

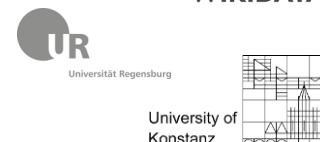
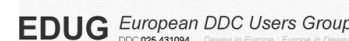
Die stetige Entwicklung in der Informationstechnologie hat den Zugang zu Informationen global erleichtert. Während die Nutzung und der Austausch von KOS¹ in den letzten Jahrzehnten erheblich zugenommen hat, sind Konkordanzen zwischen bibliothekarischen KOS weiterhin rar. Für eine nahtlose Suche in einer heterogenen Informationslandschaft ist eine semantische Integration der KOS unabdingbar. In den vorangegangenen Jahren gab es große Bemühungen, eine Vereinheitlichung der Sacherschließungslandschaft in deutschen Bibliotheken anzustoßen, dies ist leider nur teilweise erreicht worden.

Außerdem sind Mappings mit der international meistgenutzten DDC², die Entwicklung eines einheitlichen Formats für einen verbesserten Austausch der KOS und Mappings, sowie der leichte Zugang zu den vorhandenen Mappings ein langjähriges Desiderat der Fachwelt im deutschsprachigen Raum.

Angesichts dieser Gegebenheiten hat sich das VZG³-Projekt coli-conc zum Ziel gesetzt, eine Infrastruktur zur semi-automatischen Erstellung, für den Austausch, die Wartung und zur Qualitätssicherung von Konkordanzen zwischen bibliothekarischen KOS aufzubauen. Es werden dabei wiederverwendbare Module zum einheitlichen Zugriff auf KOS, Konkordanzen und Konkordanz-Bewertungen, die zusammen eine Webanwendung zur effektiven Bearbeitung von Konkordanzen bilden, entwickelt (siehe Abb. 1). Die bereits bestehenden KOS-Software, Portale, Mapping-Algorithmen, Katalogisierungssoftware, etc. werden im Rahmen des Projektes evaluiert und ggf. für deren Anwendung im coli-conc erweitert. Die im Projekt entwickelten Anwendungen stellt das Projekt für die Nachnutzung und Forschungszwecke bereit. Zudem hält das Projekt zur Qualitätssicherung und nachhaltigen Nutzung eine kollaborative Arbeit unter den Experten für sehr wichtig und fördert diese.

Das Projekt wird zunächst für zwei Jahr von der DFG⁴ gefördert und nahm seine Arbeit im Dezember 2015 auf. Viele nationale und internationale Institutionen, Fachgruppen und Projekte unterstützen das Vorhaben des Projektes coli-conc und fungieren als Kooperationspartner.

Kooperationspartner:



¹ Knowledge Organization Systems

² Dewey Dezimal Klassifikation

³ Verbundzentrale des Gemeinsamen Bibliotheksverbundes

⁴ Deutschen Forschungsgemeinschaft

Kontakt:

balakrishnan@gbv.de

coli.conc@gbv.de

coli-conc.gbv.de

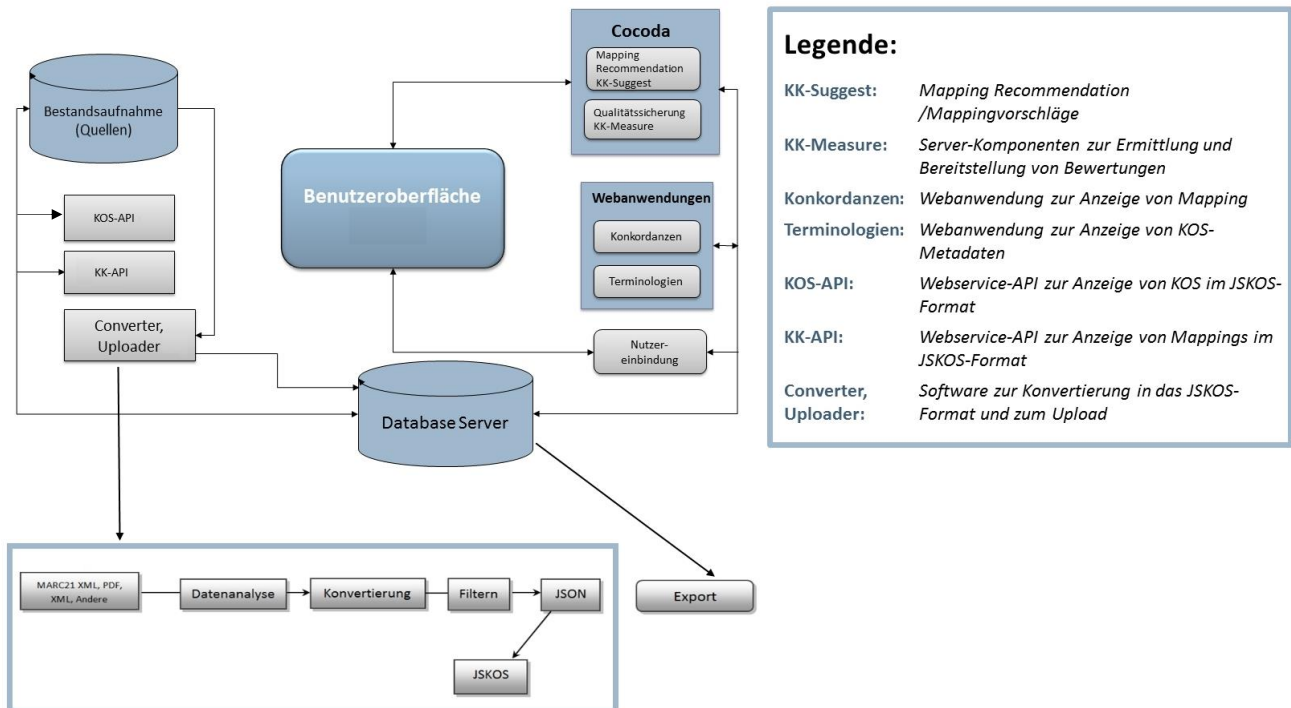


Abb. 1: coli-conc System-Architektur

Das coli-conc-Projekt sieht folgende Cocoda-Module und Stand-alone-Dienste vor:

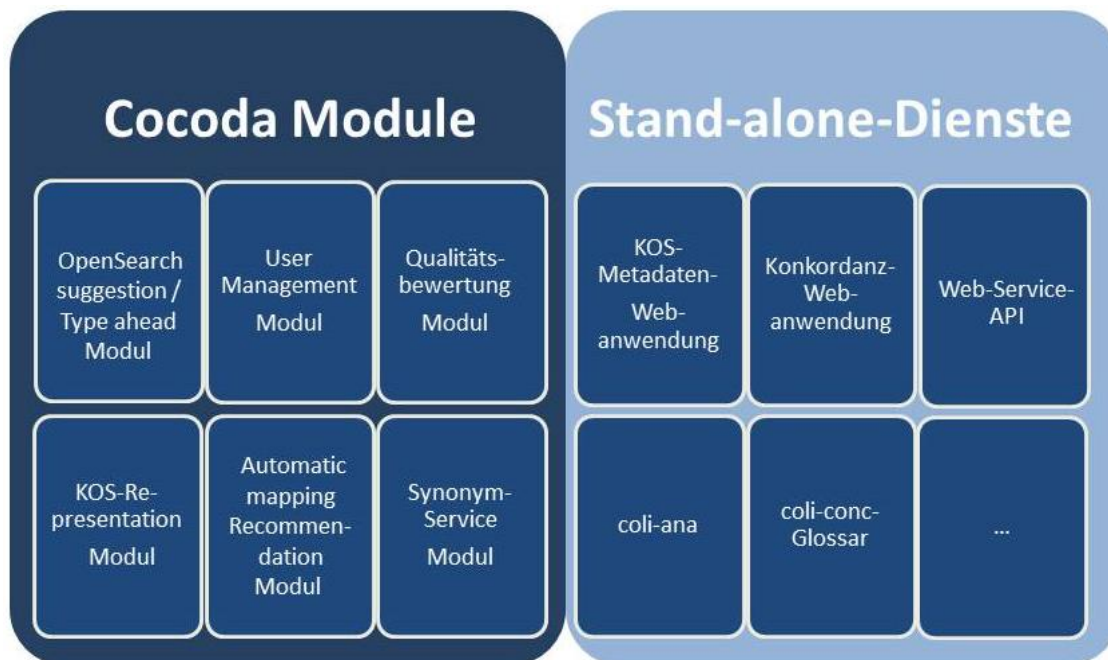


Abb. 2: Cocoda-Module und Stand-alone-Dienste

Das Projektjahr I

Im ersten Projektjahr lag der Fokus der Aufgaben auf dem Aufbau der Basis

für das Mappingtool Cocoda. Zusätzlich legte das coli-conc-Projekt die wichtigsten Grundsteine für die Cocoda-Module und konnte die ersten drei der oben genannten Stand-alone-Dienste bereits (in Beta-Versionen) verwirklichen.

Bestandsaufnahme

Zu Beginn der Projektzeit wurde eine Bestandsaufnahme durchgeführt, um den aktuellen Stand der KOS, Mappingprojekte und vorhandenen Mappings im deutschsprachigen Raum zu ermitteln. Die Ergebnisse sollten, die in Vorarbeit erstellte Datengrundlage erweitern und sie auf den neuesten Stand bringen. Ein ausführlicher Bericht dazu ist auf <https://coli-conc.gbv.de/publications> veröffentlicht und in der VZG aktuell (Ausgabe 2016_2⁵) kurz zusammengefasst.

Die ermittelten KOS und Mappings wurden danach in das einheitliche JSKOS-Format konvertiert und über die Stand-alone-Dienste in Form von Webanwendungen zur Verfügung gestellt.

Konvertierungsmodul: Das einheitliche Format JSKOS

Im Rahmen des Projektes wurde das JSKOS-Format⁶ entwickelt, um zur Standardisierung eines Formats für den Austausch der KOS und Mappings beizutragen. Dieses Format dient nun neben MARCXML (für KOS) als Standardformat für KOS und Mappings im Projekt. Die Vorzüge des JSKOS-Formats wurden letztes Jahr auf Konferenzen präsentiert und von den Experten gut aufgenommen. Das Format ist bei Projektpartnern wie BARTOC⁷ schon in Anwendung, andere Projekte und Institutionen wie ISKO⁸, SiGMaMat⁹ und CCMapper¹⁰ signalisierten auch Interesse an einer Einführung. Die Konvertierungsskripte hierfür stehen im Sinne des Open Source frei zur Verfügung.

KOS-Metadatenanzeige

Die KOS Metadaten-Dienst (siehe Abb. 3) verzeichnet die Metadaten der bibliothekarische KOS, die im Mittelpunkt des Projektes stehen und im deutschsprachigen Raum aktiv angewendet werden.

⁵ https://www.gbv.de/Verbundzentrale/Publikationen/broschueren/vzg-aktuell/VZG_Aktuell_2016_02.pdf

⁶ <https://gbv.github.io/jskos/jskos.html>

⁷ BARTOC: Basel Register of Thesauri, Ontologies & Classifications

⁸ ISKO: International Society for Knowledge Organization

⁹ SigMaMat: SIS-GND-Matching-AutoMat

¹⁰ <http://www.uio.no/for-ansatte/enhetssider/ub/prosjekter/mapping-mot-webdewey/delte-dokumenter/brosjyre-engelsk-endelig.pdf>

Die Datenbank dieser Webanwendung enthält derzeit Metadaten von 32 KOS, die laufend aktualisiert und automatisch mit Metadaten von BARTOC.org angereichert werden. Das entsprechende Skript steht für die Nachnutzung frei zur Verfügung.

Der Dienst bietet zusätzlich eine Downloadfunktion, die das Herunterladen der Metadaten in verschiedenen Formaten (Excel, JSON, etc.) ermöglicht.

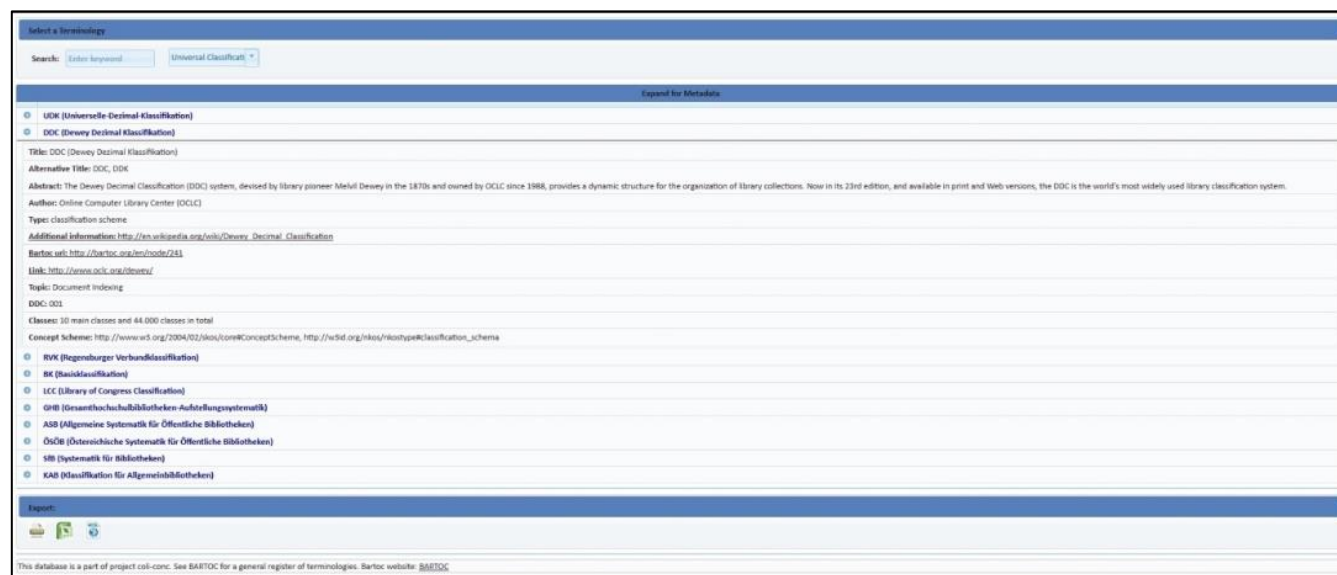


Abb. 3: KOS-Metadaten-Webanwendung

Konkordanz-Webanwendung

Dieser Service bietet eine zentrale Plattform zum Speichern, Abrufen und Verwalten von hauptsächlich im deutschsprachigen Raum erstellten Mappings. Die Datenbank enthält derzeit 137.500 uni- und bidirektionale Mappings zwischen der DDC, RVK, GND und BK¹¹. Nach der Konvertierung in JSKOS wurden die Mappings in einer Datenbank zusammengeführt, sodass ein einheitlicher Zugang dazu gewährleistet werden kann. Die Webanwendung hat damit die lange gewünschte Bereitstellung der Mappings über eine zentrale Plattform in die Wege geleitet.

Außerdem stehen durch die Einbindung von Experten deutschlandweit und die Kooperation mit Projektpartnern die kollaborative Pflege, weitere Anreicherungen und Verwaltung der Mappings im Vordergrund dieses Dienstes. Zur Realisierung dieses Vorhabens wird im zweiten Projektjahr eine Expertengruppe gegründet.

¹¹ Regensburger Verbundklassifikation (RVK);
Gemeinsame Normdatei (GND);
Basisklassifikation (BK)

Abb. 4: Konkordanz-Webanwendung-Anzeige

Web-API

Um den Zugang der KOS und Mappings zu vereinfachen und die Nutzung zu erhöhen, wurde parallel zu der Konkordanz-Webanwendung eine Webservice API¹² entwickelt (siehe Abb. 5 & 6).

Über diese API können unter Berücksichtigung der Lizenzbedingungen, verschiedene KOS aufgerufen werden. Der API ist zum Teil auch im Normdatendienst ELMA¹³ implementiert

Der Service erlaubt eine Anzeige der KOS-Daten im JSKOS-Format durch das Angeben der URI¹⁴ oder der Notation.

```

[ {
  "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"} ]
} ]

```

Abb. 5: Webservice-API - DDC-Anzeige

¹² API: Application Programming Interface
¹³ ELMA: Entity Lookup Microservice API
¹⁴ URI: Uniform Resource Identifier

```

[ {
  "uri" : "http://uri.gbv.de/terminology/rvk/AA",
  "prefLabel" : { "de" : "Bibliographien der Bibliographien, Universalbibliographien, Bibliothekskataloge, Nationalbibliographien"},
  "notation" : [ "AA"],
  "type" : [ "http://www.w3.org/2004/02/skos/core#Concept"],
  "broader" : [ { "notation" : [ "A"] , "uri" : "http://uri.gbv.de/terminology/rvk/A"} ],
  "inScheme" : { "notation" : [ "RVK"] , "uri" : "http://bartoc.org/en/node/533"}
} ]

```

Abb. 6: Webservice-API - RVK-Anzeige

Mappingtool Cocoda – Benutzeroberfläche

Die aktuellen Arbeiten an der Benutzeroberfläche von Cocoda beinhalten sowohl den Frontend- als auch Backendbereich des Mappingtools. Nachdem zunächst mit einem Prototyp verschiedene Funktionen exemplarisch erprobt wurden, wird zurzeit eine neue Umgebung für das Tool entwickelt, die eine stabile Funktionalität des Tools sicherstellt (siehe Abb. 7).

Das Kontextverständnis eines Begriffes spielt in der Konkordanzerstellung eine wichtige Rolle. Zur Erhöhung der Effizienz der semi-automatischen Konkordanzerstellung ist es daher erforderlich, alle relevanten Zusatzinformationen des Quell- und Ziel-KOS (wie Anmerkung, Mappings, Registereinträge) und die Ergebnisse der weiteren Services (Synonymservice/Mapping-Recommendation-System) auf einem Bildschirm anzuzeigen (single screen view).

Die Cocoda-Benutzeroberfläche wurde hierfür in drei Bereiche eingeteilt. Die linke und rechte Seite des Bildschirms ist für die Anzeige des Hierarchiebaums der ausgewählten Quell- und Ziel-KOS bzw. deren Notationen und Benennungen, einschließlich der notwendigen Zusatzinformationen vorgesehen. Der Mittelteil wird die automatisch generierten Mapping-Kandidaten aus verschiedenen Quellen (Online-Katalogen, dem Zielsystem, der Konkordanzdatenbank) präsentieren und auch die Möglichkeit zur Überprüfung, Bearbeitung, Speicherung bzw. zum Herunterladen bieten.

Im ersten Arbeitsschritt liegt die Priorität dieser Arbeit auf der DDC und RVK.

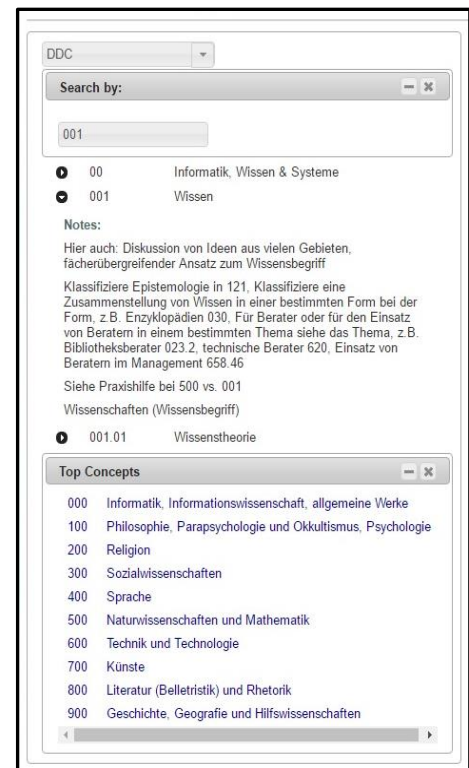


Abb. 7: Anzeige des Quell-/Zielsystems auf der Cocoda Oberfläche

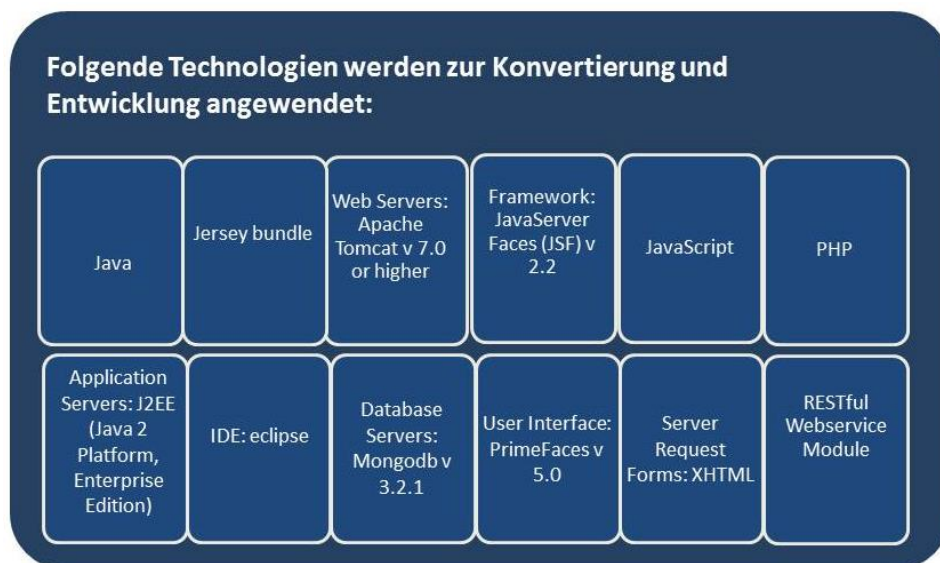


Abb. 8: Technologien für die Konvertierung und Entwicklungen im coli-conc-Projekt

Mapping-Verfahren

Für die automatische Generierung der Mappingvorschläge wurde eine Evaluation ähnlicher Verfahren anderer Projekte (CCMapper, Digitaler Assistent, Ariadne, Amalgame, SiGMaMat) durchgeführt. Diese hat ergeben, dass es noch kein Verfahren für die Konkordanzerstellung zwischen zwei bibliothekarischen Klassifikationssystemen gibt, das allen Ansprüchen des coli-conc-Projekts genügt. Zudem wurde festgestellt, dass keines davon die Einbindung eines mehrschichtigen Verfahrens vorsieht, welches für die Qualität der Konkordanzen und die Effizienz der Konkordanzerstellung unabdingbar ist. Die evaluierten Projekte werden auf dem SI & IT Workshop präsentiert.

Ausblick: Projektjahr 2017

Im zweiten Projektjahr wird der Schwerpunkt der Arbeit auf der Fertigstellung der Cocoda-Benutzeroberfläche und des Mapping-Verfahrens zur Generierung der Mapping-Kandidaten (DDC-RVK), liegen. Parallel zur Integration der einzelnen KOS (DDC, RVK, GND und BK) und der diversen Module in das Mappingtool Cocoda wird im Rahmen des SI & IT¹⁵ Workshops im Mai 2017 eine Expertengruppe gebildet, die die Weiterentwicklung, Nutzbarkeit und die Qualitätsprüfung der vorhandenen Mappings vorantreiben soll.

Zudem werden die Implementierung bzw. der Ausbau der Import-Export-Funktion, ein Usability-Test und die Überführung der Dienste in ein verbessertes Layout zu den Hauptaufgaben gehören.

Die Stand-alone-Dienste „coli-ana“ (Anzeige der automatische Analyse und Zerlegung der DDC-Notationen) und coli-conc-Glossar werden im Fall einer Verlängerung der jetzigen Projektzeit angeboten.

Ein detaillierter Zwischenbericht samt technischer Spezifikationen wird nach dem SI & IT Workshop im Mai auf der coli-conc Homepage publiziert.

¹⁵ Subject Indexing & Information Technology