The Cocoda Mapping Tool

U. Balakrishnan, J. Voß
Verbundzentrale des GBV (VZG)
Overview

- Background of Project coli-conc
- Focus of the Project
- Issues related to DDC-RVK Mapping
- Methods of Mapping
- Course Correction
- Introduction to the Software
  - Demands on the Tool
  - Web Layout
  - Software Concept
### Universal Classification Systems

<table>
<thead>
<tr>
<th>Classification System</th>
<th>No. of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>UDC (Universal Decimal Classification)</td>
<td>ca. 65,000 classes (English version)</td>
</tr>
<tr>
<td>DDC (Dewey Decimal Classification)</td>
<td>over 44,000 classes with 10 main classes</td>
</tr>
<tr>
<td>RVK (Regensburg Classification)</td>
<td>850,000 classes with 33 main classes</td>
</tr>
<tr>
<td>BC (Basic Classification)</td>
<td>2,100 classes with 89 main classes</td>
</tr>
<tr>
<td>LCC (Library of Congress Classification)</td>
<td>21 main classes</td>
</tr>
</tbody>
</table>

### Subject Classification

<table>
<thead>
<tr>
<th>Classification System</th>
<th>No. of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDC-Sachgruppen der DNB</td>
<td>10 main classes with 94 subclasses</td>
</tr>
<tr>
<td>MSC (Mathematics Subject Classification)</td>
<td>87 main classes</td>
</tr>
<tr>
<td>PACS (Physics and Astronomy Classification Scheme)</td>
<td>10 main classes</td>
</tr>
<tr>
<td>FKDigBib (Subject classification for digital library)</td>
<td>10 main classes</td>
</tr>
<tr>
<td>KfM (Classification for music library)</td>
<td>ca. 800 classes</td>
</tr>
</tbody>
</table>

### Subject Classification at the Universities

<table>
<thead>
<tr>
<th>Classification System</th>
<th>No. of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>TUM-classification (Science and technology classification of the TU Munich)</td>
<td>52 classes each with 999 notations</td>
</tr>
<tr>
<td>Subject classification of the University library Duesseldorf</td>
<td>45 classes</td>
</tr>
<tr>
<td>Bremer classification of the State and University library Bremen</td>
<td>ca. 57 main classes</td>
</tr>
<tr>
<td>GOK (Goettingen Online Classification)</td>
<td>ca. 33 main classes</td>
</tr>
<tr>
<td>Standard-Thesaurus Wirtschaft von der ZWB</td>
<td>6,000 Terms and Notations</td>
</tr>
<tr>
<td>Subject classification University library Trier</td>
<td>36 main classes</td>
</tr>
<tr>
<td>Technical University Dortmund</td>
<td>28 main classes</td>
</tr>
<tr>
<td>University library Paderborn</td>
<td>26 main classes</td>
</tr>
<tr>
<td>University library Marburg</td>
<td>35 main classes</td>
</tr>
<tr>
<td>University library Bonn</td>
<td>24 main classes</td>
</tr>
<tr>
<td>University library Heidelberg</td>
<td>22 main classes</td>
</tr>
<tr>
<td>Subject classification and nomenclature of individual languages Library of the Institute of General Linguistics at the Uni Münster</td>
<td>23 main classes</td>
</tr>
</tbody>
</table>

### Subject Classification at the public libraries

<table>
<thead>
<tr>
<th>Classification System</th>
<th>No. of classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>SEB (Scheme for protestant libraries)</td>
<td></td>
</tr>
<tr>
<td>SKB-E (Scheme for catholic public libraries)</td>
<td></td>
</tr>
<tr>
<td>KFKJ (Scheme for children and youth libraries)</td>
<td>Less than 1,000 classes</td>
</tr>
<tr>
<td>ASB (General classification for public libraries)</td>
<td>ca. 2,200 classes with 23 main classes</td>
</tr>
<tr>
<td>ÖÖSB (Austrian classification for public libraries)</td>
<td></td>
</tr>
<tr>
<td>SFB (Classification for libraries)</td>
<td>ca. 14,400 classes with 30 main classes</td>
</tr>
<tr>
<td>KAB (Classification for general libraries)</td>
<td>ca. 2,700 classes</td>
</tr>
<tr>
<td>SSD (Classification of the city library Duisburg)</td>
<td></td>
</tr>
<tr>
<td>ESSB (Single classification for South Tyrolean)</td>
<td>16 main classes</td>
</tr>
</tbody>
</table>
Primary Source and Target Schemes: DDC and RVK

Why RVK?
- wide-spread in Germany
- Local needs are better covered
- Legacy data transfer
- DDC is subject to licence

RVK
- 850,000 classes
- 33 main classes
- Granularity varies in different subject fields
- Synthesized notations are prebuilt and integrated into the online system

DDC
- ca. 46,000 classes
- 10 main classes
- not all synthesized notations/numbers are represented in the online system
Structural difference DDC : RVK
Notational Building Sequence „Medicine & health”

614.4273
Diseases--humans--incidence—United States, ...

614.42
Incidence (from Main Schedule)

73
United States (from Table T2)

RU 10585
Spreading of diseases and influence of the geographical environment (Geomedicine)

R
Geography

RU 10000
USA

585
Subject index for Geography -S1R
Spreading of diseases and influence of the geographical environment (Geomedicine)
Conceptual differences (1)

Example:
„Die Europäische Gemeinschaft als Rechtsgemeinschaft“. 
_Hrsg. von Wolfram Blomeyer_

Example:
„Grundstrukturen des Rechts öffentlicher Dienstleistungen“. 
_Autor: Krajewski, Markus_

Example:
„Jurisdiktionskonflikte bei grenzüberschreitender Kriminalität“ 
_Hrsg. von Arndt Sinn_

Example:
„Soziale Inklusion im europäisch-vergleichenden Zusammenhang aus Jurist. “
_Autor: Schulte Bernd_

Example:
„Die Instrumentalisierung des Privatrechts durch Europäische Union“. 
_Autor: Schmidt, Christoph U_

**P Law**

Europarecht; Recht der Europäischen Gemeinschaften bzw. der Europäischen Union und der sonstigen europäischen Organisationen; Europäische Rechtsvereinheitlichung

**PS**

DDC 340 Law

341 Law of nations

342 Constitutional – and administrative law

343 Military, defence, ...

344 Labor, social service, education, ...

346 Private law
Conceptual differences (2)

615
Pharmacology and therapeutics

610
Medicine and health

616.024
Domestic medicine

Example: „Selbstmedikation für die Kitteltasche,...“
Author: Lennecke, Kirsten
RVK notation: VR 5800
DDC notation: 616.024

Example: „Aspirin and other Salicylates“
Author: Vane, John, R,...
RVK notation: XI 4000, VW 5100
DDC notation: 615.783

Example: „Toxicology in the use, misuse and abuse of food, drugs and chemicals“
Author: Chambers, Philipp L,...
RVK notation: VT 5308
DDC notation: 615.9

Example: „Essentials of Pharmacology“
Author: Oldham, Frances, k.
RVK notation: XI 1400
DDC notation: 615

RVK: WW-XV
Medicine

XI
Pharmacology and Toxicology

RVK: V
Chemistry and Pharmacy

VT
Pharmacology and Toxicology

VW
Pharmaceutical Biology

VR
General Pharmacy

www.gbv.de
Percentage of DDC and RVK in the Union Catalogues

**GVK**
- Ca. 40 Mio. Title data records (2013)
- DDC: 19.8%
- RVK: 3.7%

**SWB**
- Ca. 17 Mio. Title data records (2013)
- DDC: 14.4%
- RVK: 24.3%

**BVB**
- Ca. 20 Mio. Title data records (2013)
- DDC: 17.7%
- RVK: 29.2%
Survey

- Current status of DDC-X-concordance
- Field of application and the reasons for the use of DDC
- Methods & Problems in building a DDC-X concordance
- Interest in a DDC - RVK concordance

<table>
<thead>
<tr>
<th>Existing Mapping works</th>
<th>Concordance</th>
<th>Subject area</th>
<th>Contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>DDC – BK</td>
<td>Chemistry</td>
<td>TUB TUHH</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Politics</td>
<td>SUB Hamburg</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The thousand classes of the third summary</td>
<td>VZG</td>
<td></td>
</tr>
<tr>
<td>DDC – EZB</td>
<td>41 EZB-Fachgruppen</td>
<td>VZG</td>
<td></td>
</tr>
<tr>
<td>DDC – RVK</td>
<td>Library- and Information science</td>
<td>HdM Stuttgart</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Social science</td>
<td>UB Greifswald</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medicine &amp; Health, Law, the thousand classes of the third summary level</td>
<td>VZG</td>
<td></td>
</tr>
<tr>
<td>RVK – DDC</td>
<td>Biology, Chemistry, Geology, Paleontology, Physics, Mathematics</td>
<td>GESIS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Psychology</td>
<td>SLUB Dresden</td>
<td></td>
</tr>
<tr>
<td>RVK – BK</td>
<td>German literature, Politics, Law</td>
<td>UB Wien</td>
<td></td>
</tr>
<tr>
<td>RVK – MSC</td>
<td>Mathematics</td>
<td>UB Regensburg</td>
<td></td>
</tr>
<tr>
<td>RVK – PACS</td>
<td>Physics</td>
<td>UB Regensburg</td>
<td></td>
</tr>
<tr>
<td>SWD – DDC</td>
<td>Library- and Information science</td>
<td>DNB</td>
<td></td>
</tr>
<tr>
<td>SWD-RVK</td>
<td>Library- and Information science</td>
<td>HdM Stuttgart</td>
<td></td>
</tr>
<tr>
<td>RVK-BK-MSC-PACS</td>
<td>Mathematics, Physics</td>
<td>ULB Tirol</td>
<td></td>
</tr>
<tr>
<td>DDC-MSC-BKL</td>
<td>Mathematics</td>
<td>TIB Hannover</td>
<td></td>
</tr>
</tbody>
</table>
Mapping Methods

**Work done so far:**

Complete concordance
- DDC - EZB
- DDC - BK for the thousand classes of the third summary of the DDC
- DDC - RVK for the thousand classes of the third summary of the DDC
- DDC - RVK for the DDC subject area Medicine & Health
- DDC - RVK for the DDC subject area „Law“

Partial Concordance
- DDC - RVK for the DDC subject area „Philosophy“ (ca. 14% of the current DDC-classes)

**Statistical Inference**

- Title data records
  - Catalogues and databases e.g. GVK, SWB
Facilitate exchange and use of concordances and KOS
- Collection of the existing mappings and KOS
- Provision of the above

Enhance the speed of building concordances between library KOS and ease their management
- Develop a mapping tool
- Make the concordances and KOS easily accessible
- Draft algorithms for automatic generation of mapping candidates

Improve the quality of the concordances
- Develop and implement measures for quality control
- Involve and expand the user groups
Demands on the tool

- Allow validation and storage of data
- Integration of Data from different sources
- Presentation of Data and mapping candidates on a single screen
  - Clear overview of the context of the selected term through display of
    - the hierarchical structure of the classes
    - scope notes
    - Register Index Entries
    - linked vocabularies and synonym suggestions
  - Mapping suggestions through
    - evaluation of the co-occurrences of assigned notations/terms in the title data records
    - automatic generation of mappings
    - integration of the concordance database
    - inclusion of the results of a manual mapping

Multi-user web based open source tool
- Easy access to and exchange of information
- Serve as collaboration platform
Based on SKOS extended with mapping objects

Application of JSON-LD with consistent fields
⇒ simple JSON with optional mapping to RDF

Also supports existence statements
- ∄: negation (e.g. has no related concepts)
- ∀: completeness (e.g. has narrower concepts)

Open Standardization of JSKOS, feedback appreciated: [https://gbv.github.io/jskos/](https://gbv.github.io/jskos/)
JSKOS Example: Concept

```json
{
    "@context": "http://gbv.github.io/jskos/context.json",
    "type": ["http://www.w3.org/2004/02/skos/core#Concept"],
    "inScheme": ["http://dewey.info/scheme/edition/e22/"],
    "uri": "http://dewey.info/class/612.112/e22/",
    "notation": ["612.112"],
    "prefLabel": {
        "en": "White corpuscles (Leukocytes)",
        "de": "Leukozyten (Weiße Blutkörperchen)"
    },
    "narrower": [null],
    "broader": [ {
        "notation": ["612.11"],
        "uri": "http://dewey.info/class/612.11/e22/
    } ],
    "related": null
}
```

Missing fields: maybe (Open World Assumption)
Null value: negation
Null as array element: existence
Based on JSKOS

Used in Cocoda with Web Components

Simple URL queries, possible mapping to LOD-URIs
- /mappings?from=DDC&to=RVK&notation=612
- /concepts?notation=612 ⇔ /concepts/612

Wrapping other APIs

Open preview of specification and prototype
- https://gbv.github.io/jskos-api/
- https://github.com/gbv/cocoda-db
- http://coli-conc.gbv.de/concordances/
JSKOS Example: Mapping (current draft)

```json
{
    "from": {
        "inScheme": ["http://dewey.info/scheme/edition/e22/"],
        "conceptSet": [
            {
                "uri": "http://dewey.info/class/612.112/e22/",
                "notation": ["612.112"]
            }
        ]
    },
    "to": {
        "inScheme": ["http://d-nb.info/gnd/7749153-1"],
        "conceptSet": [
            {
                "uri": "http://d-nb.info/gnd/4074195-3",
                "preflabel": { "de": "Leukozyt" }
            }
        ]
    },
    "mappingType": "closeMatch",
    "creator": "VZG"
}
```
Thank You!

http://coli-conc.gbv.de/

Vectors slide no.11: © Vallepu – fotolia.com https://de.fotolia.com/
Vectors slide no.9: © NLshop– fotolia.com https://de.fotolia.com/
Vectors slide no. 2,3,4,10,12,18: designed by Freepik.com http://www.freepik.com
Thanks to Jana Agne for creating the table at the slide no.3