

Project Coli-conc: Mapping Library Knowledge Organisation Systems

Uma Balakrishnan, Morsheda Akter.
Verbundzentrale der GBV, Germany

Abstract: Knowledge organisation systems (KOS) across libraries in German speaking countries have always struggled with a lack of homogeneity and consistent subject indexing. Various tools have been developed to map different KOS, either manually or through algorithms which allow for automatic or semi-automatic mapping^{1, 2, 3}. Considering the complexity and variety of KOS, there is a need for an overarching system which facilitates and promotes consistent subject indexing and enhances access to information and use of KOS. The first step towards designing and deploying such a system was the conversion of various KOS data from their proprietary formats to a uniform format. To meet this requirement, Coli-conc developed in the pilot phase of the project the JSKOS format⁴. The newly created format is based on SKOS and JSON LD; it eases the use of KOS in web applications and allows more flexibility in content display of the KOS data. To accelerate and increase the efficiency of the intellectual mapping process, the project proposes the mapping tool *Cocoda* that adopts a dashboard design. This approach gathers data of the KOS that are being mapped with each other from different sources, consolidates the same and displays it in a concise manner on a single screen for at-a glance-monitoring. The web-based mapping tool, *Cocoda*, besides enabling term/caption or notation search from a source or a target KOS; presents the hierarchical structure of a queried term/class; and permits browsing of the higher order concepts of the selected KOS. Additionally, the tool automatically generates mappings of a selected term/notation, gives options to edit and save the newly created mappings in the centralized VZG Mapping database integrating a feature to export them in JSKOS and various other formats. Furthermore, the Coli-conc infrastructure includes a KOS-Registry, a platform for Concordances with a web interface, JSKOS API for KOS and mappings which are provided as stand-alone services.

¹ Walter, A.-K., Mayr, P., Petras, V., Baerisch, S. (2007)

² Pfeffer, M. (2013)

³ Lauser, B., Johannsen, G., Caracciolo, C., Keizer, J., van Hage, W. R., Mayr, P. (2008)

⁴ Voß, J. (2016)

Features of the Coli-conc and technical specifications

1. Coli-conc System Architecture

The system offers an infrastructure to support semi-automatic creation of mappings. Furthermore, it facilitates the use and exchange of KOS and their mappings. The application is modular designed and consists of three core parts:

- Coli-conc Web (CCWeb)
- Data Converter
- JSKOS API for KOS and Mappings (KOS-API, KK-API)
- Database Server

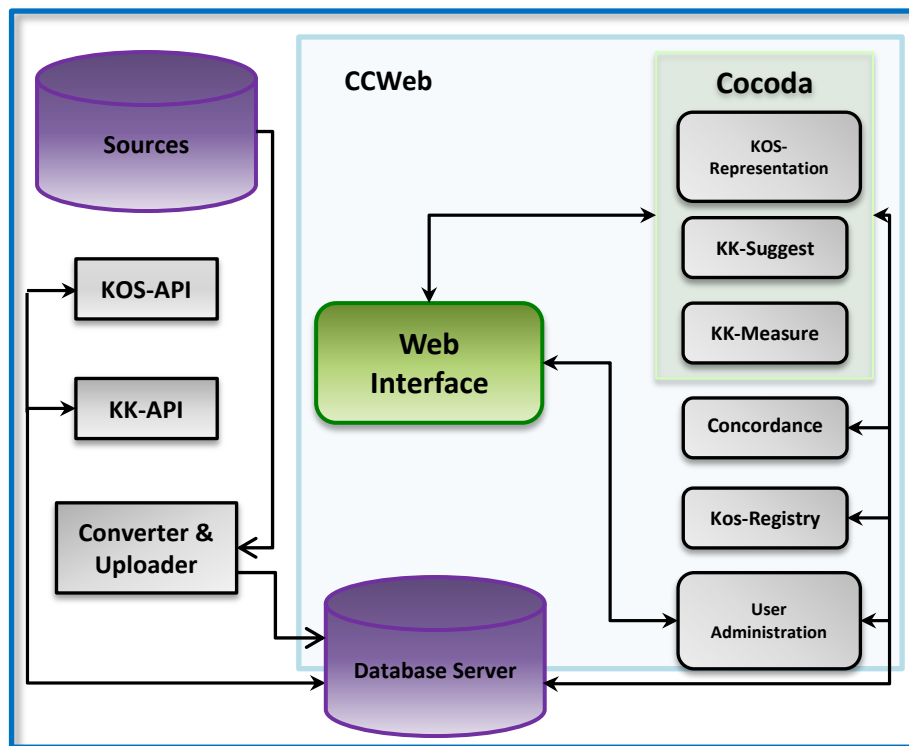


Figure 1: Coli-Conc System Architecture

1.1. CCWeb

The CCWeb is the heart of the system architecture. It comprises of:

- The Mapping tool – Cocoda
- KOS-Registry
- Concordances
- User Admin module and a Web Interface

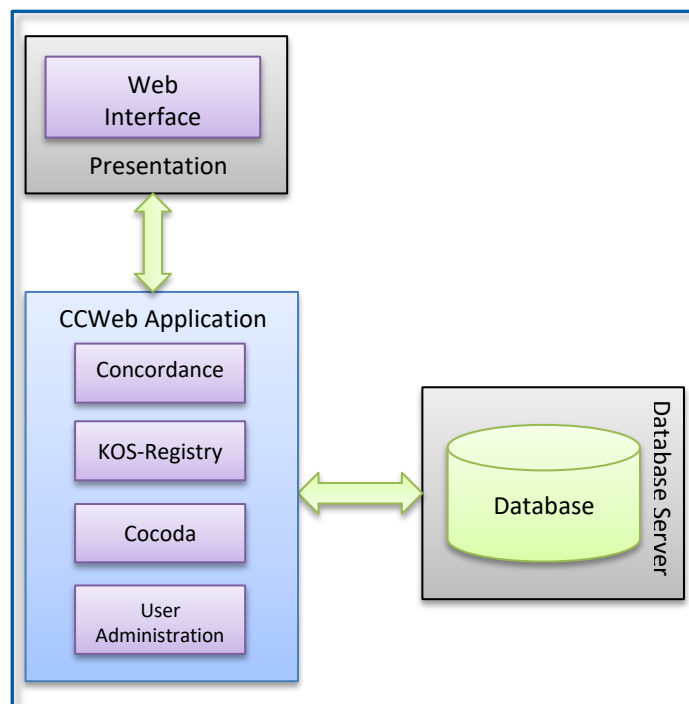


Figure 2: CCWeb

1.1.1. Mapping Tool – Cocoda

Cocoda is designed to perform multiple tasks to speed up the intellectual mapping building process between library KOS. The main components of the tool are:

- KOS-Representation Module: to search, browse and display hierarchical intra-KOS structural and content information for concept clarification.
- KK-Suggest Module: It combines several techniques and workflows into the framework (statistical co-occurrences, direct concept mapping and retrieval

of stored mappings from the concordance database) to maximize the results and generate mapping candidates for a selected source and target KOS.

- **KK-Measure Module:** to monitor and take care of quality assessment of the KOS and their mappings.

The screenshot displays the Cocoda Mapping Tool interface, which is divided into several panels. The top panel shows the browser address bar with the URL: `localhost:8080/cocoda-interface/cocoda.xhtml?sessionId=47185F15537385AAE9394A356C4CE6D1`.

The main interface is titled "Scheme View with Mapping" and contains three primary panels:

- Left Panel (Scheme View):**
 - Search by:** A search bar with the value "100". Below it, a list of search results is shown, including "281.3. Vornizänische Kirche, 100–325", "281.2. Apostolische Kirche bis zum Jahr 100 (Urchristentum)", "972.016. Klassische Zeit, ca. 100–ca. 900", and "100: Philosophie, Parapsychologie und Okkultismus, Psychologie".
 - Top Concepts Classification:** A hierarchical list of concepts, including "0 Informatik, Informationswissenschaft & allgemeine W...", "1 Philosophie & Psychologie", "2 Religion", "3 Sozialwissenschaften", "4 Sprache", "5 Naturwissenschaften", "6 Technik, Medizin, angewandte Wissenschaften", "7 Künste und Unterhaltung", "8 Literatur", and "9 Geschichte und Geografie".
 - DDC GND Mapping:** A table showing GND Notation, GND PrefLabel, GND Broader, and Relevance. The table contains three rows of data:

GND Notation	GND PrefLabel	GND Broader	Relevance
4045791-6	Philosophie	4045790-4	closeMatch
4185103-1	Theoretische Philosophie	4045791-6	closeMatch
4494545-0	Interkulturelle Philosophie	4045791-6	0.5
- Right Panel (Mapping):**
 - Concordance Database:** A table showing Source Scheme, Source Concept, Target Scheme, Target Concept, and Creator. The table contains five rows of data:

Source Scheme	Source Concept	Target Scheme	Target Concept	Creator
DDC	100	RVK	CA-CK	VZG
DDC	100	RVK	CL-CZ	VZG
DDC	100	RVK	CP 9400	VZG
DDC	100	RVK	BE 9130	VZG
DDC	100	RVK	BF	VZG
 - Library Catalogue:** A table showing DDC Notation, RVK Notation, and BK Notation. The table contains three rows of data:

DDC Notation	RVK Notation	BK Notation
100	CD 3067	08.21
296.3092B100290	DD 5541	15.9611.2008.25
100	PC 4630	

Figure 3: Mapping Tool – Cocoda

1.1.2. KOS-Registry

The KOS-Registry holds a collection of library KOS that are actively in use in the German speaking countries.⁵ The records have been classified under various types and also enriched with metadata based on the NKOS AP⁶ format. An interface enables a keyword search or selection of a specific type of KOS through a drop-down menu and retrieval of the metadata of each KOS in the selected set. The application is equipped with an export function in different formats (XLS, JSON and JSKOS).⁷ A script continually updates the registry through automatic data acquisition from BARTOC⁸.

The screenshot displays the KOS-Registry interface. At the top, there is a 'Select a Terminology' section with a search bar and a dropdown menu set to 'Universal Classification'. Below this is an 'Expand for Metadata' section. The first two items are 'UDK (Universelle-Dezimal-Klassifikation)' and 'DDC (Dewey Dezimal Klassifikation)'. The 'DDC' item is selected, and its metadata is displayed below. The metadata includes: Title: DDC (Dewey Dezimal Klassifikation), Alternative Title: DDC, DDK, Abstract: The Dewey Decimal Classification (DDC) system, devised by library pioneer Melvil Dewey in the 1870s and owned by OCLC since 1988, provides a dynamic structure for the organization of library collections. Now in its 23rd edition, and available in print and Web versions, the DDC is the world's most widely used library classification system. Author: Online Computer Library Center (OCLC), Type: classification scheme, Additional information: http://en.wikipedia.org/wiki/Dewey_Decimal_Classification, Bartoc uri: <http://bartoc.org/en/node/241>, Link: <http://www.oclc.org/dewey/>, Topic: Document indexing, DDC: 001, Classes: 10 main classes and 44,000 classes in total, Concept Scheme: <http://www.w3.org/2004/02/skos/core#ConceptScheme>, http://w3id.org/nkos/nkostype#classification_schema. Below the metadata, there is a list of other KOS: RVK (Regensburger Verbundklassifikation), BK (Basisklassifikation), LCC (Library of Congress Classification), GHB (Gesamthochschulbibliotheken-Aufstellungssystematik), ASB (Allgemeine Systematik für Öffentliche Bibliotheken), ÖsÖB (Österreichische Systematik für Öffentliche Bibliotheken), SFB (Systematik für Bibliotheken), and KAB (Klassifikation für Allgemeinbibliotheken). At the bottom, there is an 'Export' section with icons for XLS, JSON, and JSKOS. A footer note states: 'This database is a part of project coli-conc. See BARTOC for a general register of terminologies. Bartoc website: [BARTOC](http://bartoc.org)'.

Figure 4: KOS-Registry

⁵ Balakrishnan, U., Agne, J.M. (2016)

⁶ NKOS AP: <http://nkos.slis.kent.edu/nkos-ap.html>

⁷ Voß, J., Ledl, A., Balakrishnan, U. (2016)

⁸ BARTOC: <https://bartoc.org/>

1.1.3. Concordances

To store, manage, and access mappings of the Coli-conc and other related projects as well as to integrate the same into the KK-Suggest module, the project has developed a concordance platform as part of the CCWeb. The database of the platform contains currently over 200.000 mappings and will be further built up with the help of partner institutions.

The screenshot shows the 'Concordances' page of a web application. At the top, there is a navigation bar with links: 'Coli-conc', 'About', 'Technologies', 'Concordances', 'Create concordance', 'Publications', and 'Contact'. The main heading is 'Concordances', followed by a sub-heading 'This page provides an online preview of all colic-conc mapping database' and a green badge that says 'Current: 2013'. Below this is a 'Search Mappings' section with input fields for 'Name' (containing 'SIC'), 'Target' (containing 'SIC'), and 'Creator' (containing 'SIC'), and a 'Search' button. Underneath is a 'Your suggestions' section with a note: 'For suggestions, improvements or corrections, please use the form below. We are looking forward to your contributions.' This section includes input fields for 'Your Name', 'Email Address', 'Phone number', 'Target institution', and 'Comments', along with a 'Send message' button. At the bottom, there is a 'Documentation' section with a link to 'Colic-conc mapping database is available (2013) at [http://colic-conc.org/colic-conc/](#). See [CCWeb](#) for more information and technical documentation.' and a footer with '© 2013, CCWeb' and 'All rights reserved'.

Figure 5: Concordances

2. Technologies and Frameworks deployed for CCWeb

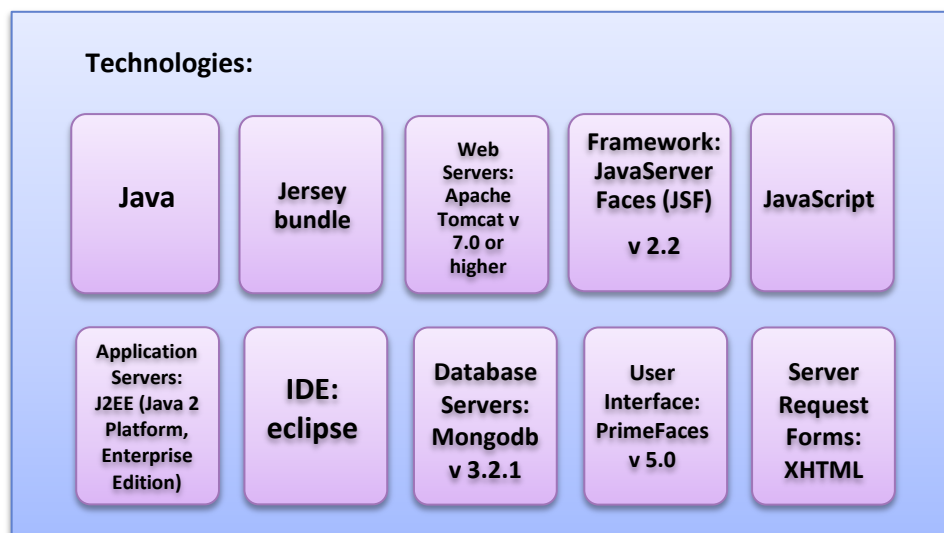


Figure 6: Technologies and frameworks for CCWeb

3. Data Conversion

One of the main components of the Coli-conc system is the JSKOS format that was modelled specifically for the project. Application for DDC (Dewey Decimal Classification) and RVK (Regensburg Classification) conversion from their proprietary formats into JSKOS has been developed using MARC4J package⁹. For other KOS (Basic Classification and Allgemeine Systematik für Öffentliche Bibliotheken,...) and mappings the mc2skos Python script¹⁰ was extended to support JSKOS. The conversion scripts for mappings are available at: <https://github.com/gbv/cocoda-mappings>.

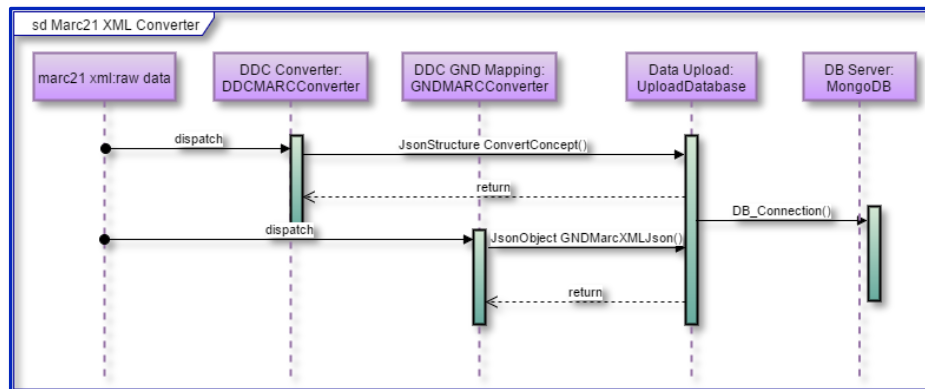


Figure 7: DDC Marc21 XML data converter Sequence Diagram

⁹ MARC4J package: <http://svn.k-int.com/default/components/marc4j/tutorial.html>

¹⁰ mc2skos Python script : <https://pypi.python.org/pypi/mc2skos>

4. JSKOS API

An additional special feature of the Coli-conc architecture is the JSKOS API. Among the key objectives of the project is to provide uniform and easy access to KOS and their mappings on the web. This has been affected by the creation of the JSKOS API. The service has been so far implemented as a database application for DDC, RVK, Basic classification (BC) and as wrappers to access Gemeinsame Normdatei (GND) Wikidata, Open Research and Contribution ID (ORCID). However, the use of the DDC API is subject to a license requirement.

```
[{
  "uri" : "http://dewey.info/class/003/e23",
  "prefLabel" : { "de" : "Systeme"},
  "notation" : ["003"],
  "type" : ["http://www.w3.org/2004/02/skos/core#Concept"],
  "broader" : [{ "notation" : ["00"], "uri" : "http://dewey.info/class/00/e23"}],
  "inScheme" : [{ "uri" : "http://dewey.info/scheme/edition/e23" }, { "uri" :
    "http://dewey.info/scheme/ddc"}]
}]
```

Figure 8: DDC- JSKOS API

Project Partner Institutions

The project is being funded by the German Research Foundation. It has received support from the German National Library, several expert groups, large academic libraries and international institutions.



Figure 9: Project Partner Institutions

References

Balakrishnan, U., Agne, J.M. (2016). Coli-conc Survey - Ergebnisse der Online-Umfrage zur Sacherschließung und Konkordanzprojekten. Retrieved from <http://coli-conc.gbv.de/publications/>

Lauser, B., Johannsen, G., Caracciolo, C., Keizer, J., van Hage, W. R., Mayr, P. (2008). Comparing human and automatic thesaurus mapping approaches in the agricultural domain. *10th International Conference on Dublin Core and Metadata Applications*. Retrieved from <https://arxiv.org/abs/0808.2246>

Pfeffer, M. (2013): Automatic creation of mappings between classification systems. *Workshop Klassifikation und Sacherschließung (LIS'2013)*, University of Luxembourg (Presentation). Retrieved from <https://publikationen.bibliothek.kit.edu/1000035777>

Voß, J., Ledl, A., Balakrishnan, U. (2016). Uniform description and access to Knowledge Organization Systems with BARTOC and JSKOS. *Proceedings of TOTH conference 2016*. Retrieved from <https://zenodo.org/record/438019>

Voß, J. (2016). JSKOS data format for knowledge organization systems. Retrieved from <https://gbv.github.io/jskos/jskos.html>

Walter, A.-K., Mayr, P., Petras, V., Baerisch, S. (2007). Kompetenzzentrum Modellbildung und Heterogenitätsbehandlung (KoMoHe). Retrieved from <https://www.gesis.org/forschung/drittmittelprojekte/archiv/komohe/>