

■ EIN SEMANTISCHES NETZ FÜR DIE SUCHE MIT DER DEWEY-DEZIMALKLASSIFIKATION – OPTIMIERTES RETRIEVAL DURCH DIE VERWENDUNG VERSIONIERTER DDC-KLASSEN

von *Claudia Effenberger*

Inhalt

1. Einleitung
2. Problematik
3. Lösung des Problems durch Versionierung der DDC und Methoden des Semantic Web
4. Erstellung einer Testdatenbank
5. Auswahl der Testfälle
6. Beispielhafte Darstellung der neuen Suche
7. Weitere Beispiele
8. Zusammenfassung

Zusammenfassung: Die Dewey-Dezimalklassifikation (DDC) wird weltweit für die Erschließung von Bibliotheksbeständen eingesetzt. Um immer auf dem aktuellen wissenschaftlichen Stand zu sein, wird die DDC regelmäßig aktualisiert. Hierdurch kann es zu Begriffsverschiebungen kommen und die Erschließung ähnlicher Titel weicht voneinander ab, wird also heterogen, was zu inkorrekten Treffermengen bei der Suche nach DDC-erschlossenen Dokumenten führen kann. Eine mögliche Lösung des Problems ist, jeder Version einer DDC-Klasse einen eindeutigen Identifier zuzuordnen und so die Klasse eindeutig identifizieren zu können, mit der ein Titel verknüpft ist. Mit Hilfe von RDF kann aus Titeldaten, Schlagwörtern der Schlagwortnormdatei und den versionierten DDC-Klassen ein semantisches Netzwerk modelliert werden und durch entsprechende Anfragen können genauere Suchergebnisse, unabhängig von Änderungen in der DDC, generiert werden. Die Untersuchungen erfolgten im Rahmen eines Werkvertrages mit der Deutschen Nationalbibliothek und wurden in ausgearbeiteter Form als Masterthesis eingereicht.

Schlagwörter: Dewey-Dezimalklassifikation, inhaltliche Erschließung, Semantic Web, Versionierung

A SEMANTIC NETWORK FOR THE SEARCH WITH THE DEWEY DECIMAL CLASSIFICATION – OPTIMIZED RETRIEVAL BY USING VERSIONED DDC CLASSES

Abstract: *The Dewey Decimal Classification (DDC) is used to catalogue library holdings all over the world. In order to keep the classification up-to-date with scientific advancement, the DDC is revised regularly. As a result, particular topics might be re-located and the original DDC notation might not represent the topic of a resource any more. This can lead to incorrect search results. A possible solution for this problem is to assign a unique identifier to each version of a DDC class. By doing that, it is possible to explicitly label which version of a DDC class was used for the classification of a particular resource. Bibliographic resources, terms from the German Subject Headings, and the different versions of the DDC classes can then be used to build a semantic network. Using the semantic network with appropriate queries allows for better search results independent of modifications within the DDC.*

Keywords: *Dewey Decimal Classification, subject indexing, Semantic Web, Versioning*

1. Einleitung

Um Informationen zu ordnen und wieder auffindbar zu machen, werden häufig Klassifikationen genutzt.[Stock] Eine methodische Revolution erlebten Klassifikationsforschung und -praxis mit der „Decimal Classification“ (1876) des amerikanischen Bibliothekars Melvil Dewey (1851–1931).[Dewey] Das Grundprinzip seiner Klassifikation baut auf dem Dezimalsystem auf und ist damit sehr einfach: Das Wissen wird auf zehn Bereiche verteilt, die durch Dezimalzahlen dargestellt werden. Diese sogenannten 10 Hauptklassen werden jeweils wieder in 10 untergeordnete Klassen unterteilt usw., sodass es auf der zweiten Ebene 100 Klassen und auf der dritten Ebene 1000 Klassen gibt. Die Aufstellung der Bücher in der Bibliothek folgt der Klassifikation, was dazu führt, dass thematisch verwandte Werke dicht beieinander stehen. Die Dewey-Dezimalklassifikation (DDC) wird, da sie ständig weiterentwickelt wurde, auch heute noch für die inhaltliche Erschließung genutzt, um Dokumente aus einer Fülle von Dokumenten herausfinden zu können. Die DDC ist eine überwiegend präkombinierte Klassifikation mit facettierten Erweiterungsmöglichkeiten. Sie ist das am weitesten verbreitete Klassifikationssystem der Welt, wird in Bibliotheken in über 135 Ländern eingesetzt und mehr als 60 Länder verwenden sie in ihren Nationalbibliografien.[DDC22] Seit dem Bibliografiejahrgang 2004 werden die Titel der Deutschen Nationalbibliographie nach DDC-Sachgruppen klassifiziert und seit 2006 wird die DDC für die Feinklassifikation von Dokumenten in der Deutschen Nationalbibliothek (DNB) genutzt.

Um die DDC in Deutschland einsetzen zu können, war ihre Übersetzung ins Deutsche eine Grundvoraussetzung. Hierzu wurde von 2002 bis 2005 das Projekt DDC Deutsch [DDCDeutsch] durchgeführt, bei dem es unter anderem darum ging, die 22. Ausgabe der DDC ins Deutsche zu übertragen. Darüber hinaus konnten einige für die deutschsprachigen Länder wichtige Erweiterungen und Überarbeitungen in die Originalausgabe eingebracht werden.

In der Deutschen Nationalbibliothek werden seit 2006 bzw. 2007 die Dokumente der Reihen A, B und H der Deutschen Nationalbibliografie mit vollständigen DDC-Notationen erschlossen.[Alex2011] Reihe A umfasst Monografien und Periodika des Verlagsbuchhandels, Reihe B die Monografien und Periodika außerhalb des Verlagsbuchhandels und Reihe H die Hochschulschriften.[Junker] In der Reihe A werden weiterhin Schlagwörter nach RSWK vergeben. Der Einsatz der DDC markiert keine Abwendung von der verbalen Sacherschließung, sondern vielmehr eine Erweiterung, bei der die verbale Erschließung nach RSWK und die klassifikatorische mit der DDC einander ergänzen.[Heiner-Freiling] Eine Verbindung der beiden Methoden wurde beispielsweise im DFG-geförderten Projekt CrissCross erstellt. In dem Projekt, in dem es um Interoperabilität von Wissensorganisationssystemen ging, wurden unter anderem Beziehungen zwischen den Schlagwörtern der Schlagwortnormdatei (SWD) und Klassen der DDC hergestellt.[Crisscross]

Mit der Einführung der DDC als Erschließungsinstrument ist die DNB mit über 100.000 mit der DDC erschlossenen Titeln pro Jahr [Alex2009] zum größten europäischen Lieferanten von DDC-Notationen geworden und nimmt weltweit nach der Library of Congress die zweite Position ein. [Ansorge]

2. Problematik

Die DDC befindet sich im ständigen Wandel. Um die Klassifikation auf dem aktuellen wissenschaftlichen Stand zu halten, muss sie laufend überarbeitet werden, und so wird regelmäßig der inhaltliche Umfang einzelner DDC-Klassen verändert. Bestimmte Wissensgebiete werden in bereits vorhandene Klassen ausgelagert und teilweise werden auch neue Unterklassen geschaffen.

Es gibt die verschiedensten Gründe, warum es in Klassifikationen zu Aktualisierungen kommt. Der wichtigste ist, dass sich Wissen laufend verändert und eine einmal aufgestellte Klassifikation nicht für einen längeren

Zeitraum dieses Wissen repräsentieren kann, wenn sie starr bleibt und nicht angepasst wird. Deweys Dezimalklassifikation war in ihrer ursprünglichen Form durch Wissensorganisationsprinzipien des späten 19. Jahrhunderts geprägt.[Stock] Ohne zahlreiche, teilweise gravierende Änderungen wäre die DDC heute in der Praxis nicht mehr einsetzbar.

Wenn sich Wissen entwickelt, dann müssen die entsprechenden Erweiterungen einen Platz in der Klassifikation finden, was sich an einem Beispiel aus der Informatik zeigen lässt. Wenn es zu technologischen Innovationen verstärkt Literatur gibt, wie im Bereich Smartphones oder neuen Datenformaten wie RDF (Resource Description Framework), muss eine eigene Systemstelle geschaffen werden. Dies ist auch bei Themen mit erhöhtem Literaturanfall erforderlich, die keine eigenständige Notation aufweisen, sondern größeren Bereichen untergeordnet sind. Dann müssen solche Themen ausdifferenziert und neue Klassen geschaffen werden, wie das z.B. im Bereich des Terrorismus geschehen ist.[Bee] Auch gibt es in auf alten Grundstrukturen beruhenden Klassifikationen immer wieder Bereiche, die nicht mehr heutigen Formen der Wissensorganisation entsprechen. Für solche Bereiche ist eine grundsätzliche Neustrukturierung angeraten, wie dies beispielsweise im Bereich der Religion geschehen ist.[Boll]

Eine andere Frage, die sich stellt, ist, wie neues Wissen verortet werden soll. Soll die Biotechnologie unter der Biologie, der Genetik, der Medizin, der Industrie oder der Landwirtschaft eingeordnet werden? Und die Nanotechnologie im Bereich der Technologie, der Computertechnologie, der Medizin, der Industrie oder der Landwirtschaft? Oft werden neue Technologien erst einmal in eine bereits bestehende Klasse mit aufgenommen. Später kann sich dann herausstellen, dass sie in einer anderen Klasse viel besser aufgehoben wären oder dass sie sogar eine eigene Klasse benötigen, da es für das entsprechende Thema ein hohes Literataufkommen gibt. [Slavicsaac] In der Vergangenheit führten auch Übersetzungsprojekte immer wieder dazu, dass zusätzliches Wissen in die DDC eingebracht werden konnte. So wurde z.B. mit der italienischen Übersetzung die geografische Hilfstafel im Bereich Italien und mit der deutschen Übersetzung der Bereich Deutschland erweitert. Ein anderes Beispiel ist der Bereich des europäischen Rechts, wo aktuell intensive Bestrebungen nach Verbesserungen im Gang sind.[EDUG] Solche Aktualisierungen werden zunächst in der amerikanischen Ausgabe der DDC übernommen [OCLC], bevor sie zeitversetzt in die deutschen Daten eingearbeitet werden.

Durch solche Anpassungen in der DDC kann es beispielsweise zu Begriffsverschiebungen kommen. Inhalte verschieben sich in andere Klassen, wodurch die einmal am Titel vergebenen Notationen manchmal nicht mehr

stimmen. Aktualisierungen gibt es in den verschiedensten Bereichen der DDC, so gab es beispielsweise im März 2009 eine Änderung in der Klasse 726.79 Teile und Nebengebäude Einschließlich: Klausen, Kreuzgänge, Refektorien. Hier wurde der Begriff Kreuzgänge ausgegliedert und eine eigene Unterklasse 726.796 Kreuzgänge gebildet. [DDCDeutschUpdate] Einschließlich-Hinweise dienen in der DDC dazu, Themen einen vorläufigen Platz zu geben. Wenn es vermehrt Literatur zu einem Thema in einem Einschließlich-Hinweis gibt, kann es zur Bildung einer neuen Klasse kommen.

726.79	Teile und Nebengebäude (2005/11/02)
	Einschließlich: Klausen, Kreuzgänge, Refektorien
726.79	Teile und Nebengebäude (2009/03/25)
	Einschließlich: Klausen, Refektorien
726.796	Kreuzgänge (2009/03/25)

Nutzer wollen relevante Ressourcen finden, auch wenn die vergebene DDC-Notation nicht mehr genau den Inhalt eines Werkes beschreibt. Kreuzgänge sind seit der Ausgliederung auf die beiden Klassen 726.79 Teile und Nebengebäude und 726.796 Kreuzgänge verteilt (bei näherer Betrachtung sogar auf drei Klassen, hierzu später mehr).

Grundsätzlich ist es aus ökonomischen Gründen nicht möglich, mehrere Millionen DDC-Notationen, die in der Vergangenheit weltweit für einzelne Dokumente vergeben wurden, zu aktualisieren; diese Problematik betrifft auch die seit 2006 in der DNB vergebenen DDC-Notationen. Durch die nicht mehr dem aktuellen Stand entsprechende DDC-Notationen kann es zu unvollständigen Treffermengen bei der Suche nach DDC-erschlossenen Daten kommen, was besonders bedauerlich ist, wenn man bedenkt, dass die DDC ein wichtiges Erschließungsinstrument in der DNB ist und die Reihen B und H der Deutschen Nationalbibliografie nur noch mit der DDC erschlossen werden. Für dieses Problem, das grundsätzlich alle Nutzer des Online-Kataloges betrifft, gibt es zwei Lösungsansätze: entweder müssen, die Notationen nachträglich automatisch angepasst werden oder der negative Einfluss der Änderungen auf die Suchergebnisse muss beseitigt bzw. minimiert werden.[Svensson] Der zweite Ansatz wird in diesem Artikel mit Hilfe von Methoden des Semantic Web verfolgt.

3. Lösung des Problems durch Versionierung der DDC und Methoden des Semantic Web

Eine mögliche Lösung des oben geschilderten Problems ist, nach jeder Aktualisierung einer Klasse die alte und die neue Version der Klasse zu speichern. Dies geschieht im Editions-system, in dem die Aktualisierungen durchgeführt werden. Momentan werden die Änderungen umgesetzt und die aktualisierte Klasse wird für die inhaltliche Erschließung verwendet. Die alte Version der Klasse wird nicht weiter benötigt. Bei einer Versionierung werden beide vorliegenden Versionen einer Klasse gespeichert und für jede Version ein eindeutiger Identifier vergeben. Nur die neueste Version der Klasse darf für die inhaltliche Erschließung genutzt werden. Mit Hilfe des Identifiers kann der zur Zeit der Vergabe gültige Klasseninhalt, also die tatsächlich verwendete Bedeutung einer DDC-Notation, am Titel identifiziert werden. Um diese Versionsinformationen optimal zu nutzen, kann aus den Beziehungen zwischen den Daten mittels der Programmiersprache RDF ein semantisches Netzwerk modelliert werden. Titeldaten und die Schlagwörter der Schlagwortnormdatei können dann mit explizit versionierten DDC-Klassen verbunden werden. Durch entsprechende Anfragen können genauere Suchergebnisse, unabhängig von Änderungen in der DDC, geliefert werden.[Svensson]

Für Untersuchungen, ob dieser Ansatz wirklich zu besseren Ergebnissen bei der Suche führt, mussten DDC-Klassen, Titel und Schlagwörter in eine Testdatenbank überführt werden. Hierfür wurden Methoden des Semantic Web genutzt. Das bisherige Web besteht aus Ressourcen und Links zwischen diesen Ressourcen. Im Semantic Web wird dieses Web um eine semantische Schicht ergänzt, d.h. die Ressourcen und ihre Beziehungen untereinander sind typisiert. Hierdurch kann festgestellt werden in welcher Beziehung Ressourcen zueinander stehen, was wiederum für die Suche genutzt werden kann.[W3CSemWeb]

Eine wichtige Sprache im Semantic Web ist RDF (Resource Description Framework), die für die Repräsentation von Metadaten genutzt wird. Die Sprache basiert auf Tripeln, gerichteten Graphen, die aus einem Subjekt, einem Prädikat und einem Objekt bestehen. An allen Positionen des Tripels kann eine URI-Referenz stehen, das Objekt kann auch ein Literal sein. Die Prädikate sind die bereits genannten typisierten Links, mit denen die verschiedenen Beziehungen zwischen den Ressourcen dargestellt werden können. Bei der graphischen Darstellung von Tripeln oder Graphen aus Tripeln werden URI-Referenzen (Uniform Resource Identifier-Reference) als Ovale, typisierte Links als Pfeile und Literale als Rechtecke dargestellt.[RDF]

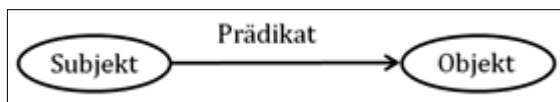


Abb. 1: Ein RDF-Triple (Quelle: <http://www.w3.org/TR/rdf-concepts/>)

In der DNB wurden im Rahmen der Projekte Linked Data und Linked Data II das erste Mal Daten der DNB für das Semantic Web aufbereitet, bei denen individualisierte Personen aus der PND, Sachschlagwörter der SWD, Körperschaften aus der GKD und die Klassen und Sachgruppen aus der deutschen DDC umgesetzt wurden. [LinkedDataService]

Linked Data baut auf mehreren Prinzipien auf. Im Kern geht es um die Identifizierbarkeit und Referenzierbarkeit von Ressourcen und darum, bestimmte Standards zu benutzen. Einer dieser Standards ist RDF. [Berners-Lee]

4. Erstellung einer Testdatenbank

Die Testdatenbank wurde erstellt, um mit realen Daten und Suchanfragen herauszufinden, ob die bisherigen Überlegungen in der praktischen Umsetzung die gewünschten, besseren Suchergebnisse liefern. Die Testdatenbank enthielt die Sachschlagwörter der SWD, die DDC und Titeldaten. Im Gegensatz zur SWD und zur DDC ist von den Titeldaten nicht der gesamte Bestand der DNB in der Testdatenbank enthalten. Hier wurden nur Titeldaten aufgenommen, die in bestimmten Notationsspannen, in denen es zu Änderungen gekommen ist, lagen. Alle aufgeführten Zahlen und Ergebnisse beziehen sich auf ebendiese Auswahl.

Die bestehenden Daten mussten in unterschiedlicher Weise bearbeitet werden, um sie für die Datenbank nutzen zu können:

Die Sachschlagwörter der SWD, wie sie im Rahmen des Linked Data Services aufbereitet wurden, konnten 1:1 übernommen werden. Es wurden hier auch die CrissCross-Verknüpfungen zur DDC mit umgesetzt. In der Testdatenbank sind alle Sachschlagwörter der SWD enthalten.

Bei der DDC musste eine Versionierung durchgeführt werden, um die verschiedenen Versionen der DDC-Klassen vorliegen zu haben. Die Versionierung der Klassen war der problematischste, jedoch gleichzeitig der entscheidende Schritt, da er eine grundlegende Veränderung der bisherigen Daten bedeutete, die bislang ohne eine Kennzeichnung ihrer zeitlichen Gültigkeit vorlagen. Durch die Bearbeitung der Klasse sind im Editionssys-

tem, in dem die Aktualisierungen durchgeführt werden, zwei Einträge für die Klasse angelegt. Diese Daten werden für die technische Erstellung der zwei Versionen einer Klasse genutzt, was keinen personellen Mehraufwand bedeutet. Die Problematik bestand in der richtigen Zuordnung von Ober- und Unterklassen verschiedener Versionen. Für die verschiedenen Versionen der DDC-Klassen wurde an die bestehende URI das Datum der Änderung angehängt, z.B. <http://d-nb.info/dewey/class/726.79/2005/11/02>. Hierdurch sind in der Datenbank von einigen Klassen zwei, in seltenen Fällen sogar drei Versionen vorhanden. Zu verschiedenen Versionen kommt es, wenn es eine Aktualisierung in einer Klasse gegeben hat. Dann liegt einmal die ursprüngliche Version der Klasse und außerdem die aktualisierte Version der Klasse vor. Wenn es weitere Aktualisierungen gegeben hat, kann es zu weiteren Versionen kommen. Es mussten Beziehungen zwischen der alten und der neuen Version einer Klasse und zwischen Ober- und Unterklassen hergestellt werden. Hierbei war besonders zu beachten, dass die Unterklasse an der richtigen Version der Oberklasse hängt, da es hier meist zwei Möglichkeiten gibt (vgl.: Abb. 2).

In der Testdatenbank ist die gesamte DDC enthalten.

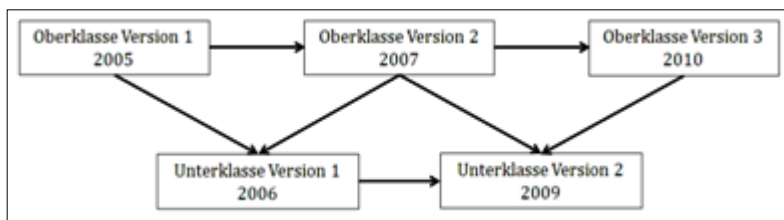


Abb. 2: Mögliche Verbindungen zwischen unterschiedlichen Versionen von Ober- und Unterklassen

Da im Rahmen des Linked Data Services noch keine Titeldaten für das Semantic Web vorlagen, wurden die Titeldaten anhand einiger PICA-Felder umgesetzt. Eine vollständige Repräsentation fand nicht statt. In der Datenbank enthalten sind Titeldaten der Reihen A, B und H der Deutschen Nationalbibliografie, da nur in diesen Reihen ein großer Teil der Dokumente mit der DDC erschlossen wird.

Die Titeldaten bilden das Kernstück der Datenbank, denn sie sollen später recherchierbar sein. Für die Datenbank wurden nur einige Felder mit Hilfe von Dublin Core umgesetzt, wobei die wichtigsten Felder die ID des Datensatzes, der Hauptsachtitel, die DDC-Notation, die WV-Nummer (Wöchentliche Verzeichnisnummer) und die Grundnotation sind.[ILTIS]

Bei der Grundnotation wurde nur das Feld verwendet, in dem die Basis-klasse aufgeführt ist, die weiteren Bestandteile von synthetischen Notation wurden bei der Suche nicht berücksichtigt.

Um feststellen zu können, mit welcher Version einer Klasse ein Dokument verknüpft ist, musste ein Datum gefunden werden, das anzeigt, wann die inhaltliche Erschließung des Dokumentes stattgefunden hat. Anhand der WV-Nummer, dem Erscheinungsdatum des Dokumentes in der Deutschen Nationalbibliografie, kann ein Datum generiert werden, bei dem feststeht, dass bis zu diesem Zeitpunkt die inhaltliche Erschließung in jedem Fall stattgefunden hat. Dieses Datum wurde dazu genutzt, die Dokumente der entsprechenden Version einer DDC-Klasse zuordnen zu können, wobei hier meist nur die richtige von zwei verschiedenen Versionen einer DDC-Klasse gefunden werden muss.

Die Abb. 3 verdeutlicht die Vernetzung der Daten untereinander. Die DDC ist grün, die Titel gelb und die SWD orange dargestellt. Der Graph zeigt einen kleinen Ausschnitt aus der Datenbank. Die Darstellung ist nicht vollständig, da zur besseren Übersicht nur eine Auswahl der in der Datenbank enthaltenen Felder aufgenommen wurde.

Die verschiedenen Versionen einer DDC-Klasse sind ebenso miteinander vernetzt wie Ober- und Unterklassen. Durch die Versionierung kann man eindeutig zuordnen, dass die Version der Klasse 726.79 Teile und Nebengebäude von 2009 die Oberklasse der neuen Klasse 726.796 Kreuzgänge ist.

Auch die Titeldaten können dank der WV-Nummer eindeutig einer Version der Klasse zugeordnet werden. So ist der Titel „Der Kreuzgang der Abtei Altenryf“ mit der Klasse 726.79 Teile und Nebengebäude verknüpft, jedoch mit der alten Version der Klasse, in der Kreuzgänge noch im Einschließlich-Hinweis enthalten waren. Der Titel „Infirmarien – Kranken- und Sterbehäuser der Mönche“ hingegen wurde zu einem späteren Zeitpunkt klassifiziert und ist deshalb mit der neuen Version der Klasse 726.79 verknüpft. Da der Titel „Der Regensburger Domkreuzgang“ erst 2010 erschienen ist, konnte er bereits mit der Klasse 726.796 klassifiziert werden.

Mit diesem Titel ist auch das Schlagwort Kreuzgang, verknüpft, das wiederum mit der Klasse 726.796 verbunden ist. Diese Verbindung entstand im Projekt CrissCross, auf die Modellierung der Verbindung wird hier nicht näher eingegangen.

Der Graph zeigt, wie stark die einzelnen Daten miteinander vernetzt sind. Für die Umsetzung der Daten wurde hauptsächlich das Vokabular des Dublin Core, SKOS (Simple Knowledge Organisation System), verwen-

det sowie eigens für die Umsetzung der DNB-Daten entworfenes Vokabular, das auf SKOS beruht.[LinkedDataService]

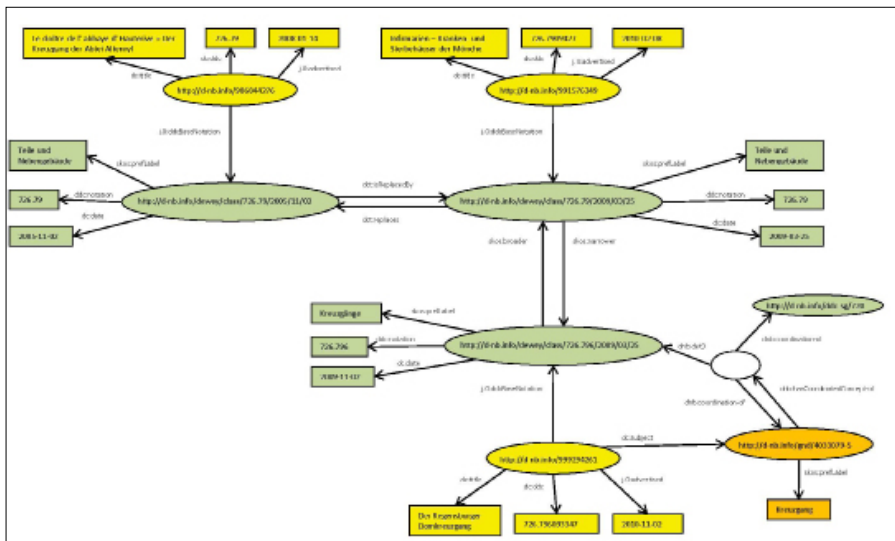


Abb. 3: Beispielgraph aus DDC-Klassen, Titeldaten und einem Sachschlagwort. Zur besseren Übersicht sind die DDC-Klassen grün, die Titeldaten gelb und das Schlagwort orange dargestellt.

5. Auswahl der Testfälle

Um geeignete Testfälle zu finden, ist wichtig zu wissen, welche Änderungen es in der DDC geben kann. Hierfür muss die Struktur und Organisation der Haupttafeln näher betrachtet werden. In den Haupttafeln der DDC gibt es Anmerkungen bei vielen Notationen. Diese Hinweise und Anweisungen können verschiedenen Zwecken dienen. So erklären sie beispielsweise den genauen Begriffsinhalt einer Klasse oder weisen auf Besonderheiten hin. Man kann die Anmerkungen nach ihrer Funktion in vier Gruppen einteilen: Hinweise zur Beschreibung des Klasseninhalts, Hinweise auf Klassifizierung in anderen Klassen, Hinweise zur Erläuterung von Änderungen oder Unregelmäßigkeiten in Haupt- und Hilfstafeln und Anweisungen für die Notationssynthese.[Chan]

Nimmt man alle Arten von Änderungen, die seit der Druckausgabe der DDC 22 Eingang in die elektronische Ausgabe der deutschen DDC gefunden haben, so ergibt sich eine lange Liste. Hierbei gibt es gravierende Än-

derungen wie beispielsweise neue Klassen, aber auch vergleichsweise kleine Änderungen wie Korrekturen in Siehe-auch-Verweisungen.

Nicht jede Änderung in der DDC hat einen Einfluss auf den Klassifizierungsprozess, der klar ersichtlich und direkt nachweisbar ist. Deshalb wurden für die Tests nur Änderungen berücksichtigt, die einen tiefgreifenden Einfluss auf die Klassifizierung neuer Titel haben:

- neue Klasse
- Stilllegung/Teilstilllegung
- Verlegung/Teilverlegung/Verlegungshinweis

Eine Übersicht über die (aktuellen) Änderungen der deutschen DDC-Daten gibt es auf der Seite „DDC Deutsch Update – Melvil-Aktualisierung“. [DDCDeutschUpdate] Anhand dieser Seite wurden die Beispiele gefunden, mit denen gearbeitet wurde. Die Auswahl erfolgte Mitte Oktober 2010, wobei nur mit geänderten Klassen gearbeitet werden konnte, mit denen bereits Titel verknüpft wurden, denn nur so ist eine praktische Suche mit vergleichbaren Ergebnissen möglich. Bei Klassen ohne Titel ist es nicht möglich nachzuweisen, dass die neuen Suchansätze bessere Ergebnisse erzielen als die bisherige Suche.

Im Bereich neue Klassen gab es 64 Änderungen mit Titeln, im Bereich Stilllegung/Verlegung 2 und im Bereich Verlegungen/Teilverlegungen/Verlegungshinweis gab es 21 Änderungen. Im Folgenden werden die neuen Suchansätze am Beispiel einer neuen Klasse dargestellt.

6. Beispielhafte Darstellung der neuen Suche

Exemplarisch für die Änderungsart neue Klassen in der DDC betrachten wir die Klasse 726.796 Kreuzgänge. Hier wurde ein Thema aus dem Einschließlich-Hinweis in eine neue Unterklasse ausgelagert. Bei diesem Beispiel werden eigentlich zwei bzw. drei Änderungen betrachtet. Die neue Klasse 726.796, die Verlegung bzw. Teilverlegung eines Aspektes der Klasse 726.69 und die aus den beiden genannten Änderungen folgende Anpassung des Einschließlich-Hinweises in der Klasse 726.79.

Um beurteilen zu können, ob eine Suche zu besseren Ergebnissen führt, muss man sich einerseits die Klassen genau ansehen und andererseits prüfen, welche Suchergebnisse es bei der bisherigen Suche gibt. Hierzu muss man sich drei Klassen ansehen, auf die das Thema verteilt ist. Diese sind Unterklassen von 726.6 Kathedrahlen und 726.7 Klosteranlagen.

In der Klasse 726.69 waren vor der Einführung der neuen Klasse 726.796 die umfassenden Werke über Kreuzgänge beheimatet. Diese sind

nun nach 726.796 verlegt worden. In der Klasse 726.79 waren die Kreuzgänge Teil des Einschließlich-Hinweises. Aus diesem wurden sie mit der Einführung der Klasse 726.796 gestrichen. Dokumente, die sich mit dem Thema Kreuzgänge als Teil von Klosteranlagen beschäftigen, können somit in diesen drei Klassen enthalten sein.

Klasse	Alte Version der Klasse	Neue Version der Klasse
726.69	726.69 Nebengebäude Hier auch umfassende Werke über Kreuzgänge 2004/11/03	726.69 Nebengebäude Umfassende Werke über Kreuzgänge verlegt nach 726.796 2009/03/30
726.79	726.79 Teile und Nebengebäude Einschließlich: Klausen, Kreuzgänge, Refektorien 2005/11/02	726.79 Teile und Nebengebäude Einschließlich: Klausen, Refektorien 2009/03/25
726.796		726.796 Kreuzgänge Hier auch: umfassende Werke über Kreuzgänge [früher 726.69] 2009/03/25

Tab. 1: Verschiedene Versionen der betroffenen Klassen

Für die Suche nach mit der DDC erschlossenen Dokumenten gibt es zwei Möglichkeiten: Zum einen kann über den Online-Katalog der DNB direkt gesucht werden (ddc=726.796)[KatalogDNB] oder man kann sich bei MelvilSearch durch die Hierarchie der DDC zur richtigen Notation navigieren.[MelvilSearch]

Bei der bisherigen Suche im Online-Katalog der DNB nach der DDC-Klasse 726.796 Kreuzgänge erhält man 3 Treffer.[KatalogDNB] (In der Datenbank sind es nur zwei Treffer, da ein Titel eine Medienkombination

ist und daher nicht im Textbestand. Ein alternativer Suchansatz über MelvilSearch liefert das gleiche Ergebnis.[MelvilSearch] Bei dieser Suche mit der für den Nutzer in diesem Moment präzisesten Notation werden durch die Aktualisierungen in der DDC nicht alle relevanten Titel gefunden, da nicht alle mit der neuen Notation 726.796 erschlossen wurden. Um ein vollständiges Suchergebnis zu bekommen müssen die Klassen, in denen das Thema Kreuzgänge früher beheimatet war, einbezogen werden. Durch die Versionierung ist es möglich, nur einen Teil einer Klasse in das Ergebnis mit einzubeziehen. Von der Klasse 726.79 Teile und Nebengebäude muss die alte Version mit in die Suche einbezogen werden, denn hier waren die Kreuzgänge Teil des Einschließlich-Hinweises. Auch von der Klasse 726.69 Nebengebäude muss die alte Version mit einbezogen werden, da hier die umfassenden Werke über Kreuzgänge eingeordnet waren. Für die neue Klasse 726.796 Kreuzgänge ist eine zeitliche Einschränkung nicht nötig.

Die Einschränkung auf einen Teil der Klasse führt zu einer Reduzierung der Treffermenge. So werden in der Klasse 726.69 nur zwei statt drei Titel gefunden und in der Klasse 726.79 nur drei statt wie bisher fünf. Da mit den Klassen nicht besonders viele Titel verknüpft sind, sind die Unterschiede relativ gering. Wenn aber in Zukunft mehr Dokumente mit den beiden Klassen 726.69 und 726.79 klassifiziert werden, so können diese Titel alle aus der Treffermenge ausgeschlossen werden, da nur der für die Suche relevante Teil der Klasse in die Suche mit einbezogen wird. In den neuen Versionen der beiden Klassen kann durch die neue Klasse 726.796 keine für die Suchanfrage relevante Literatur mehr klassifiziert werden.

Die Suchanfragen in der Testdatenbank wurden alle mit Hilfe von SPARQL (SPARQL Protocol and RDF Query Language) gestellt. Mit SPARQL können semantische Daten abgefragt werden.[SPARQL]

Das Ergebnis einer Suche mit allen Klassen, über die das Thema Kreuzgänge verteilt ist, ist in Tab. 2 zusammengestellt: Hierbei handelt es sich um zwei Treffer aus der alten Version der Klasse 726.69, drei Treffer aus der alten Version der Klasse 726.79 und die beiden Titel, die in der neuen Klasse 726.796 enthalten sind.

Die Spalte Schlagwort zeigt, dass eine Suche mit Hilfe des Schlagwortes Kreuzgang zu keinem zufriedenstellenden Suchergebnis geführt hätte, da nur zwei der Titel mit diesem Schlagwort verknüpft sind. Von den sieben gefundenen Titeln sind jedoch nicht alle relevant – schließlich wurden mit der früheren Version der 726.79 nicht nur Titel über Kreuzgänge klassifiziert. Jedoch kann mit der hier gezeigten Vorgehensweise der Ballast an nicht relevanten Titeln erheblich eingeschränkt werden. Ohne Versi-

onierung wären nicht sieben sondern zehn Titel gefunden worden. Durch die Versionierung ist der Recall im Vergleich zu einer Suche mit der Klasse 726.796 verbessert worden, ohne gleichzeitig eine zu geringe Precision zu erhalten. Je mehr Titel in Zukunft mit den Klassen 726.69 und 726.79 klassifiziert werden, desto größer ist der Nutzen der Versionierung, da zwischen möglichen relevanten Titeln (in den alten Versionen der Klassen) und auf jeden Fall nicht relevanten Titeln (in den neuen Versionen der Klassen) unterschieden werden kann.

7. Weitere Beispiele

Auf die Beispiele der Änderungsart „Stilllegung/Teilstilllegung“ und „Verlegung/Teilverlegung/Verlegungshinweis“ wird an dieser Stelle nur kurz eingegangen.

Stilllegungen gab es bis Oktober 2010 in der DDC-Deutsch erst in zwei Fällen, so beispielsweise im Bereich der Klasse 154.63 Träume. Die Unterklasse 154.634 Traumanalyse wurde stillgelegt und das Thema in die übergeordnete Klasse 154.63 Träume integriert.

Eine Suche mit der Klasse 154.634 Traumanalyse führt, wenn man Literatur zum Thema Traumanalyse finden will, nicht ans Ziel, da hier keine neuere Literatur enthalten ist. Dieses Problem kann man umgehen, indem man mit der Klasse 154.634 in der Version von 2005/11/02 sucht, da sie zu diesem Zeitpunkt noch gültig war. Außerdem kann man in die Suche die Oberklasse 154.63 Träume in der Version von 2009/06/02 einbeziehen, da ab diesem Zeitpunkt Literatur über Traumanalyse in dieser Klasse klassifiziert wurde. So kann vermieden werden, alle Titel aus der Klasse 154.63 Träume mit in der Treffermenge zu haben, da vor 2009 keine relevanten Titel in dieser Klasse enthalten sein können. Auf diese Weise wird der Recall gesteigert ohne gleichzeitig die Precision zu stark zu senken. Außerdem würde sonst aktuelle Literatur zum Thema Traumanalyse aus der Suche komplett ausgeschlossen.

Verlegungen kamen bis Oktober 2010 in der DDC-Deutsch bereits 21 mal vor. Sie beziehen sich meist auf Teilaspekte, die in eine andere Klasse ausgelagert werden, da sie dort thematisch besser passen. Aus der Klasse 845.03 Zeit des Deutschen Kaiser, 962–1122 wurde beispielsweise der Teilaspekt Umfassende Werke über das Mittelalter in die Klasse 945.01 Anfänge bis 774 verlegt.

Um trotz der Änderung umfassend suchen zu können, kann man z.B. die Registereinträge nutzen. Der Registereintrag `Mittelalter` -> `ita-`

ID	Titel	DDC (versioniert)	Schlagwort
<http://d-nb.info/987202316>	„Kreuzgänge – in Stein bewahrte Stille ; ein Reisebegleiter zu den Kreuzgängen in Nord- und Ostdeutschland / Philipp-Kaspar Pfannenstiel“	<http://d-nb.info/dewey/class/726.69/2004/11/30>	„Kreuzgang“@de
<http://d-nb.info/985901578>	„Freiburger Münster – unserer lieben Frau, Universitätskapelle / hrsg. vom Landesamt für Denkmalpflege im Regierungspräsidium Stuttgart. Mit Beitr. von Eberhard Grether ...“	<http://d-nb.info/dewey/class/726.69/2004/11/30>	
<http://d-nb.info/978016572>	„Kloster Wedinghausen – Westflügel / [Hrsg.: Stadt Arnberg, der Bürgermeister ; Stadtarchiv. Übers. ins Engl.: Michael Scuffil]. Michael Gosmann & Gerhard Kalhöfer“	<http://d-nb.info/dewey/class/726.79/2005/11/02>	
<http://d-nb.info/985674156>	„Der Westbau der ehemaligen Abtei Moissac als Beispiel eines Vorhallenturmes / ein mittelalterliches Bauwerk im Spannungsfeld funktionaler Anforderungen / vorgelegt von Günther Kälberer“	<http://d-nb.info/dewey/class/726.79/2005/11/02>	
<http://d-nb.info/986844276>	„Le cloître de l' abbaye d' Hauterive / [éd. Service des Biens Culturels. Réd. François Guex]“	<http://d-nb.info/dewey/class/726.79/2005/11/02>	
<http://d-nb.info/98251798X>	„Mittelalterliche Kreuzgänge in Europa / Rolf Legler“	<http://d-nb.info/dewey/class/726.796/2009/03/25>	
<http://d-nb.info/999294261>	„Der Regensburger Domkreuzgang / Herbert E. Berkle“	<http://d-nb.info/dewey/class/726.796/2009/03/25>	„Kreuzgang“@de

Tab. 2: Ergebnis der Suchanfrage. Gesucht wurde nach den Titeln, die mit den alten Versionen der Klassen 726.69 und 726.79, sowie Titeln, die mit der Klasse 726.796 klassifiziert worden sind.

lienische Geschichte ist von der alten Version der Klasse 945.03 zur neuen Version der Klasse 945.01 verlegt worden. In der Suchanfrage müssen die verschiedenen Versionen in diesem Fall nicht explizit angegeben werden. Es werden durch die Suche nach dem Registerbegriff alle Titel ausgegeben, die mit einer DDC-Klasse klassifiziert wurden, an der der Registerbegriff *Mittelalter* -> *italienische Geschichte* hängt.

Diese beiden Suchansätze zeigen zwei Möglichkeiten zur Verbesserung der Suchergebnisse, die das semantische Netz und die Versionierung bieten. Durch das semantische Netz aus Titeldaten, Schlagwörtern und versionierten DDC-Klassen können die unterschiedlichsten Verknüpfungen genutzt werden: zwischen Schlagwörtern und DDC-Klassen, zwischen verschiedenen Versionen einer DDC-Klasse, zwischen verschiedenen Versionen von DDC-Klassen und ihren Unterklassen, zwischen Titeln und Registerbegriffen und viele weitere. All diese Verbindungen können in die Suche einfließen und diese besser, umfassender und präziser machen.

Aufgrund des bisher geringen Beispielmaterials handelt es sich bei den evaluierten Fällen um Betrachtungen, die nicht repräsentativ sein können. Inwiefern sich der gewählte Suchansatz auf andere Beispiele übertragen lässt, muss weiter untersucht werden. Es wurden noch nicht alle Möglichkeiten, die das semantische Netz bietet, ausgeschöpft, aber die betrachteten Fälle zeigen bereits, dass sich die Suche durch eine Versionierung der DDC verbessert und Precision und Recall sich positiv beeinflussen lassen.

8. Zusammenfassung

Bibliotheken bieten den Zugang zu qualitativ hochwertigen Informationen. Um diesen Zugang zu gewährleisten, ist die Erschließung der Informationen erforderlich. Diese erfolgt auf formaler und inhaltlicher Ebene.

Durch die Vernetzung von verschiedenen Wissenssystemen ist es möglich, die Ergebnisse bei der Suche zu optimieren. Hier spielt die DDC als eines der wichtigsten Wissenssysteme eine entscheidende Rolle. Ein Schritt in Richtung Vernetzung ist die Verbindung von Schlagwörtern und den Klassen der DDC, die im CrissCross-Projekt umgesetzt worden ist. Jedoch hat die DDC ihre Stärke und gleichzeitig ihre Schwäche in den Aktualisierungen. Diese sind nötig, um die Klassifikation aktuell zu halten und eine adäquate inhaltliche Erschließung zu ermöglichen. Andererseits führen sie zu dem zu Beginn vorgestellten Problem: Durch die Aktualisierungen werden Themen an andere Stellen der Klassifikation verlegt. So können Inhalte im Laufe der Zeit über mehrere Klassen verteilt sein, was die Suche nach

eben diesen Inhalten erschwert. Um dieses Problem in den Griff zu bekommen, gibt es die Möglichkeit der Versionierung der DDC, die eine Unterscheidung der verschiedenen Ausprägungen einer Klasse greifbar macht. Wenn es Aktualisierungen in einer Klasse gab, liegen ab diesem Zeitpunkt zwei Versionen der Klasse vor. Für jeden Titel, der klassifiziert wurde, kann nachvollzogen werden, mit welcher Version er klassifiziert wurde. Das kann für eine präzisere Suche genutzt werden.

Die Versionierung wird bisher nicht eingesetzt. Die Ergebnisse haben jedoch gezeigt, dass es möglich ist, mit Hilfe von Versionierung einen besseren Recall bei der Suche zu erreichen. Es wird nicht nur mit einer einzelnen Klasse gesucht, sondern es kann berücksichtigt werden, dass ein Thema über mehrere Klassen verteilt ist, ohne dass man gleich alle Titel der entsprechenden Klassen mit einbeziehen muss. Dadurch kann verhindert werden, dass die Precision zu stark sinkt. Denn ein vollständiges Suchergebnis, mit allen relevanten Titeln, die im Katalog enthalten sind, hilft keinem, wenn gleichzeitig ein sehr hoher Anteil nicht relevanter Titel enthalten ist. Die Daten in der Testdatenbank wurden als Semantisches Netz modelliert. In diesem Netz aus Titeldaten, Schlagwörtern und versionierten DDC-Klassen können die verschiedenen Verbindungen für die Suche genutzt werden. Es ist jedoch schwierig, eine einheitliche Vorgehensweise für die verschiedenen Änderungen in der DDC zu geben. Jede Änderung hat ihre ganz spezifischen Auswirkungen auf den Klassifizierungsprozess. Manche haben einen eindeutigen und klar feststellbaren Einfluss, der sich gut in erweiterte Suchanfragen umsetzen lässt. Andere Änderungen wiederum haben keinen direkten Einfluss auf die Klassifizierung bzw. der Einfluss ist nicht eindeutig und es kommt auf den jeweiligen Einzelfall an. Für eine praktische Umsetzung im Online-Katalog müssen weitere Tests durchgeführt werden, um für die verschiedenen Änderungstypen in der DDC jeweils eine Best-practice-Suchanfrage zu finden, die ein zufriedenstellendes Ergebnis liefert.

Der Ansatz des Semantic Web auch für Titeldaten ist zukunftssträftig. Hier gibt es großes Potential, das weiter ausgebaut und genutzt werden sollte. Insgesamt muss es mehr Vernetzungen zwischen Daten geben. Dies betrifft nicht nur Verknüpfungen zwischen Titeldaten und inhaltlichen Erschließungsmethoden wie der SWD oder der DDC – auch die inhaltlichen Erschließungsmethoden untereinander müssen stärker verknüpft werden, um dem Nutzer eine umfassende Suche zu ermöglichen.

Claudia Effenberger, M.Eng. (Library Science)
Deutsche Nationalbibliothek
E-Mail: c.effenberger@dnb.de

- [Alex2011] Alex, Heidrun (2011): DDC in der Deutschen Nationalbibliografie. Online verfügbar unter <http://www.ddc-deutsch.de/anwendung/dnb.htm>, zuletzt geprüft am 17.05.2011.
- [Alex2009] Alex, Heidrun (2009): DDC in Germany : experiences and perspectives. Online verfügbar unter <http://www.onb.ac.at/events/files/alex.ppt>, zuletzt geprüft am 23.05.2011.
- [Ansorge] Ansorge, Kathrin (2008): Das war 2007. In: Dialog mit Bibliotheken, Jg. 20, H. 1, S. 7–13. Online verfügbar unter http://files.d-nb.de/pdf/dialog_08_01_vollt.pdf, zuletzt geprüft am 10.05.2011.
- [Bee] Bee, Guido (2010): DDC. Aktualisierungen und andere neue Entwicklungen. Online verfügbar unter www.opus-bayern.de/bib-info/volltexte/2010/840/pdf/DDC-Innovationen.pdf, zuletzt geprüft am 08.05.2011.
- [Berners-Lee] Berners-Lee, Tim (2009): Linked Data. Online verfügbar unter <http://www.w3.org/DesignIssues/LinkedData.html>, zuletzt geprüft am 05.05.2011.
- [Boll] Boll, Katharina (2008): From theology to religion ? – Die Klasse 200 der DDC: Aktuelle Entwicklungen und Vergleich mit der Fachsystematik B. Online verfügbar unter <http://www.qucosa.de/fileadmin/data/qucosa/documents/5660/data/boll.pdf>, zuletzt geprüft am 23.05.2011.
- [Chan] Chan, Lois Mai; Mitchell, Joan S. (2006): Dewey-Dezimalklassifikation. Theorie und Praxis ; Lehrbuch zur DDC 22. München: Saur.
- [Crisscross] Crisscross. Online verfügbar unter <http://linux2.fbi.fh-koeln.de/crisscross/index.html>, zuletzt geprüft am 15.05.2011.
- [DDC22] Dewey-Dezimalklassifikation und Register. Begründet von Melvil Dewey. DDC 22. Herausgegeben von Joan S. Mitchell unter Mitwirkung von Julianne Beall, Giles Martin, Winton E. Matthews, Jr. und Gregory R. New. Deutsche Ausgabe. Herausgegeben von Der Deutschen Bibliothek. 4 Bände. K.G. Saur, München 2005, S. xlvii–lxxvii.
- [DDCDeutsch] DDC Deutsch. Online verfügbar unter <http://www.ddc-deutsch.de>, zuletzt geprüft am 17.05.2011.
- [DDCDeutschUpdate] DDC Deutsch Update – Melvil-Aktualisierung (2010). Online verfügbar unter <http://linux2.fbi.fh-koeln.de/ddc-deutsch/updates/uebersicht.html>, zuletzt geprüft am 10.05.2011.
- [Dewey] Dewey, Melvil (1876): A classification and subject index for cataloguing and arranging the books and pamphlets of a library. Amherst, Mass.
- [EDUG] European DDC Users Group: EDUG 340: Law Working Group. Online abrufbar unter <http://www.slainte.org.uk/edug/edug340/edug340index.htm>, zuletzt geprüft am 23.05.2011.

- [Heiner-Freiling] Heiner-Freiling, Magda (2005): RSWK und DDC. Sacherschließung auf zwei Beinen. In: Dialog mit Bibliotheken, Jg. 17, H. 3, S. 4–13. Online verfügbar unter http://www.ddc.deutsch.de/publikationen/pdf/2005_3_rswk_und_ddc.pdf, zuletzt geprüft am 07.05.2011.
- [ILTIS] ILTIS-Handbuch. Teil 1 Systembeschreibung (2010). Feldverzeichnis: DNB-Titeldaten. Online verfügbar unter http://support.d-nb.de/iltis/feldverzeichnis/Titeldaten_dnblf.pdf, zuletzt geprüft am 23.05.2011.
- [Junker] Junker, Christa (2010): Deutsche Nationalbibliografie. Online verfügbar unter <http://www.d-nb.de/service/zd/dnb.htm>, zuletzt geprüft am 07.05.2011.
- [KatalogDNB] Katalog der Deutschen Nationalbibliothek. Online verfügbar unter <https://portal.d-nb.de/>, zuletzt geprüft am 12.05.2011.
- [LinkedDataService] Der Linked Data Service der Deutschen Nationalbibliothek. Online verfügbar unter http://files.d-nb.de/pdf/linked_data.pdf, zuletzt geprüft am 10.05.2011.
- [MelvilSearch] MelvilSearch: Your gateway to classified information. Online verfügbar unter <http://melvil.d-nb.de/melvilsearch?bs=dnb-portal&id=1409024>, zuletzt geprüft am 23.05.2011.
- [OCLC] OCLC: Dewey Services : Updates : New and changed entries. Online abrufbar unter <http://www.oclc.org/dewey/updates/new/default.htm>, zuletzt geprüft am 23.05.2011.
- [RDF] Resource Description Framework (RDF): Concepts and Abstract Syntax (2004). Online verfügbar unter <http://www.w3.org/TR/rdf-concepts/>, zuletzt geprüft am 05.05.2011.
- [Slavicssac] Slavic, Aida; Issac, Antoine (2009): Identifying management issues in networked KOS: examples from classification schemes. Online verfügbar unter http://www.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2009/presentations/slavic_isaac_NKOS2009_06.pdf, zuletzt geprüft am 05.05.2011.
- [SPARQL] SPARQL Query Language for RDF (2008). Online verfügbar unter <http://www.w3.org/TR/2008/REC-rdf-sparql-query-20080115/>, zuletzt geprüft am 05.05.2011.
- [Stock] Stock, Wolfgang G.; Stock, Mechtild (2008): Wissensrepräsentation. Informationen auswerten und bereitstellen. München: Oldenbourg (Einführung in die Informationswissenschaft, 2), hier S.2ff.
- [Svensson] Svensson, Lars G. (2010): Panta rei. Die Versionierung der DDC- Probleme, Anforderungen und mögliche Lösungen. Online verfügbar unter http://www.univie.ac.at/voeb/fileadmin/Dateien/Kommissionen/Sacherschliessung/2010-04-29_Panta_rei.pdf, zuletzt geprüft am 07.05.2011.

[W3CSemWeb] Koivunen, Marja-Riitta Koivunen; Miller, Eric (2001):
W3C Semantic Web Activity. Online verfügbar unter <http://www.w3.org/2001/12/semweb-fin/w3csw>, zuletzt geprüft am 12.05.2011.