Indexing KOSs in BARTOC by a disciplinary and a phenomenon-based classification: preliminary considerations

Andreas Ledl & Claudio Gnoli, 14.09.2017
Universal Decimal Classification

Link: https://www.udcdata.info/
SKOS File: https://bartoc.org/sites/default/files/skos_files/udcsummary-skos.rdf
English Abstract:
"The Universal Decimal Classification (UDC) is the world's foremost multilingual classification scheme for all fields of knowledge, a sophisticated indexing and retrieval tool. It was adapted by Paul Otlet and Nobel Prize winner Henri La Fontaine from the Decimal Classification of Melvil Dewey and first published (in French) between 1904 and 1907. Since then, it has been extensively revised and developed, and has become a highly flexible and effective system for organizing bibliographic records for all kinds of information in any medium (it is well suited to multi-media information collections). It is structured in such a way that new developments and new fields of knowledge can be readily incorporated. The code itself is independent of any particular language or script (consisting of arabic numerals have appeared in many languages in part in 40 different languages. The EuroVoc abbreviations for short - represents a selection of the EuroVoc Thesaurus of 2011 which contains over 70,000 entries in 24 languages. UDC is updated annually and UDC is a multilingual database that contains a large set of data and can be used as a retrieval tool for information retrieval purposes."

EuroVoc: document indexing

DDC: 001 Knowledge

Author: VLA
VIAF: UDC Consortium
EuroVoc: document indexing
DDC: 001 Knowledge
ILC: general class
Access: Licensed Free
Format: Microform Online Printed SKOS TXT XLS
Wikipedia: http://www.wikidata.org/entity/Q243350

Languages:
- Africans
- Arabic
- Armenian
- Basque
- Belarusian
- Bengali
- Bulgarian
- Catalan
- Chinese
- Croatian
- Czech
- Danish
- Dutch
- Esperanto
- Estonian
- Finnish
- French
- Galician
- Georgian
- German
- Hindi
- Hungarian
- Indonesian
- Irish
- Italian
- Japanese
- Kannada
- Kazakh
- Korean
- Latvian
- Lithuanian
- Malay
- Marathi
- Norwegian
- Polish
- Portuguese
- Punjabi
- Russian
- Sanskrit
- Serbian
- Sindhi
- Slovak
- Slovenian
- Spanish
- Swedish
- Tamil
- Telugu
- Thai
- Turkish
- Ukrainian
- Urdu
- Vietnamese

https://bartoc.org/node/496
Only first three levels of DDC are used

International Classification of Functioning, Disability and Health

The International Classification of Functioning, Disability and Health, known more commonly as ICF, is a classification of health and health-related domains. These domains are classified by body, individual and societal perspectives by means of two lists: a list of body functions and structure, and a list of domains of activity and participation. Since an individual's functioning and disability occurs in a context, the ICF also includes a list of environmental factors.

http://www.who.int/classifications/icf/

Maternal and Child Health Thesaurus

The purpose of developing this thesaurus is to provide the maternal and child health (MCH) professional community with a standard vocabulary in MCH program development and management, including health services, research, training, and program administration. The vocabulary focuses on public health clinical medical terms such as the names of diseases and therapeutic procedures are excluded except when they are germane to public health or MCH.

http://www.mchthesaurus.info/
https://www.unicefcmh.org/mchthesaurus/docs/MCHThesaurus.pdf

Australian Occupational Health and Safety Thesaurus

The scope of AOHOST is to cover concepts relevant to OHS in Australia at national, State and local level. Some concepts are international, but the emphasis is given to the Australian context wherever necessary. The aim is to limit the size of the thesaurus to no more than 3,000 terms. This is to ensure that it is relatively easy to review and keep up-to-date. The terminology currently included in AOHOST is OHS specific. This means that certain areas which are peripheral to the NOHSC's priorities, such as compensation and insurance, are less developed at this stage.

EuroVoc

- 2841 health (17)
- medicine (13)
- nursing care (9)
- public health (9)
- health care (4)
- medical device (4)
- biology (2)
- cancer (2)
- drug addiction (2)
- illness (2)

Show more

DDC

- 610 Medicine and health (44)
- 300 Social sciences (3)
- 570 Biology (3)
- 361 Social problems and services (2)
- 362 Social problems of and services to groups of people (2)
- 600 Technology (Applied sciences) (2)
- 630 Agriculture (2)
- 150 Psychology (1)
- 303 Social processes (1)
- 305 Groups of people (1)

Show more

ILC

- sh: health care (3)
- m30: diseases (2)
- m30:n: cancer (1)
Integrative Levels Classification (ILC)

- Experimental classification system
- Listing phenomena instead of disciplines
- Faceted
- Started in 2004 as an international research project
How does a phenomenon-based classification system perform as compared to a disciplinary one?

No substantial data available …

BARTOC is a good opportunity to index the same items by disciplinary DDC and phenomenon-based ILC:

- It has general coverage (000-999 = a-z)
- Most items are not very specialized
Applying ILC in BARTOC

- ILC is maintained as a MySQL database
- Only the basic database fields for notation, captions and synonyms have been used for this project
- Exported in a csv file (6,456 records)
Taxonomy CSV import/export

Choose the type of import. Help for each type is displayed below when format is selected.

**Important:**
- If you have a tree structure, you should import it before translations, and for each translation if you use a full translate mode.
- If you have a hierarchical vocabulary with duplicate names and fields, you should import them simultaneously.
- Import by “Term ids” uses internal identifiers (“tids”) and allows to update flat or structured vocabularies. It avoids any possible problems with duplicate term names.
- Import by “Fields” allows to import a structure too and can use terms ids (“tids”) too.

### Line format

**term name, child term name, sub-child term name...**

#### Examples (flat)

- Foo, bar, small bar

#### Examples (simple tree)

- Foo, bar, small bar
- Europe, France, Paris
- Europe, France, Lyon
- Europe, UK, London
- America, USA, Washington
- ... Hollywood

#### Examples (Polyhierarchy)

- Foo, Bar, Small Bar
- Grand-Mother, Mother, Daughter
- Grand-Father, Father, Son
- Grand-Mother 2, Father, Son
- Grand-Father 3, Father, Son
- ... Hilton
- Grand-Mother 4, Uncle
- Father, Son 3

The first term is imported as a root level parent, the second as child of first term, the third as child of second term and so on. The lower child is the last term of a line. Others are hierarchical parents.

To repeat previously imported items on each line is not needed. Import process can manage one term by line. If terms are repeated each line and lines are ordered, result and speed are same as with one term by line format.

### Type of structure

- Flat vocabulary
- Simple tree
- Polyhierarchy

Choose if terms have no parent, one parent only, or multiple parents.

#### Keep order of terms

If checked, a weight will be added to each term, so order of lines will be preserved. If unchecked, order will be alphabetic.

**Notes:**

- With polyhierarchical vocabulary (multiple parents for a term) and duplicate terms, unpredictable results may occur, because weight of a term is incremented each time it appears and a term can only have one weight, without context.
- You shouldn’t use this option if your custom format has a ‘weight’ field.
- It’s not recommended to use this option when you update a vocabulary partially.
Reorganizing the hierarchical structure of ILC

ILC contains terms grouped under parent terms. You can reorganize the terms in ILC using their drag-and-drop handles.

Adding a term:

NAME
- U: humanity
- V: general class
- a: forms
  - a: nothing
  - a: individuals
- ac: classes
  - ac: the empty class
- ad: deictics
  - ad: proper names
- ar: relations
  - ar: modality
  - ar: identity
  - ar: similarity
  - ar: difference
  - ar: belong
  - ar: inclusion
  - ar: order
  - ar: neighborhood
  - ar: transformation
  - ar: element
  - ar: component
- ar: premise
- ar: result
- ar: pattern
- ar: property
- af: predications
- af: about
- af: operations
LOD / SKOS

RDFa 1.1 Distiller and Parser

https://www.w3.org/2012/pyRdfa/Overview.html

Integrative Levels Classification

https://bartoc.org/en/node/472/visual
<table>
<thead>
<tr>
<th>TITLE</th>
<th>RATING</th>
<th>DDC</th>
<th>ILC</th>
<th>EUROVOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Classification of Diseases</td>
<td>1,737</td>
<td>616 Diseases</td>
<td>mq10: diseases, sh: health care</td>
<td>Illness</td>
</tr>
<tr>
<td>Universal Decimal Classification</td>
<td>1,723</td>
<td>001 Knowledge</td>
<td>V: general class</td>
<td>document indexing</td>
</tr>
<tr>
<td>International Classification of Primary Care</td>
<td>1,505</td>
<td>610 Medicine and health</td>
<td>sh: health care, mq10: diseases</td>
<td>health care</td>
</tr>
<tr>
<td>Global Agricultural Concept Scheme</td>
<td>1,260</td>
<td>370 Biology, 630 Agriculture</td>
<td>vj: subsistence</td>
<td>agronomy, life sciences</td>
</tr>
<tr>
<td>EuroVoc</td>
<td>1,229</td>
<td>001 Knowledge</td>
<td>V: general class, TUE: European Union</td>
<td>politics, European Union</td>
</tr>
<tr>
<td>Multilingual Classification of European Skills, Competences, Qualifications and Occupations</td>
<td>1,196</td>
<td>331 Labor economics, 370 Education</td>
<td>u9: labour, se: schools</td>
<td>professional qualifications</td>
</tr>
<tr>
<td>General Multilingual Environmental Thesaurus</td>
<td>1,178</td>
<td>577 Ecology</td>
<td>n: populations</td>
<td>ecology</td>
</tr>
<tr>
<td>Corporate bodies Name Authority List</td>
<td>1,175</td>
<td>355 Specific fields of public administration</td>
<td>TUE: European Union</td>
<td>public administration, European Union</td>
</tr>
<tr>
<td>PRODCOM List</td>
<td>1,172</td>
<td>338 Production, 622 Mining and related operations</td>
<td>x: industry</td>
<td>mining production, processing industry</td>
</tr>
<tr>
<td>Events Name Authority List</td>
<td>1,145</td>
<td>328 The legislative process</td>
<td>t6: law, TUE: European Union</td>
<td>legislation</td>
</tr>
<tr>
<td>DOH Metadata Terms</td>
<td>1,126</td>
<td>020 Library and Information sciences</td>
<td>y44oc: online documents</td>
<td>document indexing</td>
</tr>
<tr>
<td>Currencies Name Authority List</td>
<td>1,112</td>
<td>332 Financial economics</td>
<td>u4: currencies</td>
<td>money</td>
</tr>
<tr>
<td>AGROVOC</td>
<td>1,106</td>
<td>630 Agriculture</td>
<td>v: horticulture</td>
<td>agronomy</td>
</tr>
<tr>
<td>Gross Motor Function Classification System</td>
<td>1,106</td>
<td>362 Social problems of and services to groups of people</td>
<td>mq10X5a: nervous diseases, sh: health care</td>
<td>child, young person, physically disabled</td>
</tr>
<tr>
<td>Thesaurus of Clinical Signs</td>
<td>1,101</td>
<td>616 Diseases</td>
<td>mq10: diseases</td>
<td>Illness</td>
</tr>
<tr>
<td>Wikidata</td>
<td>1,100</td>
<td>001 Knowledge</td>
<td>V: general class, y: knowledge</td>
<td>data collection</td>
</tr>
<tr>
<td>Treaties Name Authority List</td>
<td>1,096</td>
<td>541 Law of nations</td>
<td>TUE: European Union, u: economies, t6: law</td>
<td>European treaties</td>
</tr>
<tr>
<td>Learning Resource Exchange Thesaurus</td>
<td>1,090</td>
<td>370 Education, 940 History of Europe</td>
<td>sa: schools, jUe: Europe</td>
<td>educational system, 728: Europe</td>
</tr>
<tr>
<td>Combined Nomenclature</td>
<td>1,084</td>
<td>381 Commerce (Trade)</td>
<td>u5mc: trade, u8: goods and services</td>
<td>industrial production</td>
</tr>
<tr>
<td>Languages Name Authority List</td>
<td>1,076</td>
<td>400 Language</td>
<td>n: languages</td>
<td>language</td>
</tr>
</tbody>
</table>
Integrative Levels Classification (ILC)

ILC® is an innovative classification listing phenomena instead of disciplines (like e.g. in DDC). This page allows to browse top-rated KOSs as classified by ILC1 (1st edition).

Publications
- Course “Overview of Systems for Knowledge Organization” at Hochschule Hannover 2017
- Taxonomy Boot Camp London 2016
- 16th NKOS Workshop 2016
- Proceedings of TKE 2016
- TOTh 2016 Conference
- European DDC Users Group 2016
- International UDC Seminar 2015
- 104th German Library Congress 2015
Results for “j: land”
ILC Search facet

International Standard Classification of Education

“Due to the different education systems across countries and the variation in structure and curricular content, it can be difficult to benchmark performance across countries over time or monitor progress towards national and international goals. In order to understand and properly interpret the inputs, processes, and outcomes of education systems from a global perspective, it is vital to ensure that data are comparable. This can be done by applying the International Standard Classification of Education (ISCED), the standard framework used to categorize and report cross-nationally comparable education statistics.

http://www.uis.unesco.org/Education/Pages/International-standard-classification...

Rate

CSO Standard Fields of Education Classification

“The CSO Standard Fields of Education Classification is a classification of the various fields of study which can be undertaken. This classification was derived from a joint project with Eurostat and UNESCO. It is based on the International Standard Classification of Education - ISCED 1997 (Fields of Education). In 1999, Eurostat and UNESCO further elaborated on the ISCED 1997 and provided a third level of about 80 categories called ‘Detailed Fields’.

http://www.cso.ie/en/surveysandmethodology/classifications/fieldsofeducacioncias...

Rate

Índice de Revistas de Educación Superior e Investigación Educativa Vocabulario Controlado

“El VOCABULARIO CONTROLADO (VC) consiste en una lista normalizada de temas e identificadores. Está integrado alrededor de 1,500 términos sobre diferentes aspectos del área educativa, que han sido clave para desarrollar el IRESIE y el Vocabulario de Documentos de Educación (VOCED).

"The VOCABULARY CONTROLLED (VC) consists of a standardized list of topics and Identifiers. It is composed of about 1,500 terms on different aspects of education, which have been compiled over the 27 years of IRESIE development. Work for the classification and indexing of articles and documents that enter the IRESIE databases and the Index of Educational Documents (VOCED).


EuroVoc

higher education (52)
education (42)
publishing (10)
research (9)
scientific research (9)
document indexing (8)
learning (7)
vocational training (7)
2106 Europe (6)
adult education (6)
Show more

DDC

001 Knowledge (26)
070 Documentary media, educational media, news media; journalism; publishing (22)
994 Australia (7)
344 Labor, social service, education, cultural law (5)
940 History of Europe (5)
993 New Zealand (5)
400 Language (4)
300 Social sciences (3)
800 Literature (Bibliography) and rhetoric (3)
026 Libraries, archives, Information centers devoted to specific subjects (2)
Show more

ILC

se: schools (4)
jUe: Europe (1)
tUe: European Union (1)
tUe: Colombia (1)
u9: labour (1)
xL: literature (1)
At first, only basic ILC classes were imported but it turned out that some concepts expressed as facets are also needed:

- «diseases» is a facet of «organisms»
- «law» is a facet of «governments»

so these were added manually

<table>
<thead>
<tr>
<th>TITLE</th>
<th>RATING</th>
<th>DDC</th>
<th>ILC</th>
<th>EUROVOC</th>
</tr>
</thead>
<tbody>
<tr>
<td>International Classification of Diseases</td>
<td>1,802 points</td>
<td>616 Diseases</td>
<td>mq30: diseases, sh: health care</td>
<td>illness</td>
</tr>
<tr>
<td>Universal Decimal Classification</td>
<td>1,784 points</td>
<td>001 Knowledge</td>
<td>V: general class</td>
<td>document indexing</td>
</tr>
<tr>
<td>International Classification of Primary Care</td>
<td>1,558 points</td>
<td>610 Medicine and health</td>
<td>sh: health care, mq30: diseases</td>
<td>health care</td>
</tr>
<tr>
<td>Global Agricultural Concept Scheme</td>
<td>1,298 points</td>
<td>570 Biology, 630 Agriculture</td>
<td>vj: subsistence</td>
<td>agronomy, life sciences</td>
</tr>
<tr>
<td>Events Name Authority List</td>
<td>1,256 points</td>
<td>328 The legislative process</td>
<td>t6: law, tUE: European Union</td>
<td>legislation</td>
</tr>
</tbody>
</table>
Examples

Thesaurus of European Education Systems

DDC: 379 940 ‘Public policy issues in education; History of Europe’
ILC: se tUE ‘schools; European Union’

CERL Thesaurus

DDC: 094 911 940 ‘Printed books; Historical geography; History of Europe’
ILC: js wl r55n yu ‘landforms; settlements; noun; humanities’
Preliminary results

- Free combination of ILC or DDC classes covers subjects with similar accuracy
- Trees of classes are arranged differently:
  - ILC scatters disciplines
  - DDC scatters phenomena
Thank you for your attention.