektur eller avsluttende gjennomgang av arbeidet som er gjort. Sluttproduktet inneholder både feil og mangler. Det vil ta tid før Humord igjen blir en skikkelig tesaurus.
- Det har delvis vært nødvendig å bryte med etablerte indekseringsregler i Humord. Hvordan dette skal løses er ikke avklart.
- Det er behov for en funksjon (hos Bibsys) for gjennomgående retting av utvalgte trefflistes. (Aktuelt behov for alle som arbeider med kontrollerte metadatasystemer, både klassifikasjon, emneord og bibliografiske data.) Dette vil spare mye manuelt arbeid.
- Konverteringsfil mangler:
Det er foreløpig ikke ressurser til å utarbeide en konverteringsfil for oppdatering av eksisterende dokumentposter med nye Humord. Dette var en viktig målsetning ved prosjektstart.

Konklusjon:

- Dette har vært et dugnadsarbeid med rammer på grensen av hva som er forsvarlig. Etterarbeid er helt nødvendig!
- Etter et møte med Nasjonalbiblioteket 5. september er det klart at en ikke kan forvente midler fra NB til etterarbeid og fullføring av UBOs emneordsprosjekter i overskuelig fremtid. Det må derfor gis rom til nødvendig etterarbeid lokalt. Konverteringsfilen bør finansieres av de samarbeidende Humord-bibliotekene i fellesskap.
- Til tross for feil og mangler:
Vi som har hatt ansvar for dette prosjektet, mener likevel at termene i Humord nå kan tas i bruk for indeksering av samfunnsvitenskap (med unntak for psykologi).
- Biprodukter:
I forbindelse med prosjektet er det utarbeidet en rekke excelfiler. Det er også utarbeidet annen dokumentasjon og sakspapirer. Dette er materiale og data som kan være til nytte i videre arbeid med emneord og terminologi, både i forhold til Bibsys og i mer generell

sammenheng. Vi som har arbeidet med dette prosjektet, håper at dette materialet vil bli brukt og arbeidet fulgt opp.

Oslo 2011-09-06

Marit Almo (prosjektansvarlig)

Berit Sonja Hougaard

Kopi: Humord koordineringsgruppe