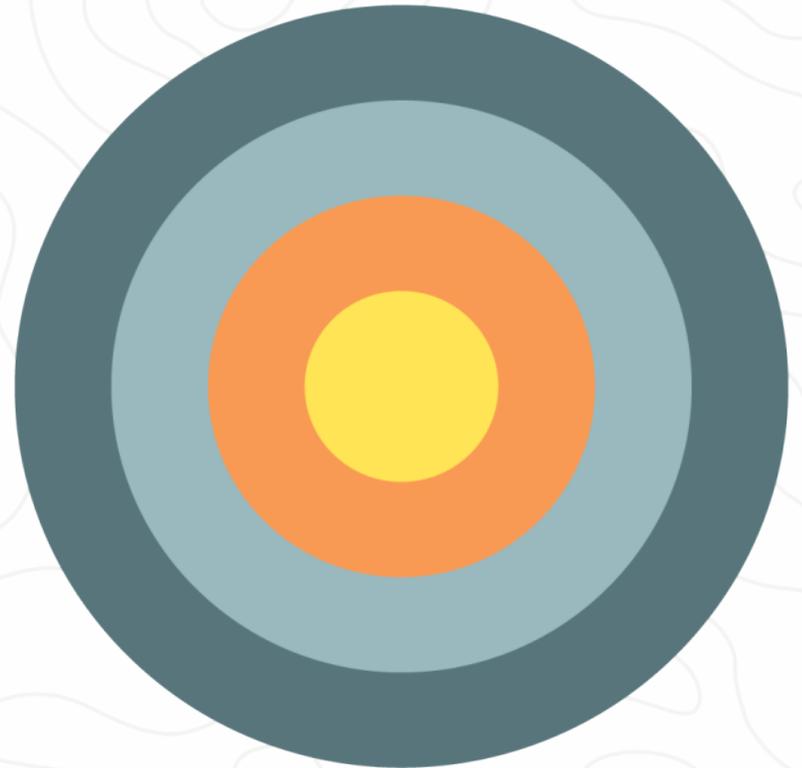


Still crazy after all these years

DDC mapping in discovery systems

Dan Michael O. Heggø, Karoline Hoff,
Unni Knutsen, Vibeke Stockinger Lundetræ



Mapping for end users

EDUG symposium 2019-05-10



Overview

- Why?
- How?
- Findings?
- Have the goals been achieved?
- Unresolved issues?
- The million-dollar question:

Are mappings helpful
for end-users?



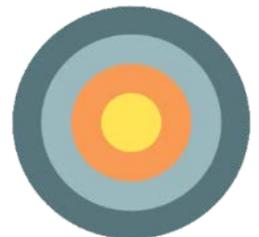
University of Oslo

- Founded 1811
- 27 700 students
- 6 000 employees
- Several sites
- Main library –
Georg Sverdrups
hus, Blindern



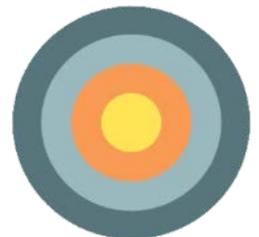
Why mapping?

- Mapping = Establishing relations between concepts in different vocabularies
- Our vocabularies (source vocabularies): Humord (18 500 concepts) and Realfagstermer (15 000 concepts)
- Target vocabulary: WebDewey (Norwegian translation)
- Main assumption: mapping between our two vocabularies and DDC will enhance topical searching in end user systems



ISO-25964-2 (2013)

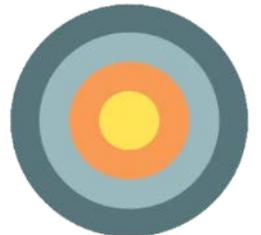
- ISO 25964-2:2013 is applicable to thesauri and other types of vocabularies that are commonly used for information retrieval
- It gives recommendations for the establishment and maintenance of mappings between multiple thesauri, or between thesauri and other types of vocabularies
- Not much about classification systems



Our solution

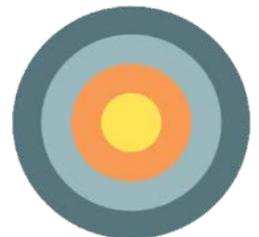


Mapping for end users



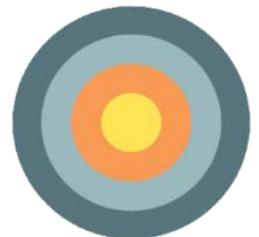
Outcome of workshop in Naples

- EDUG's recommendations for best practice in mapping involving Dewey Decimal Classification (DDC)
- 2015/2016 (last updated during EDUG's annual meeting in Göttingen)



Additional outcome for the library

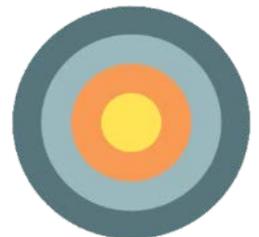
- Policy statement for the project Mapping to Norwegian WebDewey at the University of Oslo Library
 - The scope of the project
 - The choices we made
 - Develop software tool for mapping
 - Also guidelines for actual mappings with examples of use



Relationships

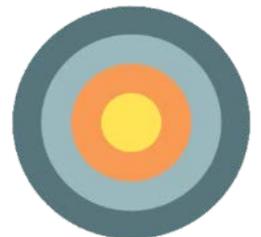
- Three primary relationships relevant to vocabularies: equivalence, hierarchical, and associative
- We decided to use the relationship types:

ISO-25964	SKOS
=EQ	exactMatch
~ EQ	closeMatch
BM	broadMatch
RM	relatedMatch

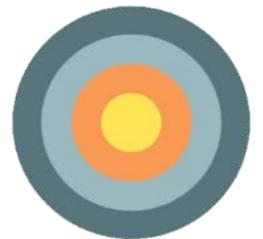


Data assisted mapping tool

- Ccmapper – collaboration between our library and Pansoft
- In use since 2017 and presented at the EDUG Paris meeting

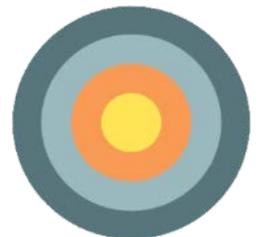


Mapping for end users



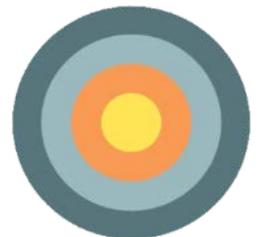
Three main activities

1. Explore how/whether our mappings can be useful in discovery systems
2. Develop a tool/protoype
3. Complete the mappings of Humord and Realfagstermer to WebDewey



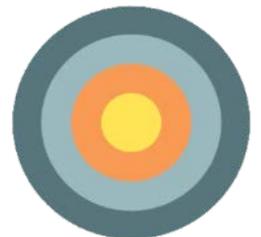
First tracks presented in Bern

- Literature review
- User observations
- Search log analysis



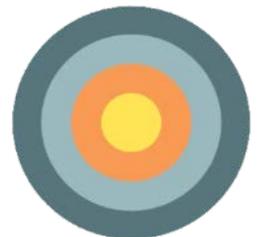
Literature review

- Students tend to use the same 4-5 search words in different combinations, rather than using synonyms
- They normally don't use terms found in subject vocabularies or thesauri
- At first very broad searches, then more specific ones
- The “Google generation” isn't automatically better at searching, or evaluating search results



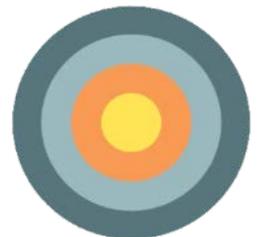
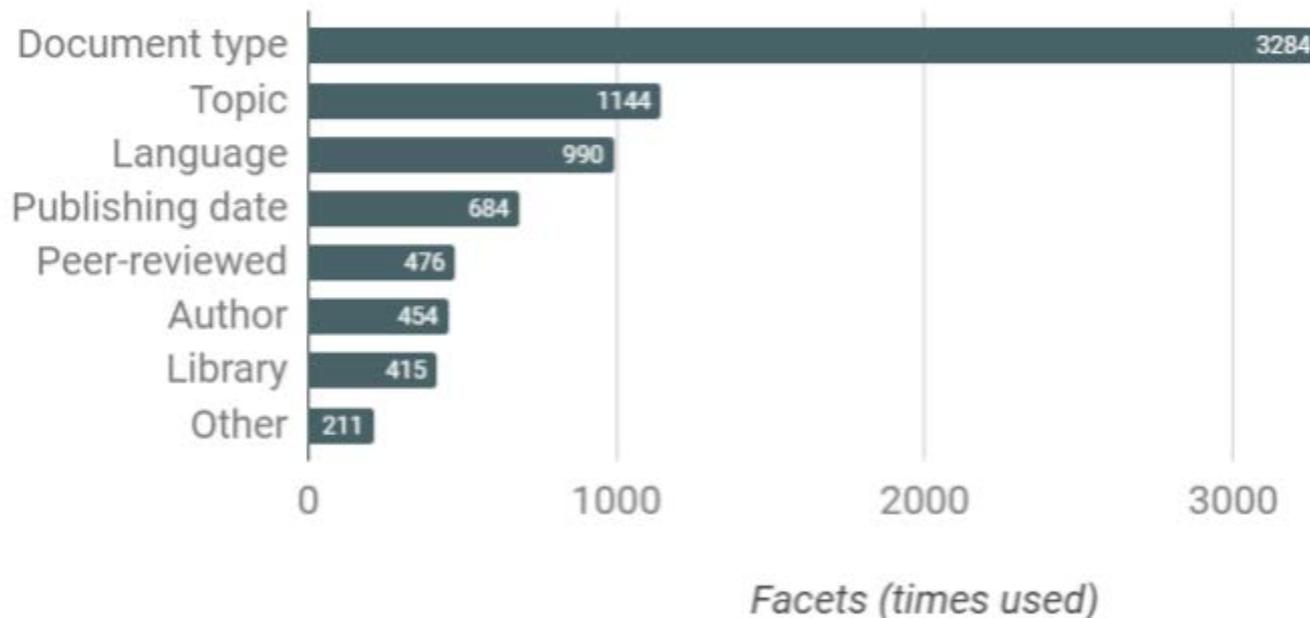
User observations

- The users hardly used synonyms or near synonyms
- Mainly broad searches
- Seemingly no plan or conscious thought behind the choice of search terms or ideas on to how achieve better search results
- Great confidence in the system

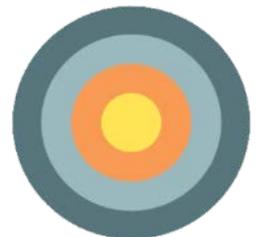
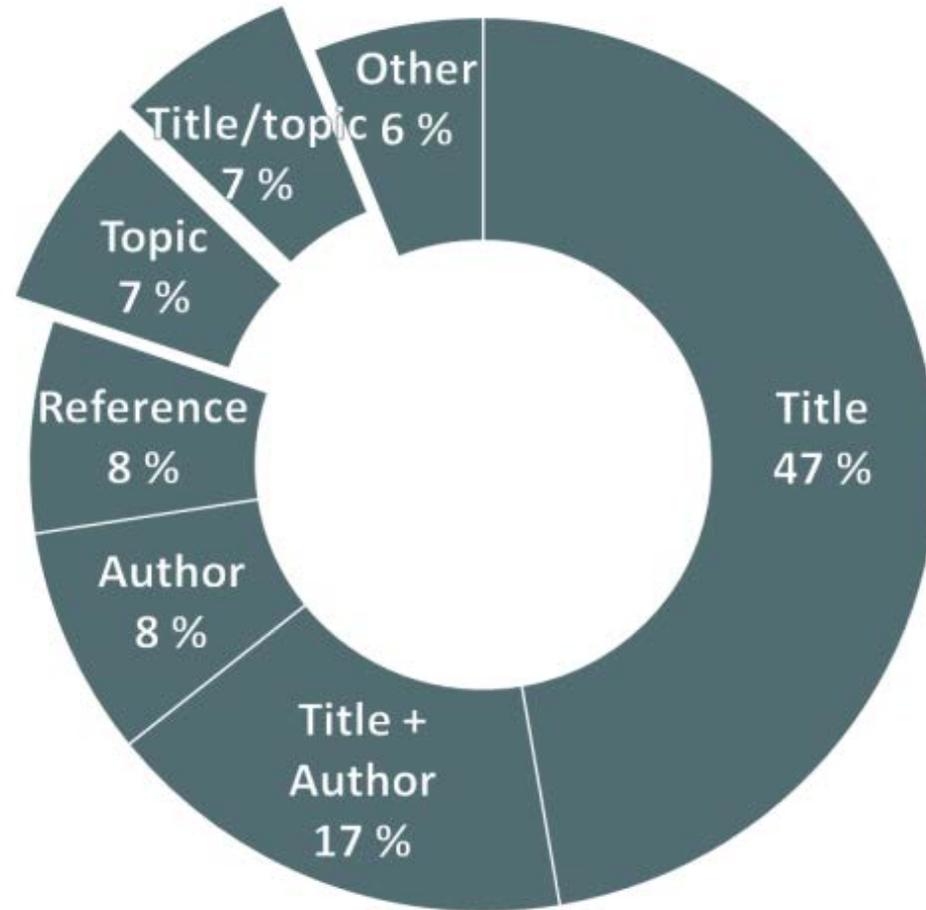


Search log analysis

- 97% of the sessions contained at least one simple search
- 8% contained at least one advanced search
- Scope: 98% «everything»
- Facets were used as limiters in 12% of the sessions

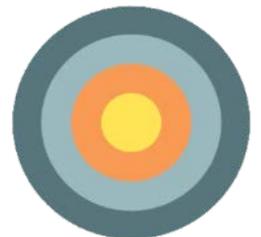


Search log analysis

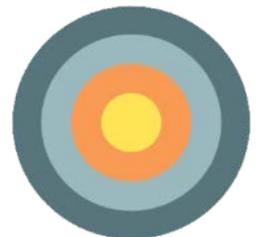
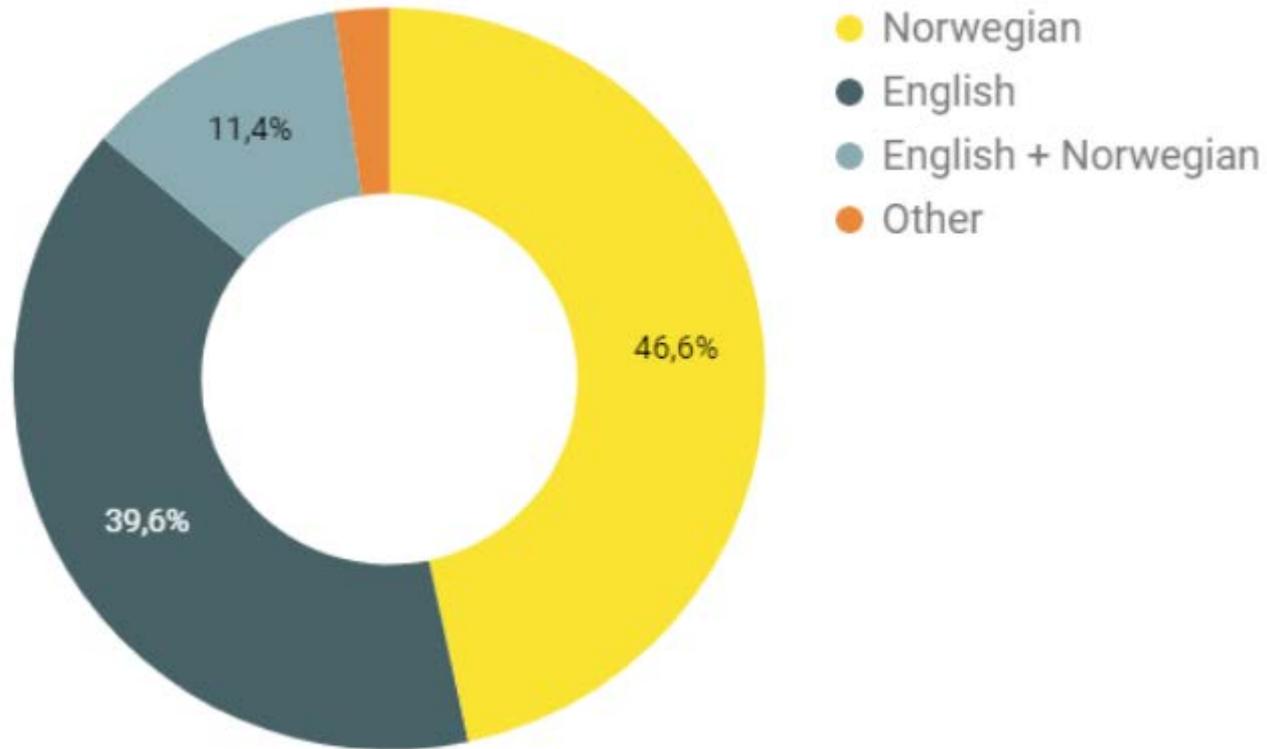


Search log analysis

- 75 % of the searches resulted in at least one opened or viewed record
- 63 % of the searches categorized as “topic” or “title/topic” resulted in at least one opened or viewed record
- Our impression: Our users miss out a lot of relevant documents when doing topical searches



Search language

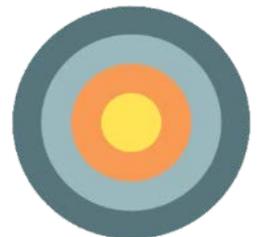


Translation of Humord

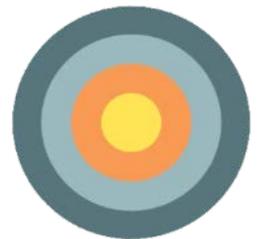
Search terms matching Humord



As of May 1st: 16 000 Humord terms are translated into English



Are mappings helpful?



fisk

X / Everything ▾



PAGE 1 95,692 Results

Personalize


BUK
[Fisk](#)
 Nikolaj Kirk; Christian Stæhr
 2005
 Ikke tilgjengelig fra UiO : Universitetsbiblioteket >


ARTIKKEL FRA OPPSLAGSVERK
[FISK](#)
 24 April 2012
 Encyclopaedia of Islam, Second Edition
 Tilgjengelig online >


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >

SIDE 2


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >


FOTOGRAFI
[Fisk](#)
 Tilgjengelig online >

American Fisheries Society (22)

Vis mer

Bibliotek ^

UIO Realfagsbiblioteket (947)

UIO HumSam-biblioteket (315)

UIO Naturhistorisk museum biblioteket (233)

Vis mer

Emne ^

Humans (9 848)

Medicine (9 455)

Female (6 197)

Vis mer

Dewey ^

597 (471)

639.3 (227)

664.94 (194)

Vis mer

Serie ^

NIVA-rapport (trykt utg.) (121)

NIVA-rapport (74)

Rapport (Universitetet i Oslo. Laboratorium for ferskvannsakonomi og innlandsfiske : trykt utg.) (65)

Vis mer

Tidsskrift/aviser ^

PLoS ONE (1 173)

Physical review letters (595)

Physical Review B (526)

Vis mer

Institusjon ^

NTNU (3 108)

Nasjonalbiblioteket (2 946)

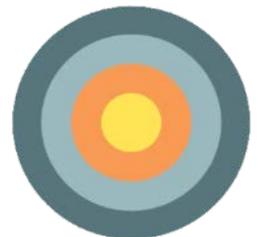
UIT - Norges arktiske universitet (2 100)



SIDE 1

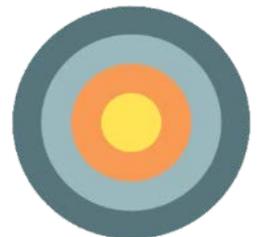
“Black box” discovery system (Primo)

- Difficult working within a closed system like Primo
- We don't fully understand the relevance ranking, it doesn't always seem consistent



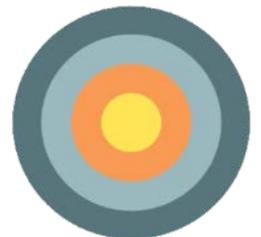
Search query expansion

- Expanding users' search queries with mapped Dewey numbers (=EQ and ~EQ) can result in more relevant documents
- BM and RM mappings could be useful in cases where the user is getting few to no results, perhaps as suggestions to further queries. We have yet to explore this.



Testing tool

Mapping for end users



Free-text search is matched with concepts in Humord/Realfagstermer

Search field: Any field | Query: **Spania** | Instance: UIO | Limit: 20

Without mappings

- any,contains,Spania
- Isr10,exact,936.6
- Isr10,begins_with,936.60*
- Isr10,exact,914.6
- Isr10,begins_with,914.60*
- Isr10,exact,946

13 213 hits after 0.038 seconds (fra cache) [SHOW IN ORIA](#)

book 255 s. ill. 27 cm (Print book@UBO) 1 → X
Spania / Cappelen Damm faktum 2011
 Subjects: Matlagning Spania
 Dewey: 641.5 641.59 641.5946

book 672 s. ill. 2 → X
Spania / Gyldendal fakta cop. 2000
 Subjects: Spania
 Dewey: 914.604
 More editions

book 462 s. ill. 3 → X
Spania / Aschehoug 2007
 Subjects: Spania
 Dewey: 914.604

With mappings

- any,contains,Spania
- Isr10,exact,936.6
- Isr10,begins_with,936.60*
- Isr10,exact,914.6
- Isr10,begins_with,914.60*
- Isr10,exact,946

13 516 hits after 0.047 seconds (fra cache) [SHOW IN ORIA](#)

book IV, 304 s. ill. (Print book@UBO) X → 1
Spania : historien, menneskene og kulturen / 2. utg. Cappelen Damm 2008
 Arnt Stefansen (1953-)\$SQArnt Stefansen (1953-)
 Subjects: Historie Kultur Reisehåndbøker Reiser Spania
 Dewey: 914.6 914.604 946
 More editions

book 377 s., pl. ill. (Print book@UBO) 4 → 2
Spania : historien, kulturen og menneskene / Historie & kultur cop. 2013
 Arnt Stefansen (1953-)\$SQArnt Stefansen (1953-)
 Subjects: Europa Historie Kultur Reiser Samfunn Spania
 Dewey: 914.6 946

book 458 s. (Print book@UBO) X → 3
The new Spaniards / 2nd ed. Penguin c2006
 John Heener\$QJohn Heener

Mapped Dewey numbers added to original query

Spania (humord)

Type: **Mappings**

- exactMatch Spania (Realfagstermer)
- closeMatch 936.6 Den iberiske halvøy fram til 415 og omkringliggende øyer fram til 415 (DDC23)
- closeMatch T2-366 Den iberiske halvøy fram til 415 og omkringliggende øyer fram til 415 (DDC23)
- exactMatch 914.6 Spania-geografi (DDC23)
- exactMatch 946 Spania, Andorra, Gibraltar, Portugal (DDC23)
- exactMatch T2-46 Spania, Andorra, Gibraltar, Portugal (DDC23)

[» Skosmos](#)

Spania (real-fagstermer)

Type: **Mappings**

- exactMatch Spania (Humord)
- closeMatch 936.6 Den iberiske halvøy fram til 415 og omkringliggende øyer fram til 415 (DDC23)

When searching for “Spøkelse” (Ghosts):

Only 2 of the 20 first hits appear in both result lists (without and with mapping expansion)

The screenshot shows a search interface with two columns of results. The left column, titled 'Uten mapper', lists 20 items. The right column, titled 'Med mapper', also lists 20 items. A yellow arrow points from the top of the 'Med mapper' column to the 'Uten mapper' column, and another yellow arrow points from the bottom of the 'Med mapper' column to the 'Uten mapper' column, highlighting that only two items are present in both lists.

Uten mapper

- Spøkelse (1915)

Med mapper

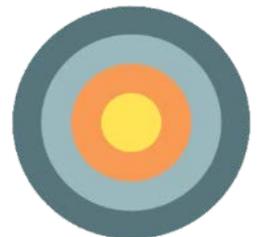
- Spøkelse (1915)

Spøkelse (humord)

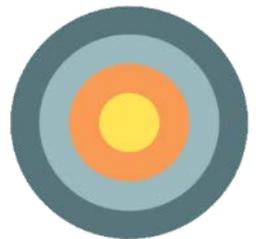
- Spøkelse (1915)
- Spøkelse (1915)
- Spøkelse (1915)
- Spøkelse (1915)

Findings

- Expanding with =EQ and ~EQ numbers seems to provide more relevant search results
- We haven't explored utilization of BM and RM, this remains an area for further research
- We hope to implement this search extension in our discovery system (Primo)



Finishing the mappings



The numbers:

29523

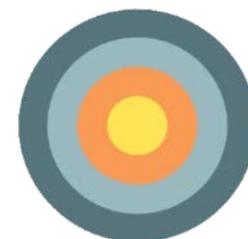
concepts mapped and reviewed; **18581** in Humord, **10942** in Realfagstermer.

65317

reviewed mappings; **16577** exact, **7130** close, **27557** broader, **14399** related.

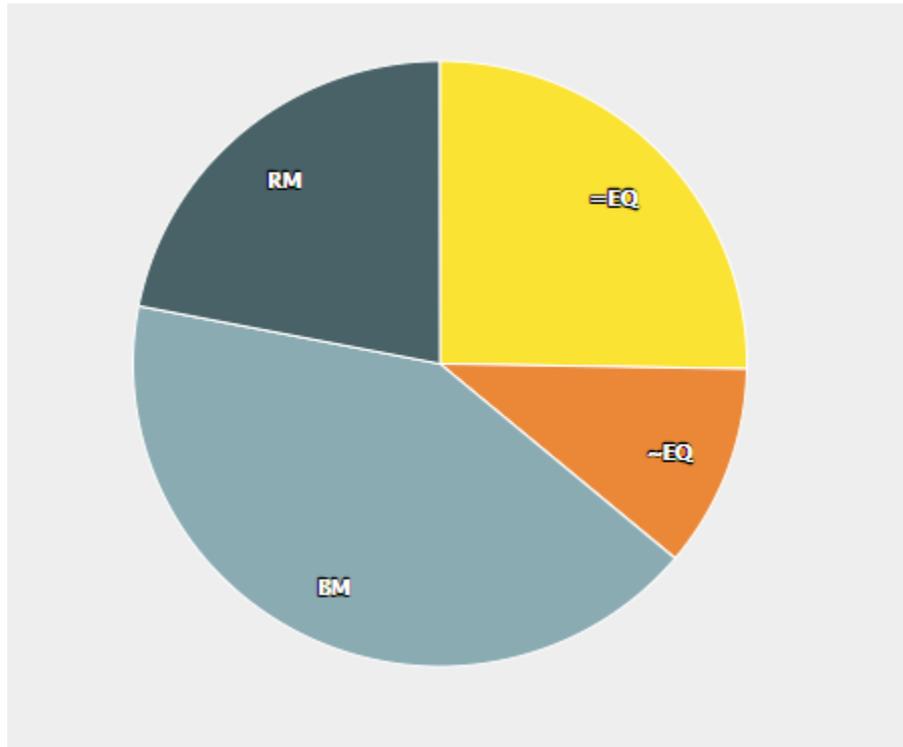
2.21

mappings per concept. **2.27** for Humord, **2.12** for Realfagstermer.

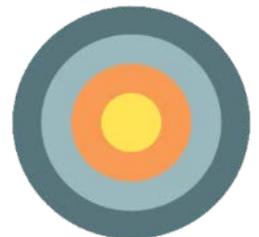


Relationship types

Distribution of relationship types



=EQ: 25%
~EQ: 11%
BM: 42%
RM: 22%



415 Standardspråkets grammatikk

[400](#) ▾ Språk

[410](#) ▾ Lingvistikk

[411-415](#) ▾ Standardspråk--beskrivelse og analyse

415 Standardspråkets grammatikk

Humord



[Formell grammatikk](#)

[Grammatikk](#)

[Grammatisk analyse](#)

[Inversjon \(Lingvistikk\)](#)

[Ordstilling](#)

[Setninger](#)

[Setningstyper](#)

[Syntaks](#)

[Kjenninger](#)

[Språknormering](#)

[Adverbiale leddsetnin](#)

[Agens](#)

[Aktiv \(Lingvistikk\)](#)

[Antipassiv \(Lingvistikk\)](#)

[Argument \(Lingvistikk\)](#)

[Asyndese](#)

[Attributt](#)

[Autoleksikalsk syntak](#)

[Bestemthet \(Lingvistikk\)](#)

[Betingelsessetninger](#)

[Degrammatikalisering](#)

[Demonstrativer](#)

[Deskriptiv grammatikk](#)

[Determinativer](#)

[Diatese \(Lingvistikk\)](#)

[Direkte objekt](#)

[Ekstraposisjon](#)

[Ellipse \(Lingvistikk\)](#)

[Enklitiske ord](#)

[Ergativitet](#)

[Etterstilte spørsmål](#)

[Feltanalyse](#)

[Flytting \(Lingvistikk\)](#)

[Formell syntaks](#)

[Formelt objekt](#)

[Formelt subjekt](#)

[Fraseologi](#)

[Fraser](#)

[Fraserstruktur](#)

[Funksjonell grammatikk](#)

[Funksjonelle kategorier](#)

[Funksjonelt setningsperspektiv](#)

[Funksjonsord](#)

[Følgesetninger](#)

[Føringsbasert grammatikk](#)

[Grammatikalisering](#)

[Grammatiske modeller](#)

[Grensesnitt \(Lingvistikk\)](#)

[Hensiktssetninger](#)

[Historisk syntaks](#)

[Mengdesledd](#)

[Metafraser](#)

[Modularitet \(Grammatikk\)](#)

[Montaguegrammatikk](#)

[Morfosyntaks](#)

[Negasjon](#)

[Neksus](#)

[Nominalisering](#)

[Nominalsetninger](#)

[Nominalsyntaks](#)

[Objekt](#)

[Operatorer \(Lingvistikk\)](#)

[Ordklasser](#)

[Parametre \(Grammatikk\)](#)

[Paratakse](#)

[Parentes \(Lingvistikk\)](#)

[Perifraser](#)

[Polaritet](#)

[Possessiv](#)

[Predikat](#)

[Predikativ](#)

[Preposisjonalobjekt](#)

[Presenteringssetninger](#)

[Redundans](#)

[Regelmessighet \(Lingvistikk\)](#)

[Reksjon](#)

[Rekursjon](#)

[Relativisering](#)

[Relativsetninger](#)

[Resiproke konstruksjoner](#)

[Sammenlikningssetninger](#)

[Semantiske roller](#)

[Setningsekvivalenter](#)

[Setningsemner](#)

[Setningsledd](#)

[Setningsmønstre](#)

[Setningsperspektiv](#)

[Spøkulativ grammatikk](#)

[Spørreord](#)

[Spørresetninger](#)

[Stratifikasjonsgrammatikk](#)

[Subjekt](#)

[Syndese](#)

[Syntagme](#)

[Syntakssentrisme](#)

[Syntaktisk analyse](#)

[Syntaktiske operatorer](#)

[Talemåter](#)

[Tallord](#)

[Tema](#)

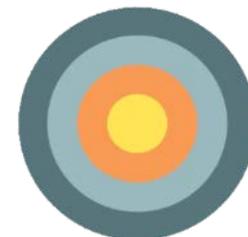
[Tidssetninger](#)

[Tomme kategorier \(Lingvistikk\)](#)

BM

Other achievements

- Mapping completed!
- The mapping is now a part of daily routines
- The quality of our vocabularies has improved
- Our data are available as linked open data (SKOS/RDF) data.ub.uio.no
- The vocabularies are visible in WebDewey



Our vocabularies are visible in WD

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 ✕ + ▾

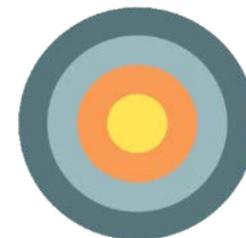
- [Fredrikstad](#) [bokmål] =EQ
- [Fredrikstad](#) [engelsk]

Humord ✕ + ▾

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

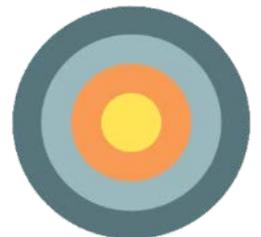
BS-emner (nynorsk) ✕ + ▾

Mapping for end users



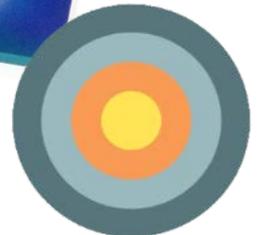
«Power builders»

- 26% of the built numbers (5604) in Norwegian WebDewey are submitted from the mapping project
- Makes the Norwegian WebDewey more «Norwegian»



Stella Dextre Clarke

«And as for the teams developing mappings to Dewey, even when supported by standards, I wish them **fortitude** and **a jar of paracetamol**, as they grapple with the intellectual challenges of mapping to a pre-coordinated scheme.»



Lastly

- Thanks to the National Library of Norway for funding us!
- Contact info:
Contact address: mappinggruppa@ub.uio.no
Homepage: <https://www.ub.uio.no/english/about/projects/mapping-for-end-users/>

