



VZG-Projekt Colibri

Bewertung von automatisch DDC-klassifizierten Titeldatensätzen der Deutschen Nationalbibliothek (DNB)

August 2008 - Februar 2009

Ulrike Reiner

VZG-Colibri-Bericht I/2008

Verbundzentrale des
Gemeinsamen Bibliotheksverbundes (VZG)
Platz der Göttinger Sieben 1
D-37073 Göttingen
<http://www.gbv.de/>



Name des Dokuments

colibri05.doc



Fotograf: Florian Vogelbacher (Fundy-Nationalpark, New Brunswick, Kanada, 25. Mai 2008)

Versionen

colibri05-08-08-21.doc	Entwurf
colibri05-08-09-01.doc	Weiterführung
colibri05-09-02-17.doc	Abschluss

Inhalt

1. Einleitung	3
2. Automatische DDC-Klassifizierung und automatische Bewertung	3
2.1 Automatische DDC-Klassifizierung	4
2.1.1 Repräsentation der <i>DDC-Datenbasis</i>	4
2.1.2 Repräsentation der automatisch zu klassifizierenden Titeldatensätze	6
2.1.3 DDC-Klassifizierung der DNB-Titeldatensätze <i>dnb_ABH</i>	7
2.1.4 <i>Sample I</i>	10
2.1.5 Qualität der automatischen DDC-Klassifizierung	11
2.1.6 Heuristische Funktion <i>cutoff_val</i>	12
2.2 Automatische Bewertung	13
2.2.1 Projekt Colibri/DDC	13
2.2.2 OCLC-Projekt Scorpion/DDC	15
2.2.3 Projekt Pfeffer/RVK	17
2.2.4 Vergleichende Betrachtung mit Bewertungsmaßen aus den Projekten 2.2.1-2.2.3	17
3. Kommentierung des DNB-Berichtes „Colibri_Auswertung_DDC_Endbericht_Sommer_2008“ ...	19
4. Weiterführende Betrachtungen und Fortführung	101
4.1 Klassifizierungsergebnisse mit <i>vc_dcl</i> -Programmänderungen	101
4.2 Klassifizierungsergebnisse – erneut betrachtet	103
4.3 Zusammenfassung der Ergebnisse und Fortführung	109



I. Einleitung

Das VZG-Projekt Colibri/DDC beschäftigt sich seit 2003 mit automatischen Verfahren zur Dewey-Dezimalklassifikation (Dewey Decimal Classification, kurz DDC). Ziel des Projektes ist eine einheitliche DDC-Erschließung von bibliografischen Titeldatensätzen und eine Unterstützung der DDC-Expert(inn)en und DDC-Laien, z. B. bei der Analyse und Synthese von DDC-Notationen und deren Qualitätskontrolle und der DDC-basierten Suche.

Der vorliegende Bericht konzentriert sich auf die erste größere automatische DDC-Klassifizierung und erste automatische und intellektuelle Bewertung mit der Klassifizierungskomponente *vc_dcl*¹. Grundlage hierfür waren die von der Deutschen Nationalbibliothek (DNB) im November 2007 zur Verfügung gestellten 25.653 Titeldatensätze (12 Wochen-/Monatslieferungen) der Deutschen Nationalbibliografie der Reihen A, B und H.

Nach Erläuterung der automatischen DDC-Klassifizierung und automatischen Bewertung in Kapitel 2 wird in Kapitel 3 auf den DNB-Bericht „Colibri_Auswertung_DDC_Endbericht_Sommer_2008“ eingegangen. Es werden Sachverhalte geklärt und Fragen gestellt, deren Antworten die Weichen für den Verlauf der weiteren Klassifizierungstests stellen werden. Über das Kapitel 3 hinaus führende weitergehende Betrachtungen und Gedanken zur Fortführung der automatischen DDC-Klassifizierung werden in Kapitel 4 angestellt. Der Bericht dient dem vertieften Verständnis für die automatischen Verfahren.

2. Automatische DDC-Klassifizierung und automatische Bewertung

Im Folgenden wird der derzeitige Stand der automatischen DDC-Klassifizierung und automatischen Bewertung der automatischen DDC-Klassifizierung skizziert.² Modellbegriffe und Bezeichnungen (für das System und seine Komponenten), die im VZG-Projekt Colibri eingeführt wurden, werden im gesamten Bericht [Reiner 2008d = colibri05.doc] kursiv wiedergegeben.

¹ *vc_dcl*: *vzg colibri_ddc classifier*. *vc_dcl* ist eine Komponente des DDC-Suchsystems *vc_ds* (*vzg colibri_ddc search system*). Vgl. z. B. [Reiner 2008b] Ulrike Reiner: Computer-aided Assignment of DDC Numbers. Gehaltener Vortrag auf der “32nd Annual Conference of the German Classification Society. Joint Conference with the British Classification Society and the Dutch/Flemish Classification Society – AG BIB: Subject Indexing 2008: Accept Progress!”. Hamburg, July 16 – 17, 2008. [<http://archiv.tu-chemnitz.de/pub/2008/0150/>], S. 17.

² Veröffentlichungen hierzu: [Reiner 2007] Ulrike Reiner: Automatische Analyse von DDC-Notationen und DDC-Klassifizierung von GVK-Plus-Titeldatensätzen. Workshop zur Dewey-Dezimalklassifikation „DDC-Einsichten und –Aus-sichten 2007“. Göttingen, 1.3.07 [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2836.pdf]; [Reiner 2008b]; [Reiner 2008e = colibri02.doc] Ulrike Reiner: VZG-Projekt Colibri - DDC-Basis: Grundlage der automatischen DDC-Klassifizierung. VZG-Colibri-Bericht 1/2004 (ruhend, bis jetzt unvollendet).



2.1 Automatische DDC-Klassifizierung

Nach Einführung in die Repräsentation der *DDC-Datenbasis* (Kapitel 2.1.1) und in die Repräsentation der automatisch zu klassifizierenden Titeldatensätze (Kapitel 2.1.2) wird die DDC-Klassifizierung der DNB-Titeldatensätze *dnb_ABH* (Kapitel 2.1.3) und die Erstellung der zu evaluierenden Stichprobe *Sample1* (Kapitel 2.1.4) erläutert. In Kapitel 2.1.5 wird auf Aspekte der Qualität der automatischen DDC-Klassifizierung und in Kapitel 2.1.6 auf die heuristische Funktion *cutoff_val* eingegangen.

2.1.1 Repräsentation der *DDC-Datenbasis*

Im Gemeinsamen Verbundkatalog (GVK) des Gemeinsamen Bibliotheksverbundes (GBV) sind ca. 28.2 Mio. bibliografische Titeldatensätze³ enthalten. Eine Teilmenge des GVK (*GVK-DDC*)⁴ umfasst ca. 4.3 Mio. Titeldatensätze⁵, die mindestens eine intellektuell vergebene DDC-Notation enthalten. Die *GVK-DDC*-Titeldatensätze stammen zu 54,7% aus der Library of Congress (LoC), zu 23,8% aus der British National Bibliography (BNB) und zu 21,5% aus nicht rekonstruierbarer Quelle. Jeder Titeldatensatz aus *GVK-DDC* wird vom PICA+-Format⁶ in die Repräsentationsform der *DDC-Datenbasis vc_DB* konvertiert.⁷ Die *DDC-Datenbasis vc_DB* ist eine Menge von DDC-Klassen und eine DDC-Klasse ist eine Menge von Deskriptorwerten. Ein Deskriptorwert ist der Wert eines Deskriptors, hier eine Pica+-Kategorie, deren Werte zur inhaltlichen Charakterisierung beitragen.⁸ Eine DDC-Klasse wird durch eine DDC-Notation und ein *vc_DB*-Datensatz als 3-Tupel $\langle dno, descr, descr_val \rangle$ ⁹ repräsentiert. Irrelevante Deskriptorwerte werden vor der Aufnahme in die *DDC-Datenbasis vc_DB* eliminiert¹⁰. Alle Grossbuchstaben aller Deskriptorwerte werden in ihre Kleinbuchstaben umgewandelt, einige Zeichen von Deskriptorwerten werden transliteriert¹¹ und Sonderzeichen¹² in Deskriptorwerten eliminiert. Weiterhin werden Personennamen in die Form $\langle 028... \rangle$ -*firstname#surname* und Verlegerdaten in die Form $\langle 033A \rangle$ -*publisher@place* gebracht.

³ GVK – Gemeinsamer Verbundkatalog 2.1 [<http://gso.gbv.de/DB=2.1/>] (November 2008)

⁴ 4.3 Mio / 28.2 Mio = ca. 15,3%

⁵ *GVK-DDC* enthält 4.319.357 Titeldatensätze (Januar 2008).

⁶ Pica+ ist das „interne bibliografische Datenformat der CBS und LBS-Software“ [<http://www.gbv.de/wikis/cls/PicaPlus>]

⁷ Vgl. [Reiner 2007], S. 83 und S. 84

⁸ Vgl. [Reiner 2007], S. 22ff. Beispiele für Deskriptoren sind $\langle 028A \rangle$ (1. Verfasser), $\langle 021A \rangle$ (Hauptsachtitel) und $\langle 045Q \rangle$ (Basisklassifikation).

⁹ *dno*: DDC-Notation; *descr*: descriptor (Deskriptor); *descr_val*: descriptor value (Deskriptorwert)

¹⁰ Z. B. „0“, „das“, „more“, „viii“, „zwischen“ und „anyone“. Die als irrelevant betrachtete Menge von Deskriptorwerten (ca. 224, Juni 2008) muß noch erheblich erweitert werden.

¹¹ Z. B.: Á -> Ae; Â -> Oe; á: -> a; öc -> c; äa -> a.

¹² Z. B.:), [, /, ., '.



Außerdem wird die *DDC-Datenbasis* *vc_DB* um die (englischen¹³) *Fakten* aus der *DDC-Wissensbasis* *vc_KB* erweitert. Die Gesamtmenge aller DDC-Klassen aus *vc_DB* und *vc_KB* wird mit *vc_DB_PLUS* bezeichnet.¹⁴ Als letzter Schritt wird *vc_DB_PLUS* für die automatische Klassifizierung invertiert und wird *vc_IDB* (für *Inverted DataBase*) genannt. Ein *vc_IDB*-Datensatz ist ein *n*-Tupel $\langle descr_val \checkmark dno_1 dno_2 dno_3 dno_4 \dots dno_m \rangle$ mit \checkmark (Pica oct \237, Dutch florin¹⁵) als Trennzeichen zwischen *descr_val* und den *m* DDC-Notationen, die *m* DDC-Klassen repräsentieren. In Abb. 1 ist ein Auszug aus der invertierten Datenbasis *vc_IDB* dargestellt. Er enthält in jeder Zeile einen mit *bgp* beginnenden Deskriptorwert, gefolgt von \checkmark und einer (mit Leerzeichen getrennten) Menge von *vc_IDB*-DDC-Klassen (durch DDC-Notationen repräsentiert):

```
bgp computer network protocol\checkmark004.62 004.6
bgp-4\checkmark004.62
bgp4\checkmark004.62
bgp\checkmark004.62 004.678 004.6
bgp\checkmark401.41 411 414.6 414
bgp\checkmark728
```

Abb. 1: Auszug aus der invertierten *DDC-Datenbasis* *vc_IDB* (417 MB, April - Oktober 2008)

In der invertierten Datenbasis *vc_IDB* (April 2008) sind nur DDC-Notationen (*schedule notations*) enthalten, die in den DDC-Haupttafeln (*schedules*) verzeichnet sind. Deshalb werden die *molekularen DDC-Notationen*¹⁶ (von *GVK-DDC*-Titeldatensätzen), die in den DDC-Haupttafeln nicht enthalten sind, von rechts nach links soweit gekürzt, bis sie in einer der DDC-Haupttafeln enthalten sind. Deskriptorwerte von *GVK-DDC*-Titeldatensätzen – die *molekulare DDC-Notationen* enthalten, die nicht in den DDC-Haupttafeln vorkommen – werden bei den Klassen der in den Haupttafeln enthaltenen DDC-Notationen (*schedule notations*, kurz: *schednos*) abgespeichert.¹⁷

¹³ Die Erweiterung um deutsche *Fakten*, d. h. deren deutsche Deskriptorwerte, wird angestrebt (vgl. Kapitel 3, DNBI-5, 3.). Im Kontext der *DDC-Daten-* und *Wissensbasis* wird von *Fakten* und *Regeln* gesprochen. Beispiel für einen *Fakt*: Die Klassenbenennung von „004.678“ ist „Internet (World Wide Web)“. Beispiel für eine *Regel*: Wenn es sich um die DDC-Notation „004.678“ handelt, ist es erlaubt, die DDC-Notation „TI--02“ aus der Hilfstafel I an die DDC-Notation „004.678“ anzufügen.

¹⁴ Dies sind 590.120 *vc_DB*-DDC-Notationen + 118.792 *vc_KB*-DDC-Notationen (Januar 2008) vgl. auch [Reiner 2007], S. 47, 48, 50, 80-82 (Juli 2004: 466.134 *vc_DB*-DDC-Notationen)

¹⁵ [http://oclpica.org/content/1399/pdf/cbs21_conversiontable.pdf]

¹⁶ Eine *molekulare DDC-Notation* (*dno_mol*) ist eine – in *atomare DDC-Notationen* – syntaktisch zerlegbare Zeichenkette.

¹⁷ Es wurde zunächst angenommen, dass DDC-Klassen mit *molekularen DDC-Notationen*, die nicht in einer der Haupttafeln enthalten sind, für die automatische Klassifizierung zuviel spezielle und ggf. irreführende Information enthalten. Ausserdem war diese Einschränkung aufgrund der damaligen, geringeren Rechnerressourcen notwendig. Ob diese Annahme sinnvoll ist, muss durch weitere Klassifizierungstests überprüft werden (vgl. *vc_dcl* / *vc_IDB*-Version von November 2008, DNB-K060 und Kap. 4.1, VC3.). Falls *dno_j* ungleich *schedno*, dient *schedno* als Vergleichs-*dno_j* zur *CP*/*CN*-Bestimmung



Beispiel 1: Die *molekulare DDC-Notation* 004.6780246171 enthält folgende *atomare DDC-Notationen*¹⁸:

```
004.6780246171 <ull_to_analyze; length: 14>
0----- Computer science, information & general works <hatzen>
00----- Computer science, knowledge & systems <hatzen>
004----- Data processing Computer science <hat>
004.6----- Interfacing and communications <hat>
004.67----- Wide-area networks <hat>
004.678----- Internet (World Wide Web) <hat>
---.---02----- Miscellany <T1--02>
---.---024----- The subject for persons in specific occupations <T1--024>
---.---0246---- Technologists--works for <T1--0246>
---.---6----- Technology <na2r7:6>
---.---61----- Medicine & health <na2r7:61>
---.---617- Miscellaneous branches of medicine Surgery <na2r7:617>
---.---6171 Surgery <na2r7:617.1>
```

Abb. 2: *vc_daygram* der *molekularen DDC-Notation* 004.6780246171 mit den *atomaren DDC-Notationen* 0, 00, 004, 004.6, 004.67, 004.678, T1--02, T1--024, T1--0246, 6, 61, 617 und 617.1¹⁹

Wie oben beschrieben, wird die *molekulare DDC-Notation* 004.6780246171 – da sie nicht in einer der Haupttafeln enthalten ist – bis zu der in Haupttafeln explizit aufgeführten DDC-Notation 004.678 gekürzt und die zu bibliografischen Titeldatensätzen mit der DDC-Notation 004.6780246171 gehörenden Deskriptorwerte werden bei der Klasse 004.678 mit der Klassenbenennung „Internet (World Wide Web)“ in *vc_IDB* abgespeichert (s. auch Bewertungsmaß S7 in Kapitel 2.2.2).

2.1.2 Repräsentation der automatisch zu klassifizierenden Titeldatensätze

GVK-Titeldatensätze (ohne DDC-Notation) werden wie GVK-DDC-Titeldatensätze von dem Format PICA+ in die *vc_DB*-Repräsentation konvertiert. Anstelle einer konkreten DDC-Notation steht die Zeichenkette „XXX“. Andere bibliografische Titeldatensätze (als GVK-Titeldatensätze) werden auf analoge Weise konvertiert. So werden z. B. die zu klassifizierenden DNB-Titeldatensätze (s. Kapitel 2.1.3) vom MAB2-Format in die *vc_DB*-Repräsentation (s. Abb. 3 und Abb. 4) konvertiert. Wegen des Deskriptorwertevergleichs beim automatischen Klassifizierungsverfahren werden die Deskriptorwerte der zu klassifizierenden Titeldatensätze genauso wie die Deskriptorwerte der GVK-DDC-Titeldatensätze transliteriert, in Kleinschreibung transformiert und es werden Sonderzeichen von Deskriptorwerten und irrelevante Deskriptorwerte eliminiert.

¹⁸ Eine *atomare DDC-Notation* (*dno_atom*) ist eine semantisch unzerlegbare Zeichenkette, die eine DDC-Klasse repräsentiert.

¹⁹ Vgl. [Reiner 2008a] Automatic Analysis of Dewey Decimal Classification Notations. In: “Data Analysis, Machine Learning and Applications” (Hrsg. C. Preisach; H. Burkhardt; L. Schmidt-Thieme; R. Decker). Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7-9, 2007. Studies in Classification, Data Analysis, and Knowledge Organization. Springer, Berlin, 2008, p. 697-704. [<http://www.springerlink.com/index/1044082243v117u6.pdf>]



deren Anzahl (absolute Häufigkeiten: Anzahl, in denen der jeweilige Deskriptorwert vorkommt) und DDC-Notationskandidaten ermittelt.²²

```

004.62<026>DDB0981454445
004.62<100>olaf#maennel
331_line Handling the complexity of BGP
335_line via characterization, testing and configuration management
004.62<331>bgp
004.62<331>handling
004.62<331>complexity
004.62<335>configuration
004.62<335>management
004.62<335>via
004.62<335>characterization
004.62<335>testing
700_line of DDB981454445: {004}
705_line of DDB981454445: {004.62}

```

Abb. 4: DDB-Titeldatensatz *in_dnb_DDB0981454445* in *vc_DB*-Repräsentation

So kommt – wie aus Abb. 1 ersichtlich – der Deskriptorwert *bgp* in den acht *vc_IDB-DDC-Klassen* 004.6, 004.62, 004.678, 401.41, 411, 414, 414.6 und 728 vor. Analog kommen z. B. der Deskriptorwert *olaf#maennel* in keiner *vc_IDB-DDC-Klasse*, der Deskriptorwert *configuration* in 109 *vc_IDB-DDC-Klassen* (u. a. in 004.6, 004.62, 004.678) und der Deskriptorwert *testing* in 2734 *vc_IDB-DDC-Klassen* vor. In den *vc_dcl*-Ergebnissen wird u. a. diese Information in der Zeile *considered descriptor values* in der Form „<Deskriptor>-Deskriptorwert [Anzahl unterschiedliche Klassen, in denen er vorkommt], ..., <Deskriptor>-Deskriptorwert [Anzahl unterschiedliche Klassen, in denen er vorkommt]“ angegeben. Für den Titeldatensatz *dnb_DDB981454445* ist das Ergebnis der

complexity, configuration, management, via, characterization, testing}. Hier fällt auf, dass der Deskriptorwert „via“ entfernt werden könnte. Wegen des Wertes „109“ (Abb. 5) der heuristischen Funktion *cutoff_val_dyn* (s. Kap. 2.1.6) wird der Deskriptorwert für die Klassifizierung (d. h. zur Bestimmung eines oder mehrerer DDC-Notationskandidaten) nicht weiter berücksichtigt.

²² Mit dem Vektorprodukt als Ähnlichkeitsmaß (*s* imilarity measure): $s_{uc} = \sum_{i=1}^{L_a} u_i c_i$; für $\forall u \in vc_DB_PLUS$

(*vc_dcl*, April 2008). Bei den beiden (binären) Vektoren *u* und *c* (Länge $L_a :=$ Anzahl der Deskriptorwerte des zu klassifizierenden Titeldatensatzes *u*) repräsentiert das *i*-te Vektorelement u_i und c_i das Gewicht des *i*-ten Deskriptorwertes eines jeden Vektors. *u* steht für den Vektor *unclassified* (ein unklassifizierter Titeldatensatz, der durch eine Menge von Deskriptorwerten repräsentiert wird) und *c* steht für den Vektor *classified* (eine aus GVK-DDC-Titeldatensätzen entstandene DDC-Klasse aus *vc_DB_PLUS*, die auch durch eine Menge von Deskriptorwerten repräsentiert wird). Ein Gewicht von 1 bzw. 0 bedeutet, dass der Deskriptorwert im Vektor bzw. nicht im Vektor vorhanden ist. DDC-Notationskandidat für einen Titeldatensatz ist der, bei dem der größte Ähnlichkeitswert $s_{uc\ max}$ erzielt wird. Bei gleichen Ähnlichkeitswerten unterschiedlicher DDC-Notationskandidaten werden alle als Menge von DDC-Notationskandidaten (*DNO_CAND_SET*) ausgegeben. Der Vektor des DDB-Titeldatensatzes *in_dnb_DDB0981454445* hat die Länge $L_a = 9$. Durch die heuristische Funktion *cutoff_val_dyn* wird die Menge der in Betracht kommenden Deskriptorwerte und damit die Anzahl L_a reduziert, im DDB-Titeldatensatz *in_dnb_DDB0981454445* auf $L_a = 2$ (s. u.).



automatischen Klassifizierung in Abb. 5 wiedergegeben. In der 1. Zeile *identifizier (dno,schedno)* steht die regionale Identifikationsnummer aus MAB2-Feld 026, danach in runden Klammern die intellektuell vergebene DDC-(ähnliche)-Notation und nach dem Komma folgend die in den Haupttafeln enthaltene DDC-Notation (*schedno*, vgl. Kapitel 2.1.1, Fußnote 17). In der 2. Zeile steht/steht die DNB-DDC-Notation/en aus MAB2-Feld 700:

```

identifizier (dno,schedno):          DNB0981454445 (004.62,004.62)
DNB DDC notation (MAB2 field 700)   {004}
calculated cutoff value:            109
title:                              Handling the complexity of BGP
title (remainder):                  via characterization, testing and configuration
management
considered descriptor values:        |9| {<100>-olaf#maennel[0], <331>-bgp[8],
<331>-handling[807], <331>-complexity[408], <335>-configuration[109], <335>-
management[7088], <335>-via[831], <335>-characterization[496], <335>-testing[2734]}
matched descriptor values:           |2| {bgp, configuration}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |3| {004.6, 004.62, 004.678}
calculated1 ddc classes (sections):  |1| {004}
calculated1 ddc classes (main):      |1| {000}
calculated2 ddc classes (subdiv):    {004.62[1], 004.678[1], 004.6[1]}
calculated2 ddc classes (sections):  {004[3]}
calculated2 ddc classes (divisions): {000[3]}
calculated2 ddc classes (main):      {000[3]}
correlation (DNB0981454445,004.62): 111.11x xxxx xxx (1)

```

Abb. 5: *vc_dcl*-Ergebnis des DNB-Titeldatensatzes in *dnb_DN0B981454445* (April 2008)

Aus der 3. Zeile (*calculated cutoff value*) ist zu erkennen, daß die heuristische Funktion *cutoff_val_dyn* den Wert 109 ermittelt hat, wodurch nur Deskriptorwerte, die in höchstens 109 unterschiedlichen *vc_IDB-DDC-Klassen* vorkommen, für die automatische Klassifizierung in Betracht gezogen werden. Somit verbleiben die beiden Deskriptorwerte *bgp* und *configuration*, die in der Zeile *matched descriptor values* stehen. Vor der Menge der Deskriptorwerte {*bgp, configuration*} steht in |...| die Anzahl der Deskriptorwerte (hier: 2), die für die automatische Klassifizierung bei diesem DNB-Titeldatensatz zur Verfügung stehen. Die Deskriptorwerte *bgp* und *configuration* kommen beide gemeinsam in den drei *vc_IDB-DDC-Klassen* 004.6, 004.62 und 004.678 vor. Alle drei *vc_IDB-DDC-Klassen* werden als gleichwertige DDC-Notationskandidaten aufgefasst (in Zeile *max. match value of matched descriptor values* steht der max. erreichbare Ähnlichkeitswert $suc_{max}=2$ wiederum in |...|²³); sie sind in den beiden Ergebnisausgaben *calculated1 ddc classes*²⁴ (*subdiv*) und *calculated2 ddc classes*²⁵ (*subdiv*) aufgeführt. *calc1* und *calc2* sind zwei unterschiedliche Ergebnisausgaben, weitere sind denkbar. In *calc1* werden zusätzlich zu den Unterklassen (*subdiv*) die DDC-Klassen der 3. Ebene (*sections*) durch

²³ Dass die Anzahl von *matched descriptor values*: |2| und *max. match value of matched descriptor values*: |2| hier gleich ist, ist ein Spezialfall. Es kann z. B. vorkommen: Anzahl von *matched descriptor values*: |6| und *max. match value of matched descriptor values*: |3|.

²⁴ kurz: *calc1*

²⁵ kurz: *calc2*



Kürzen der DDC-Notationskandidaten auf drei Stellen und die DDC-Klassen der 1. Ebene (*main*) durch Kürzen der DDC-Notationskandidaten auf eine Stelle (und Anfügen von zwei Nullen) ausgegeben. Davor steht in [...] die Anzahl der DDC-Notationskandidaten. In *calc2* werden die DDC-Notationskandidaten in Abhängigkeit der Anzahl ihres Vorkommens (Häufigkeiten in eckiger Klammer angegeben) pro Ebene aggregiert. Der DDC-Notationskandidat 004 (Ebene 3) tritt in drei DDC-Klassen auf, ebenso häufig der DDC-Notationskandidat 000 in den DDC-Klassen der 2. und 1. Ebene. Die letzte Zeile der Abb. 5 betrifft die automatische Bewertung (s. Kapitel 2.2.1), sie wird in *vc_dcl* (2008) mit *calc2*-DDC-Notationskandidaten vorgenommen.

2.1.4 *Samplel*

Von den oben erwähnten 25.653 Titeldatensätzen *dnb_ABH* enthalten 1.011 Titeldatensätze keine DDC-ähnliche-Notation (MAB2-Feld 700) und keine DDC-Notation (MAB2 Feld 705) und keinen Titel (MAB2-Felder 331, 335, 341 und 451). Deshalb sind lediglich $25.653 - 1.011 = 24.642$ Titeldatensätze automatisch klassifiziert worden, die sich wie folgt verteilen:

Anz. Klassifizierungsergebnisse													
Reihe	dno bzw. dno_like										sdno	alle	
	0	1	2	3	4	5	6	7	8	9			0-9
dnb_A	345	475	601	1852	170	270	1657	911	1784	627	8692	1813	10505
dnb_B	71	31	85	1027	17	92	395	253	82	162	2215	181	2396
dnb_H	232	211	215	1763	110	2865	5779	220	196	123	11714	27	11741
Summen	648	717	901	4642	297	3227	7831	1384	2062	912	22621	2021	24642

Abb. 6: Verteilung der Titeldatensätze *dnb_ABH* nach DDC-(ähnlichen)-Notationen (April 2008)

Aus Abb. 6 ist abzulesen, dass es z. B. 31 Klassifizierungsergebnisse in Reihe B gibt, die mit einer DDC-(ähnlichen)-Notation 1 beginnen oder dass es 5.779 Klassifizierungsergebnisse in Reihe H gibt, die mit einer DDC-(ähnlichen)-Notation 6 beginnen. 2.021 Titeldatensätze enthalten keine DDC-(ähnliche)-Notation (in Spalte *sdno* für „schedule notation“ mit den Werten KMSX), sondern enthalten ein „K“, „S“, „X“ oder „M“ (für miscellaneous: verschiedene andere Zeichen). In diesen Fällen kann keine automatische Bewertung vorgenommen werden, am Ende der *correlation*-Zeile (letzte Zeile in Abb. 5) steht dann (x) (s. Kapitel 2.2.1).

Da die Menge der Klassifizierungsergebnisse für eine intellektuelle Bewertung zu gross ist, wurde die repräsentative Stichprobe *Samplel* aus der Grundgesamtheit (24.642 Klassifizierungsergebnisse, vgl. Abb. 6) erstellt und unter [<https://wiki.d-nb.de/display/EDUG/VZG+Project+Colibri+DDC>] der DNB zur Verfügung gestellt. *Samplel* enthält – auf DNB-Wunsch – 1000 Klassifizierungsergebnisse aus den zehn DDC-Hauptklassen mit der Verteilung 20% : 40% : 40% (Reihe A : Reihe B : Reihe H).



Wie aus Abb. 6 zu erkennen, war die gewünschte Verteilung nicht in allen Fällen möglich: in Reihe B waren nur 31 (mit 1 beginnende DDC-(ähnliche)-Notationen) bzw. 17 (mit 4 beginnende DDC-(ähnliche)-Notationen) klassifizierte Titeldatensätze in der Grundgesamtheit *dnb_ABH* vorhanden. Um insgesamt 1000 Klassifizierungsergebnisse zur Verfügung stellen zu können, wurden aus der Reihe H mehr als 40, nämlich 49 (mit 1 beginnende DDC-(ähnliche)-Notationen) bzw. 63 (mit 4 beginnende DDC-(ähnliche)-Notationen) Klassifizierungsergebnisse für *Sample1* ausgewählt. Somit enthält *Sample1* 200 Reihe-A-Klassifizierungsergebnisse, 368 Reihe-B-Klassifizierungsergebnisse und 432 Reihe-H-Klassifizierungsergebnisse. Eine geringfügig veränderte Verteilung der Grundgesamtheit ergibt sich nach einer Modifikation des Programmes *vc_dcl*²⁶ (corr X := X -Werte in der *correlation*-Zeile):

Anz. Klassifizierungsergebnisse													
Reihe	dno bzw. dno_like										corr		
	0	1	2	3	4	5	6	7	8	9	0-9	X	alle
dnb_A	343	466	614	1853	172	282	1639	906	1786	632	8693	1812	10505
dnb_B	71	33	85	1028	17	92	396	247	82	163	2214	182	2396
dnb_H	233	212	215	1761	111	2894	5747	227	198	118	11716	25	11741
Summen	647	711	914	4642	300	3268	7782	1380	2066	913	22623	2019	24642

Abb. 7: Verteilung der Titeldatensätze *dnb_ABH* nach DDC-(ähnlichen)-Notationen (Oktober 2008)

Auf die durch die Programmänderung *vc_dcl* (Fußnote 26) veränderten Klassifizierungsergebnisse aus Abb. 7 wird in Kapitel 4 eingegangen.

2.1.5 Qualität der automatischen DDC-Klassifizierung

Die Qualität der automatischen DDC-Klassifizierung ist abhängig von der²⁷

- *QIDB*: Qualität der Daten in *vc_IDB* (Anzahl und Qualität der Deskriptorwerte)
- *QTIT*: Qualität der zu klassifizierenden Titeldatensätze (Anzahl und Qualität der Deskriptorwerte)
- *QSIM*: dem verwendeten Ähnlichkeitsmaß (Vektorprodukt, vgl. Fußnote 22)
- *QCUT*: der heuristischen Funktion *cutoff_val* (vgl. Kapitel 2.1.6)
- *QELIM*: Qualität der Funktion *FUELIM* zur Eliminierung irrelevanter Deskriptorwerte (Fußnote 10)
- *QCALC*: der Klassenaggregation der DDC-Notationskandidaten (*calc1* und *calc2*, vgl. Kapitel 2.1.3)

²⁶ Bei der *vc_dcl*-Programmversion (April 2008) wurde, wenn ein DNB-Titeldatensatz mehrere intellektuell vergebene DDC-Notationen enthielt, (versehentlich nur) die letzte – und nicht die erste – DDC-Notation berücksichtigt (wie es auch z. T. in den Kommentaren in den Excel-Dateien zu [Jahns/Mödden 2008] festgestellt wurde). Dies wurde in der *vc_dcl*-Programmversion (Oktober 2008) geändert: kommen bei einem DNB-Titeldatensatz mehrere MAB2-Felder „705“ hintereinander vor, wird nun angenommen, dass der Wert im 1. MAB2-Feld „705“ die höchste Priorität hat. Die automatisch ermittelten Bewertungsergebnisse verändern sich in diesen Fällen.

²⁷ Vgl. auch „Automatic classification obstacles“ [Reiner 2008b], S. 36 und S. 37



2.1.6 Heuristische Funktion *cutoff_val*

Für die Konstruktion von *cutoff_val* wird von folgender Annahme ausgegangen: Deskriptorwerte, die in zu vielen DDC-Klassen auftreten, sind für die automatische DDC-Klassifizierung ungeeignet (bis auf bestimmte Ausnahmen, s. u. Beispiel 4, H3.). Deshalb wird mit *cutoff_val* eine Obergrenze der Klassenhäufigkeiten bestimmt, bis zu der Deskriptorwerte bei der automatischen Klassifizierung von der Klassifizierungskomponente *vc_dcl* berücksichtigt werden. *cutoff_val* ist eine Funktion, die jeder sortierten Liste *S_list*²⁸ mit aufsteigenden (absoluten) Klassenhäufigkeiten der Deskriptorwerte eines zu klassifizierenden Titeldatensatzes einen konstanten Wert *cutoff_val_stat* oder einen dynamisch ermittelten Wert *cutoff_val_dyn*²⁹ zuordnet. Der Wert *cutoff_val_dyn* wird bei jedem zu klassifizierenden Titeldatensatz durch mehrere heuristische Regeln in Abhängigkeit von der ermittelten Klassenhäufigkeitsverteilung der Deskriptorwerte des zu klassifizierenden Titeldatensatzes neu bestimmt. In der Klassifizierungskomponente *vc_dcl* (realisiert durch die *gawk*³⁰-Funktion *fu_get_cutoff_val()*, kurz: *FUCUTVAL*) wird zunächst die Anzahl der Klassen mit den Häufigkeitswerten „0“ sowie die Anzahl der 1-, 2-, 3- und mehr als 3-stelligen Klassenhäufigkeiten bestimmt. Diese werden u. a. in den heuristischen Regeln verwendet. Weiterhin ist die Anzahl der bei der automatischen Klassifizierung zu berücksichtigenden Deskriptorwerte (*in_descr_val_lim*) in *vc_dcl* (2008) zunächst auf sechs gesetzt. Mit diesem Wert *in_descr_val_lim* sowie mit den heuristischen Regeln, die die Anzahl der zu berücksichtigenden Deskriptorwerte ggf. auf mehr als sechs Deskriptorwerte erhöht, muss weiter experimentiert werden – ausschlaggebend für die Qualität der automatischen Klassifizierung. In *FUCUTVAL* nimmt *cutoff_val_dyn* zur *vc_dcl*-Laufzeit unterschiedliche Werte (vom Anfangs- bis zum endgültigen Wert) an.

Es folgen drei Beispiele (Beispiel2 - Beispiel4) für in der *gawk*-Funktion *FUCUTVAL* verwendete heuristische Regeln H1.-H3.:

Beispiel 2: H1. Wenn es (mind.) eine Klassenhäufigkeit mit dem Wert „0“ gibt, dann ist der Anfangswert *cutoff_val_dyn* = Klassenhäufigkeitswert des i-ten Deskriptorwertes ($i = in_descr_val_lim + \text{Anzahl der Häufigkeiten mit dem Wert „0“}$), ansonsten ist *cutoff_val_dyn* = Klassenhäufigkeitswert des j-ten Deskriptorwertes ($j = in_descr_val_lim$).

Beispiel 3: H2. Wenn die Differenz zwischen zwei Häufigkeitswerten größer als „200“ ist, dann ist *cutoff_val_dyn* = kleinerer Wert der beiden Häufigkeitswerte (Berücksichtigung spezifischer Begriffe).

²⁸ $S_list := S_list_1 - S_list_2 - S_list_3 - \dots - S_list_{La}$, mit den Listenelementen $S_list_1, S_list_2, \dots, S_list_{La}$ und „-“ als Trennzeichen zwischen den Listenelementen.

²⁹ $0 \geq cutoff_val_dyn \leq S_list_{La}$

³⁰ "Gawk is the GNU Project's implementation of the AWK programming language" (man page of *gawk*). Die in *vc_dcl* (2008) verwendete *gawk*-Version ist 3.1.5.



Beispiel 4: H3. Wenn die Summe der 1- bis 3-stelligen Häufigkeiten kleiner ist als die Anzahl der größer als 3-stelligen Häufigkeiten, dann ist $cutoff_val_dyn = \text{Wert von } S_list_{La}$ (Berücksichtigung auch allgemeiner Begriffe, da allgemeine Begriffe im Titeldatensatz überwiegen).

Für den zu klassifizierenden DNB-Titeldatensatz $in_dnb_DNB0981454445$ ist $S_list = 0-8-109-408-496-807-831-2734-7088$ (mit $La = 9$ Listenelementen). Nach H1. ist der Anfangswert von $cutoff_val_dyn = 831$, da der 1. Deskriptorwert (olaf#maennel) die Klassenhäufigkeit 0 besitzt (d. h. olaf#maennel kommt in keiner $vc_IDB-DDC$ -Klasse vor). Zwischen dem Häufigkeitswert 408 und 109 besteht eine Differenz von 299. Nach H2. ist dann $cutoff_val_dyn = 109$. Die H3.-Regel (und andere heuristische Regeln in $FUCUTVAL$) sind nicht anwendbar und deshalb liefert $FUCUTVAL$ für die automatische Klassifizierung des DNB-Titeldatensatzes $in_dnb_DNB0981454445$ den Wert $cutoff_val_dyn = 109$ (wie in Abb. 5 in 3. Zeile *calculated cutoff value*) abzulesen.

Der Wert der heuristischen Funktion ($cutoff_val$) kann (zum Zeitpunkt des Serverstartes der Klassifizierungskomponente vc_dcl_srv) statisch festgelegt (Aufrufparameter: $cutoff_val_stat = c$, mit c : konstanter Wert) oder dynamisch (Aufrufparameter: $cutoff_val_stat = -$, mit $-$: kein Wert) ermittelt werden. Zu Testzwecken ist es nützlich, mit konstanten (= statischen) Werten der heuristischen Funktion ($cutoff_val_stat$) zu experimentieren. Der dynamische Wert $cutoff_val_dyn$ wird zur Laufzeit des Clients der Klassifizierungskomponente (vc_dcl_cli) - wie oben beschrieben - ermittelt.

2.2 Automatische Bewertung

2.2.1 Projekt Colibri/DDC

Zur Messung der Qualität der automatischen Klassifizierung wird eine automatische Bewertung vorgenommen bzw. eine automatische Bewertung ist bei sehr grossen Datenmengen wie mehreren Mio. Klassifizierungsergebnissen unumgänglich. Für die automatische Bewertung wurden das Korrelationsmuster CP (*Correlation Pattern*) und das Korrelationsmaß CN (*Correlation Number*) eingeführt und angewendet³¹ (s. Kapitel 3, DNB2-4, 1.). In CP und CN wird stellenweise von links nach rechts ein Ziffernvergleich zwischen dno_j (DDC-Notation, intellektuell vergeben) bzw. (falls keine dno_j verfügbar) dno_like_j (intellektuell vergebene DDC-ähnliche-Notation) und dno_a ³² (DDC-

³¹ [Reiner 2008c] Reiner, Ulrike: DDC-based Search in the Data of the German National Bibliography. In: New Perspectives in Subject Indexing and Classification – Essays in Honour of Magda Heiner-Freiling. Frankfurt a.M., 2008, S.126 und S. 127.

³² Das Ergebnis einer automatischen DDC-Klassifizierung mit der Klassifizierungskomponente vc_dcl ist eine Menge von DDC-Klassifizierungskandidaten $DNO_CAND_SET = \{dno_a1, dno_a2, \dots, dno_an\}$. Der Vergleich dno_j (bzw. dno_like_j) wird max. mit jedem Element dno_ai ($i = 1, \dots, n$) aus DNO_CAND_SET aus $calc2$ (s. Fußnote 22) durchgeführt.



Notation, automatisch vergeben) durchgeführt. Grundannahme ist, dass die intellektuell vergebene DDC-Notation die optimale DDC-Notation ist. In CP wird das Ergebnis als Muster (pattern) notiert:

$$CP := v_1 v_2 v_3 \cdot v_5 v_6 v_7 \ v_8 v_9 v_{10} \ v_{11} v_{12} v_{13} \ v_{14} v_{15} v_{16} \quad {}^{33}$$

mit v_s (values, Werten) an jeder Stelle $s = 1, 2, \dots, s_{\max}$ und $s_{\max} = 16$,

$$v_s(dno_i_s, dno_a_s) := \begin{cases} \cdot, & \text{falls } s = 4 \text{ (Deweypunkt)} \\ 1, & \text{falls } (dno_i_s = dno_a_s \wedge (\neg \exists v_t (v_t = 0))) \text{ für } t < s \\ x, & \text{falls } \neg \exists dno_i_s \\ 0, & \text{sonst} \end{cases}$$

mit dno_i_s (Ziffer von dno_i an Stelle s) und dno_a_s (Ziffer von dno_a an Stelle s).

Beispiele sind $CP = III.III \ 100 \ 0xx \ xxx$, $CP = I10.0xx \ xxx \ xxx \ xxx$ oder $CP = III.xxx \ xxx \ xxx \ xxx$.

Ein Spezialfall ist $CP = xxx.xxx \ xxx \ xxx \ xxx$, was bedeutet, dass keine intellektuelle DDC-(ähnliche)-Notation zur automatischen Bewertung zur Verfügung steht. In solchen Fällen steht in der *correlation*-Zeile (letzte Zeile einer jeden Ergebnisausgabe der automatischen Klassifizierung, vgl. Abb. 5) anstelle eines CN -Wertes (s.u.) ein (x). Auf CP basierend können Statistiken erstellt werden, z. B. die prozentuale Verteilung der Längen von dno_i , bei denen $CN = 1$ ist: „5% ($L1$), 52% ($L2$), 13% ($L3$), 12% ($L4$), 7% ($L5$), 7% ($L6$), 2% ($L7$), 1% ($L8$), 1% ($L9$).“³⁴

$CN (0 \leq CN \leq 1)$ ist ein auf L_i (Länge von dno_i ohne Deweypunkt „.“) normiertes Korrelationsmaß:

$$CN := C / L_i$$

mit ³⁵

$$C := \sum_{s=1}^{L_i} CP_s$$

und CP_s (CP an der Stelle s)

$$CP_s(v_s) := \begin{cases} 1, & \text{falls } v_s = 1 \\ 0, & \text{sonst} \end{cases}$$

³³ Für leichtere Lesbarkeit steht ein Leerzeichen zwischen den Stellen $s = 7$ und 8 ; 10 und 11 ; 13 und 14 .

³⁴ $L1 := dno_i$ hat Länge 1; $L2 := dno_i$ hat Länge 2; ... vgl. [Reiner 2008c], S. 127

³⁵ In *vc_dcl* (April 2008) werden beim Stellenvergleich max. 16 Stellen (= s_{\max}) (mit „Deweypunkt“) betrachtet.



CN' (z. Zt. nicht implementiert) ist ein auf die kürzere Länge der beiden Längen (L_i, L_a) normiertes Korrelationsmaß :

$$CN'(C) = C / \min(L_i, L_a)$$

mit L_i (Länge von dno_i ohne Deweypunkt „.“) und L_a (Länge von dno_a ohne Deweypunkt „.“).

Die Maße CN und CN' messen den Grad der Übereinstimmung zwischen dno_i und dno_a .

Beispiel 5:

Seien $dno_i = 150$ (Psychologie) und
 $dno_a = 158.1$ (Persönliche Weiterentwicklung and Analyse).

Dann ist $CP = 110. xxx xxx xxx xxx$ und $CN = (1+1+0)/3 = 0.66666$.

Beispiel 6:

Seien $dno_i = 158.1$ (Persönliche Weiterentwicklung and Analyse) und
 $dno_a = 158$ (Angewandte Psychologie).

Dann ist $CP = 111.0xx xxx xxx xxx$ und $CN = (1+1+1+0)/4 = 0.75$.

Beispiel 7:

Seien $dno_i = 591.513$ (Intelligenz) [Oberklasse: 590 (Tiere)] und
 $dno_a = 156.39$ (Intelligenz bei Tieren--vergleichende Psychologie, ...)

Dann ist $CP = 000.000 xxx xxx xxx$ und $CN = (0+0+0+0+0+0)/6 = 0$.

Im 7. Beispiel ist deutlich zu sehen, dass CN ein rein auf Ziffernvergleich basierendes Korrelationsmaß ist. Bei unterschiedlichen Ziffern werden thematische Nähe wie z. B. im 7. Beispiel nicht berücksichtigt. Auch im 5. Beispiel wäre es möglich, dass die automatisch ermittelte DDC-Notation besser als die intellektuell vergebene (da sie spezifischer) ist. Solche Fälle (wie im 5. und 7. Beispiel) werden durch CP und CN also ggf. automatisch „unter“bewertet (fehlbewertet). Ob dies der Fall ist, muß anhand des realen Objektes (z. B. Buch oder Zeitschriftenartikel) intellektuell überprüft werden.

2.2.2 OCLC-Projekt Scorpion/DDC

Das OCLC-Projekt Scorpion [Shafer/Subramanian/Fausey 1999]³⁶ war ein Forschungsprojekt (1996-

³⁶ [Shafer/Subramanian/Fausey 1999] Keith Shafer; Srividhya Subramanian; Jon Fausey: Measures for Evaluating Automatic Subject Assignment of Electronic Resources. OCLC Online Computer Library Center, Inc. 6565 Frantz Road, Dublin,



2000³⁷), das sich zum Ziel gesetzt hat, elektronische Web-Dokumente automatisch nach DDC zu klassifizieren. Es wurden folgende elf Bewertungsmaße eingeführt, die zwei DDC-Notationen miteinander vergleichen: “**Hundreds**”, “**Tens**”, “**Ones**”, “**More General**”, “**More Specific**”, “**Correlated**”, “**Synonym**”, “**Best Possible**”, “**Exact**”, “**Relevant**”, und “**Close**”. Die Maße (hier durch fettgedruckte Buchstaben abgekürzt und mit **S** für Scorpion von **SI-S10** benannt) werden mit Colibri-Terminologie definiert; die Beispiele (falls vorhanden) werden aus [Shafer/Subramanian/Fausey 1999], S. 4 übernommen. Seien *dno1* und *dno2* zwei DDC-Notationen, die miteinander verglichen werden. Dann sind **SI - S10** (zehn Maße, da hier **G** und **S** zusammengefasst wurde):

S1: (H) *dno1* und *dno2* stimmen an 1. Stelle überein; z. B. 3 in 351 und 307.2.

S2: (T) *dno1* und *dno2* stimmen an 1. und 2. Stelle überein; z. B. 35 in 351 und 352.5.

S3: (O) *dno1* und *dno2* stimmen an 1., 2. und 3. Stelle überein; z. B. 512 in 512.57 und 512.2.

S4: (G) und **(Sp)** *dno1* (= **Sp**) ist ein echter Präfix von *dno2* (= **G**); z. B. 307.77 (= **Sp**) und 307.7 (= **G**); 307.77 (= **G**) und 307.776 (= **Sp**).

S5: (Co) Fachliche/thematische Nähe zwischen *dno1* und *dno2* (“Dewey concept overlap”); z. B. 401.41 (Discourse analysis) und 808.0014 (Discourse analysis--rhetoric). (Vgl. Beispiel 7.)

S6: (Sy) *dno1* (= **Sy**) zu *dno2*, falls *dno1* eine optionale DDC-Notation zu *dno2* ist; z. B. 492.7 und 4A0.

S7: (B) Sei *dno1* eine *dno_mol*, die nicht in den Haupttafeln enthalten ist. Dann ist *dno2* (= **B**) die erste DDC-Notation, die in den Haupttafeln enthalten ist, wenn *dno1* von rechts nach links sukzessive gekürzt wird; z. B. *dno1* = 333.7916072 und *dno2* (= **B**) = 333.7916.

S8: (E) *dno1* = *dno2*; z. B. 351 = 351.

S9: CI ∈ {**G/Sp, Co, Sy, B, E**}

S10: R ∈ {**H, T, O, G/Sp, Co, Sy, B, E, CI**}

Ohio, USA. Last modified: Mon Jul 26 10:14:16 1999.

[<http://orc.rsch.oclc.org:6109/measures.html>] Bemerkung: nicht mehr online verfügbar.

³⁷ [Oberhauser 2004] Otto Oberhauser: Automatisches Klassifizieren - Verfahren zur Erschließung elektronischer Dokumente. Master's Thesis. Zusatzstudiengang Bibliotheks- und Informationswissenschaft, Fakultät für Informations- und Kommunikationswissenschaften Fachhochschule Köln. Wien, 2004, S. 65



2.2.3 Projekt Pfeffer/RVK

Magnus Pfeffer³⁸ beschäftigt sich mit der automatischen Vergabe von RVK-Notationen und bewertet seine Ergebnisse nach dem Grad der Übereinstimmung mit dem „Goldstandard“³⁹ und den automatisch ermittelten RVK-Notationen. „Sollten im Standard mehrere Notationen hinterlegt sein, werden diese als gleichberechtigte Lösungen angesehen. Liefert das Verfahren mindestens eine der Notationen, wird dies als Übereinstimmung der Lösung mit dem Goldstandard gewertet. Ansonsten wird für jede Kombination von Notationen aus dem Standard und ermittelten Notationen der kleinste gemeinsame Vaterknoten in der Baumdarstellung der RVK ermittelt. Die Distanz zwischen den Knoten, die den Notationen entsprechen, und dem gemeinsamen Vater ist ein Indiz für den Grad der Übereinstimmung. Je kleiner die Distanz, desto näher liegt das Ergebnis am Goldstandard.“ [Pfeffer 2007], S. 19-20. Nach [Pfeffer 2007], S. 20 und [Pfeffer 2008]⁴⁰, S. 10 gibt es ein (hier so genanntes) **P**-Maß (mit vier Ausprägungen **P1** - **P4**) :

P1 : Ergebnis perfekt,	wenn Distanz 0
P2 : Ergebnis gut,	wenn Distanz 1-3
P3 : Ergebnis mäßig,	wenn Distanz > 3, aber noch gleiches Fach
P4 : Ergebnis schlecht,	wenn anderes Fach.

2.2.4 Vergleichende Betrachtung mit Bewertungsmaßen aus den Projekten 2.2.1-2.2.3

Als Diskussionsgrundlage und Entscheidungshilfe für eine Bewertung der Klassifizierungskomponente *vc_dcl* werden die oben eingeführten Bewertungsmaße anhand von Beispielen angewendet. Die in Abb. 8 verwendeten DDC-Notationen und ihre Klassenbenennungen sind: 100 (Philosophie, Parapsychologie und Okkultismus, Psychologie), 110 (Metaphysik), 111 (Ontologie), 111.8 (Klassische Wesenseigenschaften), 111.85 (Schönheit), 111.850952 (Ästhetik--Japan), 115 (Zeit), 170 (Ethik (Moralphilosophie)), 500 (Naturwissenschaften und Mathematik), 520 (Astronomie und zugeordnete Wissenschaften), 529 (Chronologie), 529.3 (Kalender), 529.326 (Jüdischer Kalender), 570 (Biowissenschaften Biologie), 571 (Physiologie und verwandte Themen), 571.58 (Einzelne Gewebe bei Pflanzen), 571.5929 (Einzelne Pilze und Algen), 571.68 (Zellwände) und 572.6 (Proteine).

³⁸ [Pfeffer 2007] Magnus Pfeffer: Automatische Vergabe von RVK-Notationen anhand von bibliografischen Daten mittels fallbasiertem Schließen. Masterarbeit im Rahmen des postgradualen Fernstudiums Master of Arts. Humboldt-Universität zu Berlin. Philosophische Fakultät I. Institut für Bibliotheks- und Informationswissenschaft. Mannheim, Juni 2007.

[<http://www.bib.uni-mannheim.de:8080/Classification/wp-content/uploads/2007/10/main.pdf>].

³⁹ Goldstandard: die „in den Titeldaten enthaltenen Notationen, die durch Experten vergeben wurden“ [Pfeffer 2007], S. 19

⁴⁰ [Pfeffer 2008] Automatische Vergabe von RVK-Notationen (anhand von bibliografischen Daten) mittels fallbasiertem Schließens. Vortrag auf dem 97. Deutschen Bibliothekartag, 3.-6. Juni 2008, Mannheim.

[http://www.bibliothekartag2008.de/de/Donnerstag_050608.htm]



Aus dem 8.-23. Beispiel (Abb. 8) können einige Schlüsse gezogen werden. Auch wenn sie nicht verallgemeinbar sind, so kann ein erster Eindruck gewonnen werden, was die Maße (nicht) messen. So ist z. B. $CN = 0$ in $S1-S10$ nicht direkt ausdrückbar (höchstens als $\neg S10$). Von CP ist nur der relevante Teil (die restlichen Zeichen der Zeichenkette bis $S_{max} = 16$ sind „x“) dargestellt. L_i bzw. L_a sind die Längen von dno_i bzw. dno_a (ohne Deweypunkt „.“).

8.-9. Beispiel: Bei $CN = 1$ ($CN' = 1$) stimmen mindestens drei Stellen von dno_i und dno_a überein ($S1-S3$), inhaltliche Nähe ($S9$) und inhaltliche Verwandtschaft ($S10$) sind erfüllt. Das Ergebnis wird im 8. Beispiel ($S8$ erfüllt) als perfekt ($P1$) und im 9. Beispiel ($S4=G$) als gut ($P2$) bezeichnet.

10.-17. Beispiel: Im Bereich $0.75 \leq CN \leq 0.33$ ($0.6 \leq CN' \leq 1$) stimmen mindestens die beiden ersten beiden Stellen von dno_i und dno_a überein ($S1$ und $S2$). Falls auch die dritte Stelle übereinstimmt ($S3$), dann stehen (bis auf 13. Beispiel) dno_i und dno_a in einer „spezifischer- bzw. allgemeiner-als-Beziehung“ ($S4=Sp$ und damit $S9$, inhaltliche Nähe). In allen Fällen sind dno_i und dno_a inhaltlich verwandt ($S10$) und nach $P2$ wird das Ergebnis als gut bezeichnet (mit Ausnahme des 15. und 17. Beispiels: bei $CN = 0.43$ gilt $P3$ und bei $CN = 0.33$ gilt $P4$).

		Colibri/DDC							Scorpion/DDC					Pfeffer/ RVK		
Bsp.	dno_i	dno_a	L_i	L_a	CP	C	CN	CN'	$S1$	$S2$	$S3$	$S4$	$S8$	$S9$	$S10$	P
8.	529.326	529.326	6	6	111.111	6	1	1	x	x	x		x	x	x	P1
9.	529	529.3	3	4	111	3	1	1	x	x	x	G		x	x	P2
10.	529.3	529	4	3	111.0	3	0.75	1	x	x	x	Sp		x	x	P2
11.	111	115	3	3	110	2	0.66	0.66	x	x					x	P2
12.	520	529	3	3	110	2	0.66	0.66	x	x					x	P2
13.	571.68	571.58	5	5	111.00	3	0.60	0.60	x	x	x				x	P2
14.	111.8	110	4	3	110.0	2	0.50	0.66	x	x					x	P2
15.	571.5929	571	7	3	111.0000	3	0.43	1	x	x	x	Sp		x	x	P3
16.	111.85	110	5	3	110.00	2	0.40	0.66	x	x					x	P2
17.	111.850952	111	9	3	111.000000	3	0.33	1	x	x	x	Sp		x	x	P4
18.	572.6	500	4	3	100.0	1	0.25	0.33	x						x	P2
19.	111.85	100	5	3	100.00	1	0.20	0.33	x						x	P3
20.	529.326	500	6	3	100.000	1	0.16	0.33	x						x	P3
21.	571.5929	500	7	3	100.0000	1	0.14	0.33	x						x	P3
22.	100	500	3	3	000	0	0	0								P4
23.	170	570	3	3	000	0	0	0								P4

Abb. 8: Vergleichende Betrachtung mit unterschiedlichen Bewertungsmaßen aus den Projekten Colibri/DDC, Scorpion/DDC und Pfeffer/RVK

18. Beispiel: Bei $CN = 0.25$ ($CN' = 0.33$) stimmt nur die erste Stelle von dno_i und dno_a überein ($S1$), dno_i und dno_a sind lediglich inhaltlich verwandt ($S10$) und nach $P2$ ist das Ergebnis (immer noch) gut.



19.-21. Beispiel: Im Bereich $0.2 \leq CN \leq 0.14$ ($CN' = 0.33$) stimmt nur die erste Stelle von *dno_j* und *dno_a* überein (S1), nach P3 werden die Ergebnisse als mäßig bezeichnet.

22.-23. Beispiel: Bei $CN = 0$ ($CN' = 0$) stimmt kein Zeichen überein, es besteht nicht einmal inhaltliche Verwandtschaft. Nach P4 (kein gemeinsamer Vaterknoten) ist das Ergebnis schlecht.

Im nächsten Kapitel wird im DNB-Bericht von unterschiedlichen Übereinstimmungen gesprochen. Bei einer „exakten Übereinstimmung“ ist $CN = 1$ (S8 bzw. P1), bei „keiner Übereinstimmung“ ist $CN = 0$ (P4). Es ist zu klären, was unter einer „guten“, „mittleren“ und „ausreichenden Übereinstimmung“, was unter „Hauptsachgruppentreffer“ bzw. „Nebensachgruppentreffer“ zu verstehen ist (vgl. DNBI-3 und DNB5-Iff in Kapitel 3), ggf. Einführung eines neuen Maßes, z. B. *CS* (correlation subject group).

3. Kommentierung des DNB-Berichtes „Colibri_Auswertung_DDC_Endbericht_Sommer_2008“

Auf [Jahns / Mödden 2008]⁴¹ wird im Folgenden Bezug genommen und es wird ergänzende Information zur Klärung von Sachverhalten gegeben bzw. bei unklaren Sachverhalten werden Fragen formuliert. Für die vereinfachte Bezugnahme werden Aussagen aus [Jahns / Mödden 2008] mit DNBS-i (s: Seitenzahl; i = 1, 2, 3, ...), Fragen mit Fj (j = 1, 2, 3, ...) und Kommentare aus den Excel-Dateien mit DNB-Kxxx (xxx = 001, 002, ..., 104) bezeichnet.

DNBI-1> Ausgewertet wurden 1012 Titel mit jeweils ca. 100 Titeln aus den zehn Hauptklassen der DDC.

Sample1 enthält – wie gewünscht – 1000 (nicht 1012) Klassifizierungsergebnisse aus den zehn DDC-Hauptklassen (vgl. Kapitel 2.1.4)

DNBI-2> Bei 34% der Gesamtdatenmenge war keine Auswertung möglich, d.h. nur 665 Titel konnten überhaupt ausgewertet werden.

Durch diese und weitere Aussagen in [Jahns / Mödden 2008] stellt sich heraus, dass für die Auswertung nur die Ergebnisse der automatischen Klassifizierung berücksichtigt werden sollen, die mit einer intellektuellen DDC-Notation verglichen werden können. Klassifizierungsergebnisse mit einer DDC-ähnlichen-Notation bleiben unberücksichtigt.

⁴¹ [Jahns/Mödden 2008] Yvonne Jahns; Elisabeth Mödden: Automatisches Klassifizieren_DDC_Colibri. DNB-Bericht, L2/F2 AG Netzpublikation. Deutsche Nationalbibliothek, 3. Juli 2008.

[https://wiki.d-nb.de/download/attachments/8945797/Colibri_Auswertung_DDC_Endbericht_Sommer_2008.doc?version=1]



DNBI-3> Überprüfung der Qualität der automatischen Vergabe einer DDC-Notation zu einem Titeldatensatz durch die DNB.

5 %	Exakte Übereinstimmung
5 %	Gute Übereinstimmung
5 %	Mittlere Übereinstimmung
22 %	Nur Hauptsachgruppentreffer
63 %	Keine Übereinstimmung

1. Um welche Klassifizierungsergebnismenge handelt es sich in dieser Tabelle? (**F1**)
2. Was bedeuten „Gute Übereinstimmung“ und „Mittlere Übereinstimmung“, was „Nur Hauptsachgruppentreffer“? (**F2**)
3. Die oben stehende Tabelle wird so interpretiert, dass ein Vergleich zwischen intellektuell verbogener DDC-Notation bzw. DDC-ähnlicher-Notation (“Hauptsachgruppentreffer”) und automatisch ermittelter (ermittelten) DDC-Notation (DDC-Notationen) wiedergegeben wird.

DNBI-4> Ein weiterer Test mit einer besseren Datenbasis ist erforderlich. Das bedeutet, ZDB-Datensätze und Datensätze ohne Notationen müssen vorher herausgefiltert werden. Falls das nicht möglich sein sollte, müssen mind. 2000 Datensätze ausgewertet werden.

1. Solange keine zufriedenstellenden Ergebnisse erzeugt werden, müssen auf jeden Fall weitere Tests durchgeführt werden.
2. Was ist mit einer „besseren Datenbasis“ gemeint? (**F3**)
3. Selbstverständlich können Stichproben wie *Sample2*, *Sample3*, ... hergestellt werden, in denen jedes Klassifizierungsergebnis eine intellektuelle DDC-Notation enthält. Ggf. muß der Umfang des Ausgangsmaterials (Anzahl der Monats-/Wochenlieferungen) erhöht werden.

DNBI-5> Eine Modifizierung des Colibri-Systems vor einem erneuten Testlauf wäre wünschenswert. Es sollte bspw. die DDC-Basis von Colibri auf einen aktuellen Stand gebracht werden (Melvil-Daten). Die Gewichtung der von Colibri herausgefilterten einzelnen Deskriptoren könnte analysiert und für die DNB-Daten neu überdacht werden (auch in Abhängigkeit der Unterschiede zwischen den Reihen A, B und H). Darüber und ob eine Einbeziehung von weiteren Metadaten am Titeldatensatz (Abstracts, VLB-Schlagwörter) möglich (und sinnvoll) wäre, sollte mit Frau Dr. Reiner diskutiert werden.



1. Die Klassifizierungskomponente *vc_dcl* befindet sich im Anfangsstadium, die Erweiterung und Verbesserung von *vc_dcl* steht ausser Frage.

2. Ad: „DDC-Basis von Colibri auf einen aktuellen Stand gebracht werden“. Die *DDC-Datenbasis* wurde im Januar 2008 für den 1. DNB-Klassifizierungstest neu erstellt. Sie enthält 568.257 (April 2008) bzw. 572.382 (Dezember 2008) unterschiedliche DDC-Notationen (zum Vergleich: die „alte“ [Juli 2004] *DDC-Datenbasis* enthält 466.134 unterschiedliche DDC-Notationen). Die *DDC-Datenbasis* enthält des Weiteren die (englischen) *Fakten* der *DDC-Wissensbasis* (vgl. Kapitel 2.1.1, *vc_DB_PLUS*)

3. Ad: „Melvil-Daten“. Um welche Daten handelt es sich (**F4**)? Sicherlich nützlich, da u. a. deutsche Deskriptorwerte enthalten sind.

4. Ad: „Die Gewichtung der von Colibri herausgefilterten einzelnen Deskriptoren könnte analysiert und für die DNB-Daten neu überdacht werden“. Was ist mit „Gewichtung der ... einzelnen Deskriptoren“ gemeint? Ist damit die heuristische Funktion *cutoff_val_dyn* gemeint (**F5**) (vgl. Kapitel 2.1.6)? Wie unterscheiden sich die DNB-Titeldatensätze der Reihen A, B und H voneinander, wie von z. B. OLC-Titeldatensätzen? (**F6**)

5. Ad: „Darüber und ob eine Einbeziehung von weiteren Metadaten am Titeldatensatz (Abstracts, VLB-Schlagwörter) möglich (und sinnvoll) wäre, sollte mit Frau Dr. Reiner diskutiert werden.“ Schlagwörter sind sehr nützlich und sinnvoll (insbesondere, wenn es Diskriminatoren, d. h. Deskriptorwerte mit kleinen Klassenhäufigkeitswerten sind) und können die Klassifizierungsergebnisse erheblich verbessern⁴². Die Einbeziehung von Abstracts (natürlichsprachliche Texte) ist nicht notwendigerweise eine Garantie für verbesserte automatische Klassifizierungsergebnisse. Die Einbeziehung von Abstracts, die u. a. auch Effizienzprobleme verursachen kann, ist daher vorerst (vor anderen Mitteln, die zur Verbesserung beitragen könnten) zurückzustellen. Allerdings wäre die Einbeziehung von Abstracts in den Fällen nützlich, wenn ansonsten zuwenig inhaltsreiche Information zur automatischen Klassifizierung vorliegt, wie es z. B. bei dem DNB-Titeldatensatz *dnb_DNB984784829* der Fall ist. Hier steht lediglich „Der Apfel“ (MAB2 Feld 331) zur Klassifizierung zur Verfügung. „Der“ wird von der Klassifizierungskomponente *vc_dcl* eliminiert und es verbleibt „Apfel“. Es ist offensichtlich, dass diese Klassifizierungsinformation für eine automatische DDC-Klassifizierung nicht ausreichen kann, vor allem bei einem so allgemeinen Deskriptorwert wie „Apfel“. *vc_dcl* ermittelt in *calc2* (*sections*) den DDC-Notationskandidaten „823“ („Englische Erzählprosa“).

DNBI-6> Nach diesem Test scheint eine automatische Sachgruppenvergabe nach Verbesserung des Programms eventuell möglich.

⁴² Vgl. [Reiner 2008b], S. 34 (CN=0 . 428571) und S. 35 (CN=1) mit den zusätzlichen Deskriptorwerten: „bioinformatik“ und „zellularer automat“.



Ja, ev. kann eine zufriedenstellende automatische DDC-Klassifizierung in die tausend Klassen der dritten Ebene erreicht werden, vielleicht in einigen Fällen darüber hinaus.

DNB1-7> Parallel mit dem nächsten Test sollte eine Zusammenarbeit mit der Universitätsbibliothek Mannheim erwogen werden. Dort wird ein Projekt zur automatischen Klassifizierung mit der RVK (Herr Pfeffer) durchgeführt. Hierzu wird DNB mit Herrn Pfeffer in Kontakt treten und prüfen, ob eine Bereitschaft für eine Kooperation besteht und wie diese möglich sein kann.

Eine Zusammenarbeit mit Herrn Magnus Pfeffer ist erstrebenswert (vgl. auch Kapitel 2.2.3 und 2.2.4).

DNB2-1 Der GBV hat ca. 3 Millionen Titeldaten mit DDC, vor allem aus der LoC

Vgl. Kapitel 2.1.1, Fußnote 5.

DNB2-2 Das Verfahren wurde bisher an der VZG für ca. 21.000 GVK-Titel mit DDC-Notationen und eine Stichprobe von ca. 600 Online Contents Artikeln⁴³ angewandt. Eine Auswertung erfolgte bisher nur rein statistisch hinsichtlich der Übereinstimmung von intellektuell und automatisch vergebenen Notationen, eine intellektuelle Überprüfung dieser Ergebnisse steht noch aus.

1. Um die Treffsicherheit (accuracy) der automatischen Klassifizierungskomponente *vc_dcl* zu testen, wurde im Februar 2007 eine DDC-Testklassifizierung mit ca. 21.000 intellektuell DDC-klassifizierten GVK-Titeldatensätzen der *DDC-Datenbasis* aus allen DDC-Klassen durchgeführt. Diese diente dazu, die Korrektheit und Treffsicherheit der Klassifizierungskomponente *vc_dcl* zu testen, d. h. in wieviel Prozent der Fälle die automatische Klassifizierung die intellektuelle Klassifizierung „trifft“. Das damalige Ergebnis war, dass in 0,3% aller 21.000 DDC-Klassifizierungen die intellektuelle und automatische DDC-Notation nicht übereinstimmen. Die nicht 100%-ige Übereinstimmung ist in der heuristischen Funktion *cutoff_val_dyn* begründet. *vc_dcl* hat im Februar 2007 demnach ca. 21.000 DDC-intellektuell klassifizierte GVK-Titeldatensätze mit einer 99,7%-igen Treffsicherheit automatisch klassifiziert. Eine im August/September 2008 durchgeführte Testklassifizierung aller *GVK-DDC-Titeldatensätze*⁴⁴ ergab eine 98,6%-ige Treffsicherheit (CN=1). Eine vollständige Testklassifizierung aller *GVK-DDC-Titeldatensätze* mit intellektuell vergebenen DDC-Notationen (nur) I. Priorität ist geplant.

2. Um die Qualität (quality)⁴⁵ der Klassifizierungskomponente *vc_dcl* zu testen, wurde eine

⁴³ Der GBV hat ca. 18 Millionen solcher OLC-Titel (i.d.R. ohne DDC-Notationen)

⁴⁴ 4.469.875 Klassifizierungsergebnisse, entstanden aus *GVK-DDC*, d. h. 4.319.357 Titeldatensätzen mit mehreren DDC-Notationen. Diese Testklassifizierung hat intellektuell vergebene DDC-Notationen jeder Priorität berücksichtigt.

⁴⁵ Die intellektuell vergebene DDC-Notation wird als Qualitätskriterium verwendet. Die Qualität von *vc_dcl* ergibt sich hiermit durch einen Vergleich zwischen intellektuell und automatisch vergeben[er/en] DDC-(ähnlichen) Notation[en].



repräsentative Stichprobe von 600 Online Contents (OLC)-Aufsatztiteln aus einer Grundgesamtheit von ca. 18 Millionen OLC-Titeldatensätzen erstellt, die mit *vc_dcl* automatisch klassifiziert wurde. Das Ergebnis wurde – wie in DNB2-2 erwähnt – bislang intellektuell noch nicht überprüft, sondern stattdessen wurde auf Wunsch der DNB die Qualitätsüberprüfung an den DNB-Titeldatensätzen *dnb_ABH* vorgenommen.

DNB2-3 Der GBV hat ca. 18 Millionen solcher OLC-Titel (i.d.R. ohne DDC-Notationen)

„In OLC sind über 28,4 Mio. Aufsatztitel aus über 21.010 Zeitschriften nachgewiesen.“ (November 2008) [<http://gso.gbv.de/DB=2.3/>]. Davon sind 49 Aufsatztitel mit mindestens einer DDC-Notation versehen (2008): 4 DDC-Notationen (mit 0 beginnend), 1 DDC-Notation (mit 1 beginnend), 16 DDC-Notationen (mit 2 beginnend) 10 DDC-Notationen (mit 3 beginnend), 1 DDC-Notation (mit 5 beginnend), 10 DDC-Notationen (mit 6 beginnend), 3 DDC-Notationen (mit 7 beginnend), 1 DDC-Notation (mit 8 beginnend) und 3 DDC-Notation (mit 9 beginnend).

DNB2-4 Das von Frau Dr. Reiner statistisch gewonnene Ergebnis des Durchlaufs klang zunächst viel versprechend⁴⁶ hielt aber der intellektuellen Überprüfung durch die DNB-Fachreferenten nicht stand.

1. Hier wird zunächst ein Teil des Originaltextes aus [Reiner 2008c], S. 126 und S. 127 zitiert:

“All title records of *in_dnb_ABH* have been automatically DDC-classified with *vc_dcl* ... – these classification results are now ready to be intellectually evaluated by the DNB. As almost all title records of the file *dnb_ABH* contain at least one DDC (-like)⁴⁷ number, it is also possible to measure the correlation between the intellectually (*dno_i*) and automatically determined *dno(s)_a* (cf. last line of Fig. 11) ... *CN* is a normalized measure between “0” and “1” considering the length of *dno_i*. Currently, for the *in_dnb_ABH* title records, 8% have *CN* value “X”, 33% *CN* value “0”, 37% *CN* value greater “0”, and 22% *CN* value “1”. ...”

2. Ad: „Das von Frau Dr. Reiner statistisch gewonnene Ergebnis des Durchlaufs klang zunächst viel versprechend“

a) Ad: „viel versprechend“.

Nach Meinung der Autorin ist das Klassifizierungsergebnis zu verbessern. Es ist zu klären, was ein zufriedenstellendes Klassifizierungsergebnis wäre.

b) Ad: „22 % völlige Übereinstimmung aller Notationen u. mind. bei der Grundnotation“ (Fußnote 46).

⁴⁶ Nach ihrem rein zahlenmäßigen Vergleich der Notationen erreichte sie 22 % völlige Übereinstimmung aller Notationen u. mind. bei der Grundnotation.

⁴⁷ DDC-like number: subject group based on the divisions of the DDC.



Die vorstehende Aussage wird in [Reiner 2008c], S. 127 so nicht getroffen, sondern es wird gesagt: „Currently, for the *in_dnb_ABH* title records, 8% have CN value „X“, 33% CN value „0“, 37% CN value greater „0“, and 22% CN value „1“. ...“. Damit wird gesagt, dass 22% (5.314)⁴⁸ aller (24.642) automatisch klassifizierten *in_dnb_ABH*-Titeldatensätze den CN-Wert „1“ haben. Entsprechend haben 8% aller automatisch klassifizierten *in_dnb_ABH*-Titeldatensätze den CN-Wert „X“⁴⁹ (2.021), 33% (8.156) den CN-Wert „0“ und 37% (9.151) einen CN-Wert größer „0“. Werden die Klassifizierungsergebnisse mit den „X“-Werten eliminiert, reduziert sich die Menge der automatisch klassifizierten *in_dnb_ABH*-Titeldatensätze auf 22.621 (24.642 - 2.021 = 22.621). Es ergibt sich dann folgende prozentuale Verteilung (vgl. auch Abb. 83):

Grundgesamtheit res in dnb ABH (April 2008)

23,49% (531400 / 22621) haben den CN-Wert „1“, 40,45% (915100 / 22621) haben einen CN-Wert größer „0“ und 36,06% (815600 / 22621) haben den CN-Wert „0“.

Stichprobe Sample I (April 2008)

23,4% (234) haben den CN-Wert „1“, 31,2% (312) haben einen CN-Wert größer „0“ und 45,4% (454) haben den CN-Wert „0“.

Bezüglich des Maßes CN liegt der Fehlerbereich von *Sample I* (1000 Klassifizierungsergebnisse, Stichprobenumfang) zwischen ca. 9,25% und 9,34%.

3. Ad: „hielt aber der intellektuellen Überprüfung durch die DNB-Fachreferenten nicht stand.“

Eine intellektuelle Überprüfung sollte dasselbe Ergebnis (wie die automatische Bewertung) erzielen. Das Missverständnis entsteht aufgrund unterschiedlicher Annahmen. Die Annahmen, die die Autorin getroffen hat, sind offensichtlich andere, als diejenigen der DNB-Fachreferenten. Zum einen basiert das Ergebnis CN-Wert = 1 (22%), wie eben unter 2. genannt, auf der Grundgesamtheit mit 24.642 DNB-Titeldatensätzen⁵⁰ und nicht der Stichprobe *Sample I*⁵¹ mit 1000 DNB-Titeldatensätzen. Zum anderen herrschen offensichtlich unterschiedliche Auffassungen darüber, welche Klassifizierungsergebnisse ausgewertet werden sollen. Die Autorin hatte angenommen, dass alle zur Verfügung gestellten DNB-Titeldatensätze (und nicht nur eine Teilmenge) automatisch klassifiziert werden sollen. Bei der automatischen Bewertung der Klassifizierungsergebnisse mit CP und CN (vgl. Kapitel 2.2.1) werden *dno_i* und *dno_a* miteinander verglichen⁵². Die Autorin ging davon aus, dass DDC-

⁴⁸ In Klammern absolute Anzahlen von klassifizierten Titeldatensätzen

⁴⁹ X: es gibt weder eine DDC- noch eine DDC-ähnliche Notation und/oder es gibt keinen Titel im DNB-Titeldatensatz, d. h. kein MAB2-Feld „331“, „335“, „341“ oder „451“.

⁵⁰ „All title records of *in_dnb_ABH* have been automatically DDC-classified with *vc_dcl*“ [Reiner 2008c], S. 126

⁵¹ vgl. DNBI-I

⁵² Hier könnte eingewendet werden, dass in mehr als der Hälfte der Fälle das Klassifizierungsergebnis nicht nur eine DDC-



ähnliche-Notationen zumindest in der 1. Stelle, ev. auch in der 2. und (falls vorhanden) ggf. auch in der 3. oder 4. Stelle (z. B. DNB983854742, DNB984908609, DNB976330210, DNB981906990) mit der entsprechenden DDC-Notation, die vergeben wurde oder vergeben werden würde, übereinstimmen.

Beispiel 24:

Der Titeldatensatz *dnb_DNB983297096* aus der Datei *res_in_dnb_A0745_A0746_A0747_A0748-I-20*⁵³ mit 20 Klassifizierungsergebnissen enthält im MAB2 Feld 700 die DDC-ähnliche-Notation 150 und im MAB2 Feld 705 die DDC-Notation 158.1. *vc_dcl* ermittelt als Klassifizierungskandidaten: {100[I], 150[I]}⁵⁴ und als Ergebnis der automatischen Bewertung

„correlation (DNB0983297096,158.1): 110.0xx xxx xxx xxx (0.5)“.

Der CN-Wert=0.5 ist aus Sicht der Autorin sinnvoll, da die ersten beiden Stellen der DDC-Notation (*dno_i*) und DDC-ähnlichen Notation (*dno_like_i*) übereinstimmen. Es ist jedoch kein Problem, die Fälle *dno_i-dno_a* und *dno_like_i-dno_a* voneinander getrennt zu bestimmen (vgl. Kapitel 4.3, Abb. 86ff).

DNB2-5 Anfang Juni 2008 wurden der DNB Daten ... zur Verfügung gestellt.

Sample1 wurde am 5.5.2008 unter [<https://wiki.d-nb.de/display/EDUG/VZG+Project+Colibri+DDC>] der DNB zur Verfügung gestellt.

DNB2-6 Ausgewertet wurden in L2/F2 1012 Titel mit jeweils ca. 100 Titeln aus den zehn Hauptklassen der DDC.

Sample1 enthält 1000 Klassifizierungsergebnisse wie in Kapitel 2.1.4 ausgeführt.

DNB3-I Die Titeldaten mit den automatisch vergebenen Colibri-Notationen wurden von Frau Reiner aus Göttingen auf dem DNB-Wiki zur Verfügung gestellt und sehen für jeden Titel folgendermaßen aus:

Notation *dno_a* ist, sondern eine Menge von automatisch ermittelten DDC-Notationen *dno_a*'s enthält und dass dann für CP bereits eine „I“ vergeben wird / der CN-Wert erhöht wird, auch wenn *dno_i* „nur“ in der Ergebnismenge vorkommt.

⁵³ *res_in_dnb_A0745_A0746_A0747_A0748-I-20* steht kurz für: *res* (result); *in_dnb* (input file of DNB); A0745, ..., A0748 (DNB-Dateiname für Wochen-/Monatslieferung: 07 für das Jahr; 45, ..., 48 für die jeweiligen Kalenderwochen); I: mit I beginnende DDC-Notationen; 20: Anzahl der in dieser Datei enthaltenen Klassifizierungsergebnisse.

⁵⁴ Ergebnis in Zeilen *calc2* (*sections*) und *calc2* (*divisions*).



```

...
identifizier (dno,schedno):          DNB0984491740 (779.28092,779)
DNB DDC notation (MAB2 field 700)  {770}
...
calculated1 ddc classes (subdiv):   |1| {779}
calculated1 ddc classes (sections): |1| {779}
calculated1 ddc classes (main):     |1| {700}
calculated2 ddc classes (subdiv):   {}
calculated2 ddc classes (sections): {779[1]}
calculated2 ddc classes (divisions):{770[1]}
calculated2 ddc classes (main):     {700[1]}
correlation (DNB0984491740,779): 111.xxx xxx xxx (1)

```

Abbildung 1: Titel mit automatischer Colibri-Notation

Der hier gezeigte Teil von „Abbildung 1“ aus [Jahns / Mödden 2008] stellt die derzeitige Ergebnisausgabe der automatischen Klassifizierung des DNB-Titeldatensatzes *dnb_DNB984491740* mit *calc1* und *calc2* (vgl. Kapitel 2.1.3, Fußnoten 24 und 25)⁵⁵ dar. Die automatische Klassifizierung mit der Klassifizierungskomponente *vc_dcl* liefert in *calc1* die Unterklasse 779⁵⁶ (*subdiv*) bzw. 779 als Klasse der 3. Ebene (*sections*) und die Hauptklasse 700 (*main*); *calc2* kommt hier zu dem gleichen Ergebnis: 779 (*sections*), 770 (*divisions*)⁵⁷ und 700 (*main*). In der letzten Zeile steht das Ergebnis der automatischen Bewertung (*correlation*) mit *CP* und *CN* (vgl. Kapitel 2.2.1).

DNB3-2 Es stellte sich bei der Auswertung der Daten heraus, dass bei 34% der Gesamtdatenmenge keine Auswertung möglich war, d.h. nur 665 Titel konnten überhaupt ausgewertet werden. Von den 347 Titeln die nicht ausgewertet werden konnten, hatten 70% keine Notation (zum Beispiel Sammelbände) und 26% waren ZDB-Daten.

1. Wieso ist nur bei 665 Titeln eine Auswertung möglich (vgl. Argumentation unter DNB2-4, 3.) **(F7)**

2. Was ist bei ZDB-Daten der Fall? Wieso ist da keine Auswertung möglich? Sollen diese niemals automatisch klassifiziert werden? **(F8)**

DNB4-I Fehlerursache, wenn keine Auswertung möglich war: 4% sonstige Ursache; 26% ZDB-Daten; 70% keine Notation

⁵⁵ Da in diesem Beispiel nur jeweils eine Klasse pro Ebene ermittelt wurde, wurde keine Klassenaggregation vorgenommen. Würde z. B. folgende Menge von DDC-Notationskandidaten {610, 617, 617.47, 617.47044} ermittelt, ergäben sich folgende Ergebnisse: *calculated2 ddc classes (subdiv): {617.47044[1], 617.47[1]}*, *calculated2 ddc classes (sections): {617[3]}*, *calculated2 ddc classes (divisions): {610[4]}* und *calculated2 ddc classes (main): {600[4]}*.

⁵⁶ „779“ ist keine Unterklasse (*subdivision*), sondern eine Klasse der 3. Ebene und sollte deshalb hier nicht aufgeführt werden (bei Gelegenheit in *vc_dcl* zu korrigieren).

⁵⁷ In *calc1* erfolgt z. Zt. keine Ausgabe der Klassen der 2. Ebene (*divisions*).



1. Was sind sonstige Ursachen? Woran können diese Fälle erkannt werden? (F9)

2. Sollen Titeldatensätze mit ZDB-Daten und „keine Notation“ niemals automatisch klassifiziert werden, vgl. DNB3-2, 2. (F10)?

DNB5-1 Bei 62% der Titel gab es keine Übereinstimmung, hier waren keine Stelle oder nur eine Stelle korrekt, d.h. die maschinell gefundene Notation ist inhaltlich fast immer abwegig bzw. irreführend. Bei 22% stimmten 2 oder 3 Stellen der Colibri-Notationen mit den DNB-Notationen überein, dies entspricht in etwa den Hauptsachgruppen. Eine mittlere Übereinstimmung der Notationen gab es bei 5% der Titel, eine gute Übereinstimmung ergab sich ebenfalls bei 5% der Titel. **Eine exakte Übereinstimmung, d.h. alle Stellen der Colibri-Notationen stimmten mit allen Stellen der DNB-Notationen überein, hatten nur 5% der Titel.**

1. Ad: „**Bei 62% der Titel gab es keine Übereinstimmung**, hier waren keine Stelle oder nur eine Stelle korrekt“. Aus Sicht der Autorin bedeutet „Keine Übereinstimmung“ := keine Stelle stimmt überein; eine Übereinstimmung := „eine Stelle stimmt überein“. Heisst „eine Stelle korrekt“, 1. Stelle stimmt überein (F11)?

2. Ad: „Bei 22% stimmten 2 oder 3 Stellen der Colibri-Notationen mit den DNB-Notationen überein, dies entspricht in etwa den Hauptsachgruppen.“ Bedeutet das: die ersten beiden oder ersten drei Stellen von *dno_j* und *dno_a's* bzw. *dno_like_j* und *dno_a's* stimmen überein, vgl. Argumentation unter DNB2-4, 3. (F12)?

3. Ad: „Eine mittlere Übereinstimmung der Notationen gab es bei 5% der Titel, eine gute Übereinstimmung ergab sich ebenfalls bei 5% der Titel.“ Was wird unter „mittlere Übereinstimmung“, was unter „gute Übereinstimmung“ verstanden (F13)?

4. Ad: „**Eine exakte Übereinstimmung, d.h. alle Stellen der Colibri-Notationen stimmten mit allen Stellen der DNB-Notationen überein, hatten nur 5% der Titel.**“ In Kapitel 4 wird dieses Ergebnis genauer analysiert.

DNB5-2 Vergleich der Colibri-Notationen mit den DNB-Notationen: 5% exakte Übereinstimmung; 5% gute Übereinstimmung; 5% ausreichende Übereinstimmung; 22% HSG getroffen; 63% keine Übereinstimmung.

Hier Abweichungen zum Text in DNB5-1, 62% bzw. 63% keine Übereinstimmung, 5% „ausreichende Übereinstimmung“. Was ist eine „ausreichende Übereinstimmung“? (F14)

DNB5-3 Anhaltspunkte für Ursachen des schlechten Ergebnisses lassen sich kaum erkennen. Man



kann jedoch davon ausgehen, dass das Wortmaterial aus Titelstichwörtern, Reihenangaben und z. T. vorhandenen Schlagwörtern zu gering für die Ableitung einer Notation ist bzw. die gegenüberstehende DDC-Datenbasis noch zu klein oder veraltet ist. Oftmals war sogar schwer nachvollziehbar, wie das Programm innerhalb der 5%-Marge zu den exakten Notationen kam.

Aufschluss kann hier wohl nur der direkte Austausch mit der Programmiererin bringen.

1. Ad: „Anhaltspunkte für Ursachen des schlechten Ergebnisses lassen sich kaum erkennen“. Hierzu müssen die Klassifizierungsergebnisse intellektuell gesichtet, Hypothesen und Klassen gebildet und danach das Klassifizierungsverfahren entsprechend verändert werden (vgl. Kapitel 4).

2. Ad: „Man kann jedoch davon ausgehen, dass das Wortmaterial aus Titelstichwörtern, Reihenangaben und z.T. vorhandenen Schlagwörtern zu gering für die Ableitung einer Notation ist.“ Hierzu vgl. z. B. [Reiner 2008b], S. 36 und 37.

3. Ad: „die gegenüberstehende DDC-Datenbasis noch zu klein oder veraltet ist.“ Die *DDC-Datenbasis* ist nicht veraltet, sondern wurde im Januar 2008 erstellt (vgl. DNBI-5, 2.). Die Größe der *DDC-Datenbasis* ist nicht das einzige Kriterium, sondern es kommt hauptsächlich darauf an, welche Deskriptorwerte in der *DDC-Datenbasis* (und in den zu klassifizierenden Titeldatensätzen) enthalten sind (vgl. Kapitel 2.1.5). Ev. spielt die Sprache der Deskriptorwerte (deutsche, englische oder anderssprachige Deskriptorwerte), die in der *DDC-Datenbasis* enthalten sind, eine Rolle.

4. Ad: „Oftmals war sogar schwer nachvollziehbar, wie das Programm innerhalb der 5%-Marge zu den exakten Notationen kam.“ Siehe Kapitel 4.

DNB5-4 Bei fast der gesamten Datenbasis von 1012 Titeln war eine intellektuell vergebene Hauptsachgruppe vorhanden, also auch bei den Titeln, bei denen ansonsten keine Auswertung möglich war, wegen einer fehlenden intellektuell vergebenen DDC-Notation. So ergab sich für die gesamte Datenbasis eine Hauptsachgruppentrefferquote von 39% durch Colibri.

1. Ad: „gesamten Datenbasis“? Ist damit *Sample I* gemeint (**FI5**)?

2. Ad: „fast“. 999 DNB-Titeldatensätze aus *Sample I* (1000 DNB-Titeldatensätze) enthalten eine oder mehrere DDC-ähnliche-Notation(en), lediglich der DNB-Titeldatensatz *dnb_DNB985396679* enthält kein MAB2-Feld 700 (DDC-ähnliche-Notation), wohl aber eine DDC-Notation (MAB2-Feld 705).

Wenn mehrere DDC-ähnliche-Notationen vergeben wurden, wie ist die Prioritätenreihenfolge (**FI6**)?



Beispiele: Beispiel 25 (*dnb_ZDB2299336-8*): {610,370,260,280}, Beispiel 26 (*dnb_ZDB2299456-7*): {920,100,305,z681}, Beispiel 27 (*dnb_ZDB716177-3*): {620,500,Y,TB 1030, AB 43080, 900}, Beispiel 28 (*dnb_ZDB1174458-3*): {500,18,31.la,27.la,700,800,900} und Beispiel 29 (*dnb_DNB981654398*): {840,B}. Was bedeuten in den Beispielen z681, Y bzw. B? (**F17**).

Zur Unterscheidung werden in der *vc_dcl*-Programmversion (November 2008) die Indikatoren des MAB2-Feldes 700 folgendermassen mit ausgegeben: für Beispiel 25: {610,370,z-260,z-280}, für Beispiel 26: {920,z-100,z-305,z-z681}, für Beispiel 27: {620,500,Y,g-TB 1030,g-AB 43080,z-900}, für Beispiel 28: {500,d-18,d-31.la,d-27.la,z-700,z-800,z-900} und für Beispiel 29 wie oben.

3. Ad: „So ergab sich für die gesamte Datenbasis eine Hauptsachgruppentrefferquote von 39% durch Colibri.“ Siehe Kapitel 4.

DNB5-5 Sollten sich hier in der Weiterentwicklung des Programms bessere Ergebnisse einstellen, wäre immerhin eine Perspektive für die Zuordnung von Webseiten zu Sachgruppen und somit zu ihrer bibliografischen Ordnung gegeben.

Ja, das wäre ein erstes, anzustrebendes Ziel.

DNB6-1 Abbildung 6: Hauptsachgruppentreffer der gesamten Datenmenge.

Ad: „gesamten Datenmenge“. Ist hiermit die Stichprobe *Sample1* (1000 Klassifizierungsergebnisse) oder die Grundgesamtheit *dnb_ABH* (24.642 Klassifizierungsergebnisse) gemeint? (**F18**)

DNB6-2 Alle Auswertungen werden im DNB-Wiki gespeichert und mit Frau Dr. Reiner diskutiert

Mit diesem Bericht wird u. a. die im Mai 2008 begonnene Diskussion weitergeführt. Weitere Fragen sind:

- Wieviele RVK-Klassen gibt es? (**F19**)
- Was ist ein HSG-, was ein NSG-Treffer? (**F20**)
- Wieviele Stellen hat eine HSG (Hauptsachgruppe?), eine NSG (Nebensachgruppe?) mindestens, höchstens? Wo stehen HSG, NSG? (**F21**)

DNB6-3 Ein erneuter Testlauf wird – nach Optimierung des Colibri-Tools an der VZG – für 2009 anvisiert. Für die Programmoptimierung sollten sämtliche DNB-klassifizierte Titel in die Datenbasis integriert werden (liegen in der VZG vor). Außerdem sollte Frau Dr. Reiner ein Abzug der aktuellen DDC (Melvil einschl. CrissCross-Anreicherung mit SWD-Schlagwörtern) zur Verfügung gestellt werden.



1. Ad: „sämtliche DNB-klassifizierten ... liegen in der VZG vor“. Wo sich diese Daten befinden, muss in der VZG geklärt werden.

2. Ad: „Abzug der aktuellen DDC (Melvil einschl. CrissCross-Anreicherung mit SWD-Schlagwörtern)“. Ja, das wäre sicherlich von Nutzen.

Im Folgenden wird auf die DNB-Kommentare (**DNB-K001** – **DNB-K104**) in den Excel-Dateien Bezug genommen.

DNB-Excel-Datei-Kommentare zur Hauptklasse 000

DNB-K001 {ZDB2167911-3} Vorschlag Colibri: SG 780, hier sicher wg. Salzburg

Ja, wegen „<331>-salzburger[82]“. Da weder der Deskriptorwert „museumshefte“ des Deskriptors „Hauptsachtitel in Vorlageform oder Mischform“ noch der Deskriptorwert „museum@salzburg“ des Deskriptors „Verleger@Ort“ in *vc_IDB* enthalten sind (erkennbar an in eckiger Klammer stehendem Häufigkeitswert „0“), verbleibt für die Klassenbestimmung als einziger – in 82 unterschiedlichen DDC-Klassen vorkommender – Deskriptorwert „salzburger“ des Deskriptors „Hauptsachtitel“. Von *vc_dcl* wird statt *dno_like_i=060* („Organisationen, Museumswissenschaft“) *dno_a=780* („Musik“) ermittelt, da diese Klasse in *calc2 (sections)* am häufigsten vorkommt. Um das Klassifizierungsergebnis zu verbessern, muß *QTIT* (vgl. Kapitel 2.1.5) verbessert werden, da ein einzelner Deskriptorwert im seltensten Fall für eine (eindeutige) DDC-Klasse stehen wird.

DNB-K002 {DNB0985281928} Colibri-Treffer hier 005.133 ("Einzelne Programmiersprachen"), DNB-Notation 005.1 ("Programmierung"). Die Richtigkeit der spezifischeren C-Notation könnte nur anhand des Buches beurteilt werden, der Pfad ist jedenfalls korrekt("005.1").

Ja, *dno_a=005.133* ist spezifischer als *dno_i=005.1*. *dno_a* müßte am Original auf Angemessenheit überprüft werden. Zum Ergebnis beigetragen haben die beiden Deskriptorwerte des „Hauptsachtitel-Deskriptors“: „<331>-feature[179]“ und „<331>-oriented[216]“. *CN=1*.

DNB-K003 {DNB0971645779} DNB-Altdaten ohne DDC; Colibri klassifiziert korrekt 005.1 ("Programmierung"); vielleicht aufgrund des mit der DDC-Nummer "005.1" hinterlegten SW "Evolutionärer Algorithmus"?

DNB971645779 mit dem Titel „Evolutionary multi-objective optimization – on the distribution of offspring in parameter and fitness space. Berichte aus der Informationstechnik“ enthält *dno_like_i=004*



(„Informatik“) und *dno_like_j*=510 („Mathematik“). WorldCat⁵⁸ liefert als „Related Subjects“: „Genetic algorithms. | Evolutionary programming (Computer science) | Mathematical optimization.“. Ist *dno_a*=005.1 („Programmierung“) somit vertretbar (**F22**)? *dno_a*=005.1 wird aufgrund der drei Deskriptorwerte, die alle gemeinsam in ihr vorkommen, ermittelt : „matched descriptor values: |3| {verteilter algorithmus, mehrkriterielle optimierung, multi-objective}“. Die ersten beiden Deskriptorwerte gehören zum Deskriptor „Sachschlagwort“, der dritte zum Deskriptor „Hauptsachtitel“. Die Häufigkeiten der Deskriptorwerte sind mit 14 bzw. 16 relativ klein: „<902s1>-verteilter algorithmus[14]“, „<902s>-mehrkriterielle optimierung[16]“ und „<331>-multi-objective[14]“. *CN*=0.67 (gerundet).

DNB-K004 {DNB0971870934} Colibri-Notation "004.6" vertretbar

dno_like_j=004 („Informatik“) wird für den Titeldatensatz DNB971870934 mit dem Titel: „Self configuring services for extensible networks - a routing integrated approach“ in *calc2* (*sections*) exakt ermittelt. Da alle drei Stellen von *dno_like_j* und *dno_a* übereinstimmen, ist *CN*=1. Zusätzlich werden in *calc2* (*subdiv*) die spezifischeren DDC-Notationskandidaten *dno_a*=004.6 („Rechnerkopplung und Nachrichtenübermittlung“) und *dno_a*=004.36 („Verteilte Verarbeitung“) mit jeweils der Häufigkeit 1 vorgeschlagen. Zwei Deskriptorwerte des Deskriptors „Sachschlagwort“ mit kleinen Häufigkeitswerten „<902s2>-betriebsmittelverwaltung[8]“ und „<902s1>-dienstverwaltung[3]“ und ein Deskriptorwert des Hauptsachtitel-Deskriptors: „<331>-configuring[41]“ tragen gemeinsam zu *dno_a*=004.6 bzw. *dno_a*=004.36 bei. Sind beide vertretbar (**F23**)? Im Katalog der Library of Congress (LoC)⁵⁹ sind für diesen Titeldatensatz folgende Schlagwörter vergeben: „Computer network architectures“, „Computer network protocols“, „Routers (Computer networks)“ und „Internet“. Die LoC klassifiziert diesen Titeldatensatz fragwürdigerweise mit der LCC (Library of Congress Classification)-Notation „TK5105.52“ (TK5101-6720: „Telecommunication - Including telegraphy, telephone, radio, radar, television“), was der DDC-Klasse „384.1-384.6“ („Telecommunication“) entspricht.

DNB-K005 {DNB0971894892} Reihe B Datensatz ohne Notation; HSG getroffen

Der Titeldatensatz DNB971894892 ist in Reihe H (Datei: H0710ti2.dis) mit *dno_like_j*=004 („Informatik“) enthalten. In *calc2* (*sections*) kommt *dno_a*=004 vor. Die vier Deskriptorwerte (hauptsächlich des Hauptsachtitel-Deskriptors): „<331>-geodaten[3]“, <451>-schriftenreihe[122], <331>-austausch[46] und <331>-xml[171] führen zu diesem Ergebnis. *CN*=1, da alle drei Stellen von *dno_like_j* und *dno_a* übereinstimmen.

DNB-K006 {DNB0985417900} Findet 004.21 in calculated| ddc classes; entscheidet sich allerdings

⁵⁸ [http://www.worldcat.org/search?qt=worldcat_org_all&q=tatsuya+okabe+Evolutionary+multi-objective+optimization]

⁵⁹ [<http://lccn.loc.gov/2008531164>]



für 620ziger Notation in *calculated2 ddc classes*, offenbar da hier mehr Treffer in der 620ziger SG gefunden?!

DNB985417900 wurde mit *dno_like_i=004*, *dno_i=004.210285117* und *schedno=004.21* klassifiziert. Ja, in diesem Fall ist die *calc1*- der *calc2*-Ergebnisausgabe (Verfahren s. Kapitel 2.1.3, Fußnoten 24 und 25) überlegen. Da sich die automatische Bewertung der *sample1*-Klassifizierungsergebnisse auf *calc2* bezieht und beim *calc2*-Verfahren die häufigsten DDC-Notationskandidaten bevorzugt und die anderen (aus Reduktionsgründen) fallengelassen werden, wird hier bei der Bewertung *dno_a=004.21* aus *calc1* (*subdiv*) nicht berücksichtigt, daher ist *CN=0*. Es muß untersucht werden, welche Ergebnisausgaben (*calc1*, *calc2* oder andere) im Allgemeinen die besten Ergebnisse liefern, was eine Frage von *QCALC* ist (vgl. Kapitel 2.1.5).

DNB-K007 {DNB0983974810, DNB0985814144, DNB0985435259} C-Notation nicht einsehbar ("More than 15 *calculated classes*") lt. *CN*-Wert exakter Treffer.

Beim Titeldatensatz DNB983974810 mit dem Titel „Programmieren in C“ stehen nur die Deskriptorwerte „<331>-programmieren[140], <412@410>-<033A>-springer@wien[116]“ mit hohen Häufigkeiten zur Verfügung. Deshalb werden sehr viele DDC-Notationskandidaten ermittelt. *dno_a=005.133* ist in der Menge der (hier) 100 DDC-Notationskandidaten enthalten, daher *CN=1*: „{001.3[1], 001.44[1], 001.4[1], 001.94[1], 003.3[1], 003.54[1], 004.015113[1], 004.0151[1], 004.03[1], 004.11[1], 004.125[1], 004.165[1], 004.16[1], 004.21[1], 004.24[1], 004.35[1], 004.36[1], 004.62[1], 004.678071[1], 004.678[1], 004.67[1], 004.68[1], 004.692[1], 004.6[1], 005.114[1], 005.115[1], 005.117[1], 005.11[1], **005.133**[1], 005.14[1], 005.16[1], 005.1[1], 005.265[1], 005.268[1], 005.26[1], 005.2762[1], 005.2768[1], 005.276[1], 005.2[1], 005.369[1], 005.3[1], 005.432[1], 005.446[1], 005.4476[1], 005.7136[1], 005.713765[1], 005.71376[1], 005.72[1], 005.74[1], 005.7565[1], 005.757[1], 005.7585[1], 005.758[1], 005.82[1], 005.8[1], 006.32[1], 006.33[1], 006.3[1], 006.454[1], 006.686[1], 006.68[1], 006.696[1], 006.6[1], 006.786[1], 006.7[1], 011.62[1], 016.33[1], 016.973[1], 021.2[1], 025.04[1], 025.0661[1], 025.06[1], 025.1714[1], 025.21[1], 025.3432[1], 025.47[1], 025.49[1], 025.524[1], 025.527[1], 025.52[1], 025.54[1], 025.84[1], 027.5[1], 027.626[1], 027.70973[1], 027.7[1], 028.5[1], 028.9[1], 069.52[1], 070.195[1], 070.409[1], 070.41092[1], 070.4332[1], 070.444[1], 070.486[1], 070.4[1], 070.5092[1], 070.593[1], 070.5[1], 070.92[1]}“. Um die Ausgabemenge aller Klassifizierungsergebnisse zu reduzieren, wurde die Ausgabe verkürzt, wenn mehr als 15 DDC-Notationskandidaten ermittelt wurden, da die automatische Klassifizierung in diesen Fällen von sehr beschränktem Nutzen ist. Wenn gewünscht, kann die Ausgabe selbstverständlich (wie oben) ausführlich erfolgen. Für intellektuelle Klassifizierer ist die Klassifizierung dieses Werkes sehr einfach: „C“ zusammen mit „programmieren“ führt direkt zur DDC-Klasse *dno_i=005.133* („Einzelne Programmiersprachen“). In der Klassifizierungskomponente *vc_dcl* wird jedoch das für den Menschen bedeutungstragende „C“ eliminiert, da es aus einem einzelnen Zeichen besteht. Sollte „C“ nicht automatisch eliminiert werden, müssten z. B. (aufwendige) Kontextregeln (Textverstehen) eingeführt werden.

Ähnliche Sachverhalte (Deskriptorwerte mit großen Häufigkeitswerten) sind in den Titeldatensätzen anzutreffen, in denen mehr als 15 DDC-Notationskandidaten ermittelt werden. So ist dies auch beim



Titeldatensatz DNB985814144 der Fall: von den drei Deskriptorwerten „fachstelle, oeffentliche, bibliothekswesen“ tragen zwei zur Ermittlung der DDC-Klassenkandidaten bei. $CN=1$.

Beim Titeldatensatz DNB985435259 sind die Deskriptorwerte „{<100>-timo#wolf[0], <331>-engineering[3538], <331>-model[3141], <331>-rationale-based[0], <331>-unified[492], <331>-software[1257]}“ mit Häufigkeitswerten „0“ (= Deskriptorwert nicht in *vc_IDB*) oder sehr hohen Häufigkeitswerten (= Deskriptorwerte in sehr vielen DDC-Klassen vertreten) verfügbar. $CN=0.5$.

Wie sollen Titeldatensätze mit sehr allgemeinen Deskriptorwerten in sehr spezifische DDC-Klassen wie z. B. in DNB985435259 mit *dno_i=005.12028* („Unterstützende Techniken und Verfahren; Geräte, Ausstattung, Materialien“) automatisch klassifiziert werden können (**F24**)?

DNB-K008 {DNB0971670889, DNB0971870713, DNB0971870934, DNB0971457123, DNB0972095330} DNB-Altdateien von 2004, keine Notation vorhanden.

Von den fünf Titeldatensätzen werden die *dno_like_i*'s in zwei Fällen ermittelt ($CN=1$).

DNB-K009 {DNB0985435259} lt. CN -Wert offenbar 005.1 getroffen

Wie unter DNB-K007 beschrieben, die Deskriptorwerte sind: „{<100>-timo#wolf[0], <331>-engineering[3538], <331>-model[3141], <331>-rationale-based[0], <331>-unified[492], <331>-software[1257]}“. Wegen *dno_i=005.12028* und (mind. einem) mit „005.1“ beginnenden *dno_a* aus 40 DDC-Notationskandidaten in *calc2 (subdiv)* ist $CP=111.100\ 00x\ xxx\ xxx$ und $CN=0.5$. Bei Verbesserung von *QTIT* (s. Kapitel 2.1.5) könnte der CN -Wert erhöht werden (Abb. 10): es sollten Deskriptorwerte, wie z. B. „system design“ und „software modeling“ des Deskriptors „Sachschlagwort“ aus dem nach „005.1“ folgenden Themenfeld hinzugefügt werden (Abb. 9):

```
005.12028 <dno_DNB0985435259_to_analyze; length: 9>
0----- Computer science, information & general works <hatzen>
00----- Computer science, knowledge & systems <hatzen>
005----- Computer programming, programs, data <hat>
005.1---- Programming <hat>
005.12--- Software systems analysis and design <hat>
005.12028 Auxiliary techniques and procedures; apparatus, equipment, materials
           <hat>
-05.----- Serial publications <T1--05>
---.1---- Programming <nfu:005.1>
---.12--- Software systems analysis and design <nfu:005.12>
---.12028 Auxiliary techniques and procedures; apparatus, equipment, materials
           <nfu:005.12028>
```

Abb. 9: *vc_daygram* der *dno_i=005.12028* mit der Analysekomponente *vc_day*

```
dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_dnb_DNB0985435259
number of ddc-classified title:      1
```



```

identifizier (dno,schedno):      DNB0985435259 (005.12028,005.12028)
DNB DDC notation (MAB2 field 700): {004}
DDC notation (MAB2 field 705):  {005.12028}
calculated cutoff value:       212
title:                          Rationale-based unified software engineering
model
considered descriptor values:   |8| {<100>-timo#wolf[0], <331>-
engineering[6869], <331>-model[4967], <331>-rationale-based[0], <331>-unified[592],
<331>-software[2267], <902s>-system design[212], <902s>-software modeling[1]}
matched descriptor values:      |2| {system design, software modeling}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |1| {005.12028}
calculated1 ddc classes (sections): |1| {005}
calculated1 ddc classes (main):     |1| {000}
calculated2 ddc classes (subdiv):   |1| {005.12028[1]}
calculated2 ddc classes (sections): {005[1]}
calculated2 ddc classes (divisions): {000[1]}
calculated2 ddc classes (main):     {000[1]}
correlation (dnb_H0711_DNB0985435259,005.12028): 111.111 11x xxx xxx (1)

```

Abb. 10: Klassifizierungsergebnis von *in_dnb_DNB0985435259* bei Verbesserung von *QTIT* mit „system design“ und „software modeling“

DNB_K010 {ZDB2280197-2} "CN"-Wert ergibt sich aus der von Colibri gefundenen SG 016

Der Titeldatensatz ZDB2280197-2 enthält *dno_like_i=050* („Zeitschriften, fortlaufende Sammelwerke“) und *dno_like_i=100* (Philosophie“). Aufgrund der drei Deskriptorwerte „<331>-monat[23]“, „<335>-styria[71]“ und „<335>-lifestyle[304]“ wird *dno_a=016* („Bibliografien und Kataloge von Werken über einzelne Themen oder in einzelnen Fachgebieten“) ermittelt. Daher ist *CP=10x.xxx xxx xxx xxx* und *CN=0.5*. Spricht etwas gegen diese Bewertung? Vgl. (F10).

DNB-K011 {DNB0985352671} CN-Wert ergibt sich aus gefundener SG 005 und ist daher falsch

Der Titeldatensatz DNB985352671 ist mit *dno_like_i=020* („Bibliotheks- und Informationswissenschaft“) und *dno_i=025.0655* („Geowissenschaften--Informationssysteme“) intellektuell klassifiziert worden. Die ermittelte DDC-Notation ist *dno_a=005* („Computerprogrammierung, Computerprogramme, Daten“). Daher ist *CP=100.000 0xx xxx xxx* und *CN=1/7=0.142857*. Wieso ist *CN* falsch (F25)? *CN=1* kann erreicht werden, wenn *QTIT* mit den Sachschlagwörtern: „earth sciences--information systems“ und „gis“ (geographic information system) angereichert wird (Abb. 11):

```

dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_dnb_DNB0985352671
number of ddc-classified title: 1
identifizier (dno,schedno):      DNB0985352671 (025.0655,025.0655)
DNB DDC notation (MAB2 field 700): {020}
DDC notation (MAB2 field 705):  {025.0655}
calculated cutoff value:       147
title:                          Verwaltung geowissenschaftlicher 3D-Daten in
mobilen Datenbanksystemen
considered descriptor values:   |8| {<100>-wolfgang#baer[1], <331>-mobilen[6],
<331>-datenbanksystemen[4], <331>-verwaltung[522], <331>-geowissenschaftlicher[3],

```



```

<331>-3d-daten[0], <902s1>-earth sciences--information systems[1], <902s2>-
gis[147]}
matched descriptor values:          |5| {mobilen, datenbanksystemen, gis, earth
sciences--information systems, geowissenschaftlicher}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |7| {004, 005, 005.74, 025.0655, 330, 550, 551}
calculated1 ddc classes (sections):  |3| {004, 005, 025}
calculated1 ddc classes (main):      |1| {000}
calculated2 ddc classes (subdiv):    |2| {005.74[1], 025.0655[1]}
calculated2 ddc classes (sections):  {005[2]}
calculated2 ddc classes (divisions): {000[3]}
calculated2 ddc classes (main):      {000[4]}
correlation (dnb_H0711_DNB0985352671,025.0655): 111.111 1xx xxx xxx (1)

```

Abb. 11: Klassifizierungsergebnis von *in_dnb_DNB0985352671* bei Verbesserung von QTIT mit „earth sciences--information systems“ und „gis“

DNB-K012 {DNB0984787046} CN-Wert "0,4" errechnet sich aus Colibri-"Treffer" 004

Bei diesem Titeldatensatz ist *dno_like_j=004* („Informatik“) und *dno_j=005.37* („Programme für einzelne Verarbeitungsarten“). Mit *dno_j=005.37* und *dno_a=004* ist $CP=110.00x\ xxx\ xxx\ xxx$ und $CN=2/5=0.4$. Dass *dno_j=005.37* nicht ermittelt wird, liegt daran, dass *dno_j* in *vc_IDB* nur bei den Deskriptorwerten „processing modes“ und „programs specific processing modes“ vorkommt, diese Deskriptorwerte jedoch nicht im Titeldatensatz vorkommen. Alle fünf Deskriptorwerte „{middleware, cluster <rechnernetz>, skalierbare, skalierbarkeit, autonomic computing}“ kommen jedoch in *dno_a=004* vor. Eine QTIT- und QIDB-Verbesserung könnte ev. bessere Ergebnisse liefern.

DNB-K013 {DNB0973315601} nur HSG vorhanden, daher CN-Wert "1"

Hier sind (nur) *dno_like_j=070* („Nachrichtenmedien, Journalismus, Verlagswesen“) und *dno_like_j=004* („Informatik“) gegeben; da *dno_a=070* („Dokumentarische Medien, publizistische Medien, Unterrichtsmedien; Journalismus; Verlagswesen“) ermittelt wird, ist $CN=1$. Fünf Deskriptorwerte tragen zum Klassifizierungsergebnis bei (ähnlicher Titel von Thorsten Quandt in *vc_IDB*): „<100>-thorsten#quandt[2]“, „<331>-journalisten[33]“, „<335>-journalistischen[13]“, „<412@410>-<033A>-vs verl. sozialwiss.@wiesbaden[128]“ und „<902s2>-journalismus[153]“.

DNB-K014 {DNB0985568674} Colibri vergleicht sein Ergebnis ("650") hier mit der zweiten von DNB vergebenen Notation 001.40684. Der richtige CN-Wert ist somit 0,4.

Ja. Der Titeldatensatz DNB985568674 enthält *dno_like_j=650*, *dno_like_j=000*, *dno_j=658.57* und *dno_j=001.40684*. In verbesserter *vc_dcl*-Programmversion (vgl. Fußnote 26) ist *dno_a=650* und somit $CP=110.00x\ xxx\ xxx\ xxx$ und $CN=5/2=0.4$. Die Deskriptorwerte des Titeldatensatzes DNB985568674 sind allerdings mit *dno_like_j=650* (in Pica+-Kategorie <045F> und <045G>) in *vc_IDB* enthalten, deshalb ist *dno_a=650* (und nicht spezifischer). Die Deskriptorwerte {f+e-outsourcing, forschungssystem, innovationschance, robert#gassner, <033A>-izt@berlin} tragen zum Klassifizierungsergebnis bei.



DNB-K015 {DNB0985295473} Vergleicht mit zweiter DNB-Notation

Ja. Der Titeldatensatz DNB985295473 ist folgendermaßen klassifiziert: *dno_like_i=004* („Informatik“), *dno_like_i=020* („Bibliotheks- und Informationswissenschaft“), *dno_i=006.7* („Multimediasysteme“) und *dno_j=025.04* („Informationsspeicherungs- und Information-Retrieval-Systeme“). Die fünf Deskriptorwerte des Deskriptors „Hauptsachtitel“: „{ranking, link, applications, emerging, analysis}“ führen nur zu DDC-Notationskandidaten, die mit “3” beginnen, deshalb ist *CN=0*. Wird *QTIT* mit den Schlagworten, die der Autor selbst in seiner Dissertation angegeben hat⁶⁰, angereichert, ermittelt *vc_dcl* u. a. die DDC-Notationskandidaten *dno_a=025.04* („Informationsspeicherungs- und Information-Retrieval-Systeme“) und *dno_a=006.74* („Auszeichnungssprachen“), wodurch *CN=1* wird (Abb. 12):

```
dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_dnb_DNB0985295473
number of ddc-classified title:      1
identifier (dno,schedno):            DNB0985295473 (006.7,006.7)
DNB DDC notation (MAB2 field 700):   {004,020}
DDC notation (MAB2 field 705):       {006.7,025.04}
calculated cutoff value:              181
title:                                Emerging applications of link analysis for
ranking
considered descriptor values:         |9| {<100b>-paul-alexandru#chirita[0], <331>-
link[511], <331>-analysis[6848], <331>-ranking[181], <331>-emerging[1426], <331>-
applications[4266], <902s>-data mining[107], <902s>-informationswiedergewinnung[3],
<902s>-internet[2193]}
matched descriptor values:            |3| {ranking, informationswiedergewinnung, data
mining}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):     |18| {003.56, 004, 005.1, 005.2762, 005.74,
006.3, 006.31, 006.312, 006.33, 006.74, 016, 025.04, 330, 519.5, 519.54, 650,
658.403, 658.84}
calculated1 ddc classes (sections):   |6| {003, 004, 005, 006, 016, 025}
calculated1 ddc classes (main):       |1| {000}
calculated2 ddc classes (subdiv):     |10| {003.56[1], 005.1[1], 005.2762[1],
005.74[1], 006.312[1], 006.31[1], 006.33[1], 006.3[1], 006.74[1], 025.04[1]}
calculated2 ddc classes (sections):   {006[5]}
calculated2 ddc classes (divisions):  {000[10]}
calculated2 ddc classes (main):       {000[12]}
correlation (dnb_H0711_DNB0985295473,006.7): 111.1xx xxx xxx (1)
```

Abb. 12: Klassifizierungsergebnis von *in_dnb_DNB0985295473* bei Verbesserung von *QTIT* mit „data mining“ und „informationswiedergewinnung“

DNB-K016 {DNB0985286083} Colibri vergleicht hier mit der zweiten DNB-Notation, die erste "629.8924019" wird nicht berücksichtigt. In "calculated1 ddc classes " findet sich neben 13 weiteren Notationen (größtenteils aus der 400) auch die SG 620.

Ja. Für den Titeldatensatz DNB985286083 mit dem Titel: „Robustes Verstehen gesprochener Sprache in einem multimodalen Roboter-Szenario“ werden von *vc_dcl* nur DDC-Notationskandidaten

⁶⁰ [http://deposit.d-nb.de/cgi-bin/dokserv?idn=985295473&dok_var=dok_ext=pdf&filename=985295473.pdf], S. 4
Schlagwörter: Informationswiedergewinnung, Data mining, Internet



ermittelt, die mit „4“ beginnen. Z. B. *dno_a=401.41* („Diskursanalyse“) oder *dno_a=410.285* („Datenverarbeitung Computereinsatz“), aber auch *dno_a=495.6* („Japanisch“). Intellektuell ist DNB985286083 mit *dno_like_i=620* („Ingenieurwissenschaften“), *dno_like_i=004* („Informatik“), *dno_j=629.8924019* und *dno_j=004.01902856454* klassifiziert (Abb. 13a und 13b). Deshalb ist *CN=0*.

```
629.8924019 <dno1_DNB985286083_to_analyze; length: 11>
6----- Technology <hatzen>
62----- Engineering <hatzen>
629----- Other branches of engineering <hat>
629.8----- Automatic control engineering <hat>
629.89----- Computer control <hat>
629.892---- Robots <hat>
---.---4--- Data processing      Computer science <nalr1:004>
---.---4019 Psychological principles <nalr1:004.019>
```

Abb. 13a: *vc_daygram* der *dno_j=629.8924019* mit der Analysekomponente *vc_day*

```
004.01902856454 <dno2_DNB985286083_to_analyze; length: 15>
0----- Computer science, information & general works <hatzen>
00----- Computer science, knowledge & systems <hatzen>
004----- Data processing      Computer science <hat>
004.019----- Psychological principles <hat>
---.01----- Philosophy and theory <T1--01>
---.019----- Psychological principles <T1--019>
---.---02----- Miscellany <T1--02>
---.---028----- Auxiliary techniques and procedures; apparatus, equipment,
materials <T1--028>
---.---0285----- Data processing      Computer applications <T1--0285>
---.---028564-- Pattern recognition--computer science <T1--028564>
---.-----6--- Special computer methods <nalr1:006>
---.-----64-- Computer pattern recognition <nalr1:006.4>
---.-----645- Acoustical pattern recognition <nalr1:006.45>
---.-----6454 Speech recognition <nalr1:006.454>
```

Abb. 13b: *vc_daygram* der *dno_j=004.01902856454* mit der Analysekomponente *vc_day*

```
dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_dnb_DNB0985286083
number of ddc-classified title:      1
identifier (dno,schedno):            DNB0985286083 (629.8924019,629.892)
DNB DDC notation (MAB2 field 700):  {620,004}
DDC notation (MAB2 field 705):      {629.8924019,004.01902856454}
calculated cutoff value:            207
title:                               Robustes Verstehen gesprochener Sprache in
einem multimodalen Roboter-Szenario
considered descriptor values:        |9| {<100>-sonja#huewel[0], <331>-
sprache[1251], <331>-multimodalen[3], <331>-roboter-szenario[0], <331>-robustes[1],
<331>-verstehen[508], <331>-gesprochener[10], <902s1>-robot[207], <902s2>-automatic
speech recognition[34]}
matched descriptor values:           |4| {robot, multimodalen, gesprochener,
automatic speech recognition}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |8| {001, 006.3, 006.37, 338.47, 410.285, 620,
621.399, 629.892}
calculated1 ddc classes (sections):  |5| {001, 006, 620, 621, 629}
calculated1 ddc classes (main):      |2| {000, 600}
calculated2 ddc classes (subdiv):    |4| {006.37[1], 006.3[1], 621.399[1],
629.892[1]}
```



```

calculated2 ddc classes (sections): {006[2]}
calculated2 ddc classes (divisions):{000[3], 620[3]}
calculated2 ddc classes (main):      {000[3], 600[3]}
correlation (dnb_H0709_DNB0985286083,629.8924019): 111.111 000 0xx xxx (0.6)

```

Abb. 14: Klassifizierungsergebnis von *in_dnb_DNB0985286083* bei Verbesserung von *QTIT* mit „robot“ und „automatic speech recognition“

Mit einschlägigen Sachschlagworten kann das Ergebnis auf $CN=0.6$ verbessert werden (Abb. 14). Ein größerer CN -Wert ist nicht erreichbar, da *vc_IDB* (Januar 2009) keine Deskriptorwerte zur DDC-Klasse 629.8924019 enthält, sondern erst ab DDC-Klasse „629.892“ („Roboter“) aufwärts. Mit Aufnahme von Titeldatensätzen, die mit *dno_i=629.8924019* klassifiziert wurden, kann diese Art von Lücke geschlossen werden, d. h. durch Verbesserung von *QIDB*.

DNB-K017 {DNB0972095330} DNB-Altdateien ohne Notation, Reihe A

Der Titeldatensatz DNB972095330 ist in Reihe H (Datei: H0711ti2.dis) enthalten und nur mit *dno_like_j=004* („Informatik“) klassifiziert. $CN=1$, da *dno_a=004* mithilfe der Deskriptorwerte „{automaten, jens#schwarzer, nervennetz}“ ermittelt wird.

DNB-Excel-Datei-Kommentare zur Hauptklasse 100

DNB-K018 {DNB0984531548} Hauptnotation wurde nicht erfasst

S. Fußnote 26. Der Titeldatensatz DNB984531548 mit dem Titel: „Comenius' Pansophie - die dreifache Offenbarung Gottes in Schrift, Natur und Vernunft“ enthält *dno_like_j=230* („Theologie, Christentum“), *dno_like_j=100* („Philosophie“), *dno_like_j=220* („Bibel“), *dno_i=230.01* („Philosophie und Theorie des Christentums, der christlichen Theologie“) und *dno_i=121* („Epistemologie (Erkenntnistheorie)“). In beiden Programmversionen *vc_dcl* (April und Oktober 2008) ist $CN=0$. Eine Verbesserung von *QELIM* (vgl. Kapitel 2.1.5) durch Eliminierung der Deskriptorwerte „unvollendete“ und „habilitationsschrift“ ergibt $CN=0.8$:

```

dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_dnb_DNB0984531548
number of ddc-classified title:      1
identifier (dno,schedno):            DNB0984531548 (230.01,230.01)
DNB DDC notation (MAB2 field 700):   {230,100,220}
DDC notation (MAB2 field 705):       {230.01,121}
calculated cutoff value:             153
title:                               Comenius' Pansophie
title (remainder):                  die dreifache Offenbarung Gottes in Schrift,
Natur und Vernunft ; unvollendete Habilitationsschrift
considered descriptor values:         |11| {<100>-daniel#neval[0], <331>-
comenius'[0], <331>-pansophie[3], <335>-gottes[570], <335>-schrift[631], <335>-
natur[2759], <335>-vernunft[629], <335>-dreifache[28], <335>-offenbarung[153],
<412@410>-<033A>-theol. verl.@zuerich[77], <540a>-3-290-17435-2[0]}
matched descriptor values:           |4| {offenbarung, dreifache, <033A>-theol.
verl.@zuerich, pansophie}
max. match value of matched descriptor values: |2|

```



```

calculated1 ddc classes (subdiv): |27| {193, 200, 200.1, 215, 221, 225, 225.6,
226.406, 226.5, 227.83, 228, 229.8, 230, 230.044, 230.044092, 230.2, 231.7, 232,
232.1, 242, 252, 261.7, 262, 284, 299.2, 831.6, 914}
calculated1 ddc classes (sections): |17| {200, 215, 221, 225, 226, 227, 228, 229,
230, 231, 232, 242, 252, 261, 262, 284, 299}
calculated1 ddc classes (main): |1| {200}
calculated2 ddc classes (subdiv): |13| {200.1[1], 225.6[1], 226.406[1], 226.5[1],
227.83[1], 229.8[1], 230.044092[1], 230.044[1], 230.2[1], 231.7[1], 232.1[1],
261.7[1], 299.2[1]}
calculated2 ddc classes (sections): {230[4]}
calculated2 ddc classes (divisions): {220[8]}
calculated2 ddc classes (main): {200[24]}
correlation (dnb_A0745_DNB0984531548,230.01): 111.10x xxxx xxxx xxxx (0.8)

```

Abb. 15: Klassifizierungsergebnis von *in_dnb_DNB0984531548* bei Verbesserung von *QELIM* durch Löschung von „unvollendete“ und „habilitationsschrift“

DNB-K019 {DNB0985287896} Primäre Notation nicht erfasst

S. Fußnote 26. Der Titeldatensatz DNB985287896 enthält den (italienischen) Titel und enthält keine Sachschlagwörter: „... A ... commentary on the "De predestinatione et prescientia, paradiso et inferno" by Giles of Rome on the Basis of MS Cambrai BM 487 455. Er ist mit *dno_like_i=230* („Theologie, Christentum“), *dno_like_j=100* („Philosophie“), *dno_i=230.2* („Römisch-Katholische Kirche--Lehren, . . .“) und *dno_j=189.4* („Scholastische Philosophie“) intellektuell klassifiziert. Von den zehn betrachteten Deskriptorwerten führen die vier „{giles, cambrai, inferno, paradiso}“ zu mit „8“ beginnenden DDC-Notationskandidaten, z. B. „823“ („Englische Erzählprosa“), weshalb *CN=0* ist. Hier muß *QTIT* verbessert werden.

DNB-K020 {DNB0985531282} im Grunde reiner Sachgruppenabgleich

Was ist mit „reiner Sachgruppenabgleich gemeint (**F26**)? *dno_like_i* „100“ (Philosophie“) wird durch *vc_dcl* mithilfe der Deskriptorwerte „{herzog, folia, ueberlegungen}“ ermittelt; deshalb *CN=1*.

DNB-K021 {DNB0985887842} Keine Colibri-Daten generiert

Durch *vc_dcl* werden folgende Daten generiert:

```

number of ddc-classified title: 12699
identifier (dno,schedno): DNB0985887842 (158.1,158.1)
DNB DDC notation (MAB2 field 700) {150}
calculated cutoff value: 0
title: NLP-Abendseminar
title (remainder): Live-Mitschnitte aus 2005
considered descriptor values: |5| {<100b>-stephan#landsiedel[0], <331>-nlp-
abendseminar[0], <335>-live-mitschnitte[0], <412@410>-<033A>-Âªs.Âª
landsiedel@wiesentheid[0], <540a>-3-940692-00-X[0]}
matched descriptor values: |0| {}
max. match value of matched descriptor values: |0|
calculated1 ddc classes (subdiv): |0| {}
calculated1 ddc classes (sections): |0| {}

```



```

calculated1 ddc classes (main):      |0| {}
calculated2 ddc classes (subdiv):    {}
calculated2 ddc classes (sections): {}
calculated2 ddc classes (divisions): {}
calculated2 ddc classes (main):      {}
correlation (DNB985887842,158.1): XXX.Xxx xxx xxx xxx (0)

```

Abb. 16: Klassifizierungsergebnis von *in_dnb_DNB0985887842*

Wie aus Abb. 16 zu sehen, ist die Menge der Deskriptorwerte, die zu (irgend)einer DDC-Klasse beitragen leer. Dies liegt daran, dass kein einziger Deskriptorwert in *vc_IDB* vorkommt (alle Häufigkeitswerte sind 0). Nur durch *QTIT*- und/oder *QIDB*-Verbesserung ist das Klassifizierungsergebnis ev. zu verbessern.

DNB-K022 {DNB0983516634} C-Notation zwar spezifischer, richtig ist aber wohl die DNB-Notation

Vergeben wurde *dno_like_i=100* („Philosophie“) und *dno_i=121* („Epistemologie (Erkenntnistheorie)“). Der Titel ist „Objektive Wirklichkeit und sinnliche Erfahrung - zum Verhaeltnis von Geist und Welt“. *vc_dcl* ermittelt mit den vier Deskriptorwerten „{skeptizismus, erkenntnistheorie, erfahrung, verhaeltnis}“ u. a. *dno_a=121* in *calc2 (sections)* und *dno_a=121.68* („Bedeutung, Interpretation, Hermeneutik“) in *calc2 (subdiv)*.

DNB-K023 {DNB0971590931} C-Notation ist hier gelungen

Keine besondere *vc_dcl*-Leistung, da (Deskriptorwerte des) Titeldatensatz(es) DNB971590931 (Datei: H0710ti2.dis) als *ppn_0389987832* mit *dno=170.92* („Ethik--Personen“) in *vc_IDB* enthalten. Mit den Deskriptorwerten „{unsaying, ewa#rychter, emmanuel, levinas's}“ wird *dno_a=170.92* ermittelt, daher *CN=1*.

DNB-K024 {DNB0985341882} Hier Bestergebnis von insgesamt 5 Subdivisions berücksichtigt

Das Klassifizierungsergebnis zum Titeldatensatz DNB985341882 ist:

```

number of ddc-classified title:      21392
identifier (dno,schedno):            DNB0985341882 (198,198)
DNB DDC notation (MAB2 field 700)   {100}
calculated cutoff value:             253
title:                               Kierkegaard und Aesthetik?
title (remainder):                  im Lichte der Erfahrung
considered descriptor values:        |5| {<100>-lin#riedesel[0], <331>-
kierkegaard[130], <331>-aesthetik?[0], <335>-erfahrung[250], <335>-lichte[253]}
matched descriptor values:           |3| {erfahrung, kierkegaard, lichte}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):    |5| {170.92, 192, 301, 370, 833.912}
calculated1 ddc classes (sections):  |4| {170, 192, 301, 370}
calculated1 ddc classes (main):      |2| {100, 300}
calculated2 ddc classes (subdiv):    {170.92[1]}

```



```

calculated2 ddc classes (sections): {170[1], 192[1], 301[1], 370[1]}
calculated2 ddc classes (divisions): {170[1], 190[1], 300[1], 370[1]}
calculated2 ddc classes (main):      {100[2], 300[2]}
correlation (DNB985341882,198): 110.xxx xxx xxx xxx (0.666667)

```

Abb. 17: Klassifizierungsergebnis von *in_dnb_DNB0985341882*

Wie am Ende des Kapitels 2.1.3 erwähnt, wird die automatische Bewertung zwischen *dno_i* (falls vorhanden) bzw. *dno_like_i* (falls kein *dno_i* verfügbar) und *calc2*-DDC-Notationskandidaten vorgenommen. Hier ist *dno_i*=198 und ein *dno_a*=190 (vgl. Fußnote 52) in *calc2* (*divisions*), deshalb ist $CN=2/3=0.67$.

DNB-K025 {DNB0979329787, DNB0983297096, DNB0985392584, DNB098053464X, DNB0982167733, DNB0984248048, DNB0979405130, DNB098089588X, DNB0982508069, DNB0982898851, DNB0983297134, DNB0985130032, DNB0985230320, DNB0985291052, DNB0985440287, DNB0985759267, DNB0985958278, DNB0985975601, ZDB1458852-3, DNB0985048298, DNB0985130059, DNB0985829168, DNB0984095594, DNB0984546863, DNB0980455448, DNB0981838219, DNB0984077502, DNB0971308691, DNB0984403310, DNB0985263334, DNB0985132698, DNB0985260831, DNB0985601809, DNB0985589744, DNB0985531282, DNB0974279315, DNB0971286884, DNB0971686688, DNB0971776369, DNB0969780850, DNB0972001409, DNB0985850191} Keine Subdivision vorhanden

Bei 18 (von 42) Titeldatensätzen wurde auch keine mehr als 3-stellige *dno_(like)_i* vergeben. Bei den anderen 24 Titeldatensätzen muß untersucht werden, ob es mit einer Verbesserung von *vc_dcl* möglich ist, häufiger spezifischere (mehr als 3-stellige) *dno_a*'s zu ermitteln. Eine leere *calc2* (*subdiv*)-Menge ist keine Eigenschaft der DDC-Klasse 100, sondern tritt auch bei den anderen DDC-Klassen auf: bei den DDC-Klassen 000, 200-900 liegt die Anzahl zwischen 31 und 62.

DNB-K026 {DNB0983529167, DNB0972057552, ZDB2389239-0, ZDB1500201-9, DNB0985177772, DNB0985759682, DNB0985768002, ZDB2287942-0, DNB0985589744, DNB0985531193} mehr als 15 C-Notationen; mehr als 15 C-Klassen

Bei diesen Titeldatensätzen ist die Anzahl der Deskriptorwerte entweder gleich Eins, die in einer DDC-Klasse vorkommen, und/oder diese Deskriptorwerte sind fallweise inhaltlich nicht aussagekräftig. DNB983529167: {geist}; ZDB2287942-0: {grupos}; DNB972057552: je einer von beiden Deskriptorwerten kommt in vielen DDC-Klassen vor: {phantasiereise, meditationen}; ZDB2389239-0: {epistemology}; ZDB1500201-9: {schriftenreihe}; DNB985177772: je einer der drei Deskriptorwerte kommt in vielen DDC-Klassen vor: {uebereinstimmung, persoenlichkeitsforschung, emotionaler}; DNB985531193: je einer von sechs Deskriptorwerten kommt in vielen DDC-Klassen vor: {absurde, grenzsituationen, erleben, folia, gendiagnostik, absurdaet}; DNB985589744: je zwei von fünf Deskriptorwerten kommen in vielen DDC-Klassen vor: {kulturwissenschaften, chinesische, trendforschung, Schriftenreihe, ning#huang}. Bei DNB985759682 und DNB985768002 kommen



jeweils alle fünf Deskriptorwerte in vielen DDC-Klassen vor: DNB985759682: {evolutionary, approaches, psychology, case, personality} und DNB985768002: {year, during, joint, attention, first}.

DNB-K027 {DNB0981950043, DNB0985291044, DNB0985531193, DNB0985531282, DNB0974279315, DNB0971252025, DNB0971286884, DNB0971590931, DNB0971686688, DNB0971776369, DNB0978184912, ZDB2287942-0, DNB0969780850, DNB0972001409, DNB097174372X} keine DDC-Notation vorhanden; keine DNB-DDC im eigentlichen Sinn vorhanden; keine DNB-DDC vergeben; Keine DNB-DDC vorhanden.

Nur *dno_like_i*'s vergeben. Bei neun Titeldatensätzen ist *CN=1*: DNB981950043, DNB985291044, DNB985531282, DNB974279315, DNB971590931 (in *vc_IDB!*), DNB971686688, DNB971776369, DNB978184912 und DNB969780850; bei sechs Titeldatensätzen ist *CN=0*: DNB985531193, DNB971252025, DNB971286884, ZDB2287942-0, DNB972001409 und DNB97174372X.

DNB-K028 {DNB0971286884, DNB0971590931, DNB0971686688, DNB0971776369, DNB0978184912, DNB0969780850, DNB0972001409, DNB097174372X} Datensatz von 2004; Titel von 2004

Wieso sind hier keine *dno_j* vergeben (**F27**)? Die Datenlieferung stammt aus dem Jahr 2007. Bei fünf (von acht) Titeldatensätzen ist *CN=1*: DNB971590931, DNB971686688, DNB0971776369, DNB978184912 und DNB969780850; bei drei (von acht) Titeldatensätzen ist *CN=0*: DNB0971286884, DNB972001409 und DNB97174372X.

DNB-Excel-Datei-Kommentare zur Hauptklasse 200

DNB-K029 {DNB098573390X, DNB0968944728} more than 15 calculated ddc classes, ich muss C also einfach glauben, denn die Notation sehe ich nicht.

Vgl. Argumentation in DNB-K007. Die ausführliche Ergebnisausgabe enthält bei DNB98573390X in *calc2 (subdiv) dno_a=231.044 (CN=1)* und bei DNB968944728 kommt *dno_a=238...* in *calc2 (subdiv)* vor (wegen *dno_j=238.2* ist *CN=3/4=0.75*). Der einzige Deskriptorwert „trinitaet“ (DNB98573390X) kommt bzw. die beiden Deskriptorwerte „glaube“ und „leben“ (DNB968944728) kommen gemeinsam in mehr als 15 DDC-Klassen vor.

DNB-K030 {DNB0982818386} keine Notation, Colibri kann nicht zwischen SG und Notation unterscheiden

vc_dcl kann zwischen *dno_i*'s und *dno_like_i*'s unterscheiden (Ausnahme: Argumentation unter DNB-K103). Wie in Kapitel 2.2.1 und an anderen Stellen erwähnt, wird (nur) dann eine automatische



Bewertung zwischen *dno_like_i* und *dno_a* vorgenommen, wenn keine *dno_i* verfügbar ist (um überhaupt eine automatische Bewertung zu erhalten). Wie in Beispiel 24 angemerkt, können diese Fälle selbstverständlich voneinander getrennt bestimmt werden. Für den Titeldatensatz DNB982818386 wurde die 2-stellige *dno_i*=230 vergeben, die *dno_a*'s beginnen alle mit „2“. Die automatische Bewertung ist daher $CP=10x.xxx\ xxx\ xxx\ xxx$ und $CN=1/2=0.5$ (Abb. 18). In *calc2* (*subdiv*) wird *dno_a*=282.092 („Kardinäle, ...“) und *dno_a*=284.1092 („Luther, Martin ...“) von *vc_dcl* ermittelt.

```

number of ddc-classified title:      10728
identifier (dno,schedno):           DNB0982818386 (230,230)
DNB DDC notation (MAB2 field 700)  {230}
calculated cutoff value:           93
title:                              Peking brueskiert den Vatikan
title (remainder):                 eigenmaechtige Ernennung eines Bischofs in
Kunming
considered descriptor values:       |9| {<100>-anton-heinz#schmidt[2], <331>-
vatikan[93], <331>-peking[374], <331>-brueskiert[0], <335>-bischofs[24], <335>-
kunming[23], <335>-eigenmaechtige[1], <335>-ernennung[5], <412@410>-<033A>-Ãªa.-
h.Ãª schmidt@
en-voglhuh[0]}
matched descriptor values:         |3| {kunming, bischofs, vatikan}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):  |5| {282.092, 284.1092, 327, 915, 943}
calculated1 ddc classes (sections): |4| {282, 284, 915, 943}
calculated1 ddc classes (main):     |2| {200, 900}
calculated2 ddc classes (subdiv):   {282.092[1], 284.1092[1]}
calculated2 ddc classes (sections): {282[1], 284[1], 915[1], 943[1]}
calculated2 ddc classes (divisions):{280[2]}
calculated2 ddc classes (main):     {200[2], 900[2]}
correlation (DNB982818386,230): 10x.xxx xxx xxx xxx (0.5)

```

Abb. 18: Klassifizierungsergebnis von *in_dnb_DNB0982818386*

DNB-K031 {DNB0984780521} kann nicht zw. HSG und Notation unterscheiden, wertet sich hier mit 0.5 Treffern, aber es ist gar keine Notation vorhanden

Argumentation, *dno_like_i*, *dno_a*, *CP* und *CN* wie bei DNB-K030.

DNB-K032 {ZDB2383435-3} wertet hier wieder SG als Notationstreffer

Argumentation, *dno_like_i*, *dno_a*, *CP* und *CN* wie bei DNB-K030.

DNB-K033 {DNB0984741852} wertet SG als Notation

Argumentation wie bei DNB-K030. *dno_like_i*=230 und *dno_like_i*=360, in *calc2* (*divisions*) *dno_a*=230, $CP=11x.xxx\ xxx\ xxx\ xxx$ und $CN=1$.



DNB-K034 {DNB0986173746} Nimmt hier die Zweitnotation und übergeht die Hauptnotation, hier hätten zumindest die ersten beiden Stellen getroffen.

Ja, s. Fußnote 26. In geänderter Programmversion *vc_dcl* (Oktober 2008) ist das Klassifizierungsergebnis (CN=0.5):

```

number of ddc-classified title:      12301
identifier (dno,schedno):           DNB0986173746 (330.901,330.9)
DNB DDC notation (MAB2 field 700):  {330}
DDC notation (MAB2 field 705):      {330.901,211.34}
calculated cutoff value:            3899
title:                              Monotheism from a sociopolitical and economic
perspective
title (series):                     Discussion paper series / IZA ; No. 3116
considered descriptor values:        |10| {<100b>-murat#iyigun[0], <331>-
sociopolitical[119], <331>-economic[3539], <331>-perspective[3899], <331>-
monotheism[53], <412@410>-<033A>-iza@bonn[1], <451>-no.[14], <451>-
discussion[1416], <451>-paper[2462], <451>-series[3528]}
matched descriptor values:          |7| {discussion, sociopolitical, series, paper,
monotheism, perspective, economic}
max. match value of matched descriptor values: |6|
calculated1 ddc classes (subdiv):    |13| {230, 301, 305.896073, 306, 306.09,
333.75, 370.19, 621.3678, 657, 658, 658.3124, 658.4038, 822.33}
calculated1 ddc classes (sections):  |5| {301, 305, 306, 333, 370}
calculated1 ddc classes (main):      |1| {300}
calculated2 ddc classes (subdiv):    {305.896073[1], 306.09[1], 333.75[1],
370.19[1]}
calculated2 ddc classes (sections):  {306[2]}
calculated2 ddc classes (divisions): {300[4]}
calculated2 ddc classes (main):      {300[6]}
correlation (dnb_B0747_DNB0986173746,330.9): 110.0xx xxx xxx xxx (0.5)

```

Abb. 19: Klassifizierungsergebnis von *in_dnb_* DNB0986173746 mit *vc_dcl* (Oktober 2008)

DNB-K035 {DNB097401852X} Wertet permanent SG übereinstimmungen mit CN-Wert 1, daher die unglaublich hohe Trofferquote von 20%, die so also keineswegs stimmt, ich schätze maximal 5-10%, mehr auf keinen Fall.

Vgl. Argumentation unter DNB-K030. Hier sind *dno_like_i's*={220,230,480}, es gibt kein *dno_i*. In *calc2* (*divisions*) ist *dno_a*=220, daher *CP*= 11x.xxx xxx xxx xxx und *CN*=1. Ad. „die unglaublich hohe Trofferquote von 20%, die so also keineswegs stimmt, ich schätze maximal 5-10%, mehr auf keinen Fall.“ Die Prozentzahlen können exakt bestimmt werden, s. Kapitel 4.

DNB-K036 {DNB0974864315} HSG und NSG nicht getroffen und dennoch CN-Wert 0,5, das ist nach meiner Meinung unlauterer Wettbewerb.

Vgl. Argumentation unter DNB-K030. Hier sind *dno_like_i's*={230,290,930}, kein *dno_i* gegeben. In *calc2* (*main*) ist *dno_a* = 200, daher ist *CP*= 10x.xxx xxx xxx xxx und *CN*=1/2=0.5. Welche *dno_a* sollte für den Titeldatensatz DNB974864315 mit dem Titel „Christliche Identitaet und Assimilation in



der Spaetantike - Probleme des Christseins in der Reflexion der Zeitgenossen.“ vergeben werden? Ist nicht zu vermuten, dass diese mit „2“ beginnt? *vc_dcl* ermittelt in *calc2 (sections)* die *dno_a=274* („Christentum in Europa Christliche Kirchen in Europa“). Ist die automatische Klassifizierung in diesem Fall nicht treffend (**F28**)?

DNB-Excel-Datei-Kommentare zur Hauptklasse 300

DNB-K037 {DNB0984840664} lt. CN-Wertt offenbar 330.1 getroffen

Ja, *dno_a=330.1* ist in *calc2 (subdiv)* enthalten, vgl. Argumentation unter DNB-K007. In *calc2 (sections)* wird *dno_a=338* mit einer Häufigkeit von „21“; in *calc2 (divisions)* *dno_a=330* mit einer Häufigkeit von „43“ und in *calc2 (main)* *dno_a=300* mit einer Häufigkeit von „108“ ermittelt. *CP=111.1xx xxx xxx xxx* und *CN=1*. Ursache für die großen Häufigkeitswerte liegen im Titel „Information and incentives in organizations“ mit allgemeinen Deskriptorwerten – die in sehr vielen DDC-Klassen vorkommen – begründet: „<100b>-julia#nafziger[0]“, „<331>-organizations[1428]“, „<331>-information[4285]“, „<331>-incentives[313]“. Für den Titeldatensatz wurden keine Sachschlagwörter vergeben. Die drei Deskriptorwerte „information“, „incentives“, „organizations“ kommen alle gemeinsam in den von *vc_dcl* ermittelten DDC-Klassen vor.

DNB-K038 {DNB0982155506} keine Feinddc und kein SW am Titel

Ja, es wurde nur *dno_like_j=320* („Politik“) und keine Sachschlagwörter vergeben, *vc_dcl* ermittelt mit dem einzigen (fragwürdigen) Deskriptorwert „sackgasse“ u. a. *dno_a=320.947* („Osteuropa--Politik, . . .“). *CN=2/2=1*.

DNB-K039 {ZDB2385040-1} anhand des Wortes "Geschäftsbericht" findet Colibri die für den Geschäftsbericht korrekte Notation. Für diesen Titel trifft sie allerdings nicht zu;

Es wurde *dno_like_j= 330* („Wirtschaft“) vergeben. In diesem Fall ist in *calc1 (subdiv)* *dno_a=330* („Wirtschaft“) enthalten (Abb. 20). Da die automatische Bewertung jedoch mit *dno_a's* aus *calc2* und nicht aus *calc1* durchgeführt wird (vgl. Kapitel 2.1.3, letzter Satz), ist *CN=0* und nicht *CN=1*.

```

number of ddc-classified title:      11095
identifier (dno,schedno):           ZDB2385040-1 (330,330)
DNB DDC notation (MAB2 field 700)  {330,001}
calculated cutoff value:            6
title:                              Geschaeftsbericht
considered descriptor values:       |2| {<331>-geschaeftsbericht[6], <412@410>-
<033A>-industriellenvereinigung tirol@innsbruck[0]}
matched descriptor values:          |1| {geschaeftsbericht}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv):   |6| {330, 331.89, 657.3, 657.32, 658.404,

```



```

808.066651}
calculated1 ddc classes (sections): | 2 | {657, 658}
calculated1 ddc classes (main):      | 1 | {600}
calculated2 ddc classes (subdiv):    {657.32[1], 657.3[1], 658.404[1]}
calculated2 ddc classes (sections):  {657[2]}
calculated2 ddc classes (divisions): {650[3]}
calculated2 ddc classes (main):      {600[3]}
correlation (ZDB2385040-1,330): 00x.xxx xxx xxx xxx (0)

```

Abb. 20: Klassifizierungsergebnis von *in_dnb_ZDB2385040-1*

DNB-K040 {DNB0978604490} zieht unbesetzte DDC-Klassen als möglich heran (574, 589)

Die DDC-Notationen „574“ und „589“ sind seit Ausgabe 21 „Unbesetzt“. *vc_dcl* schlägt (trotzdem) DDC-Notationskandidaten vor, die mit „574“ und „589“ beginnen, da sie in der *DDC-Datenbasis vc_DB* enthalten sind. Durch eine Programmänderung könnten die Deskriptorwerte der *GVK-DDC*-Titel datensätze in *vc_DB* ausgeschlossen werden, deren DDC-Notationen stillgelegt wurden. Nach Bildung einer neuen DDC-Datenbasis könnte *vc_dcl* dann keine DDC-Notationskandidaten mehr vorschlagen, deren Klassen unbesetzt sind.

DNB-K041 {ZDB2391837-8, DNB0982155506, DNB0984216308, DNB0985132639, ZDB2176386-0, DNB0985811056, ZDB1302632-x, DNB0983907641, DNB0986100102, DNB0984840664} more than 15 calculated ddc classes.

Vgl. Argumentation unter DNB-K007. Die automatische Bewertung liefert in diesen Fällen:

```

correlation (ZDB2391837-8,320): 00x.xxx xxx xxx xxx (0)
correlation (DNB982155506,320): 11x.xxx xxx xxx xxx (1)
correlation (DNB984216308,390): 00x.xxxx xxx xxx xxx (0)
correlation (DNB985132639,363.46): 111.11x xxx xxx xxx (1)
correlation (ZDB2176386-0,330): 11x.xxx xxx xxx xxx (1)
correlation (DNB985811056,331.114): 110.000 xxx xxx xxx (0.333333)
correlation (ZDB1302632-x,360): 11x.xxx xxx xxx xxx (1)
correlation (DNB983907641,338.1762): 111.000 0xxx xxx xxx (0.428571)
correlation (DNB986100102,333.79): 111.11x xxx xxx xxx (1)
correlation (DNB984840664,330.1): 111.1xx xxx xxx xxx (1)

```

DNB-K042 {DNB0972751319, DNB0973043199, DNB097356105X, DNB0973575913, DNB0973577770, DNB0974271993, DNB0974325899, DNB0974447870, DNB0974454877, DNB0975048147, DNB0975119605} wertet aber SG als Notation, Reihe B statt H.

Bei diesen Titeldatensätzen wurde jeweils nur eine *dno_like_i* (in *correlation*-Zeile hinter dem Komma stehend) vergeben, *CP* und *CN* sind:

```

correlation (DNB972751319,340): 11x.xxx xxx xxx xxx (1)
correlation (DNB973043199,370): 11x.xxxx xxx xxx xxx (1)
correlation (DNB97356105X,340): 11x.xxxx xxx xxx xxx (1)
correlation (DNB973575913,380): 10x.xxx xxx xxx xxx (0.5)

```



```

correlation (DNB973577770,330): 00x.xxxx xxxx xxxx xxxx (0)
correlation (DNB974271993,330): 11x.xxxx xxxx xxxx xxxx (1)
correlation (DNB974325899,340): 11x.xxx  xxx  xxx  xxx  (1)
correlation (DNB974447870,340): 11x.xxxx xxxx xxxx xxxx (1)
correlation (DNB974454877,320): 11x.xxxx xxxx xxxx xxxx (1)
correlation (DNB975048147,340): 11x.xxxx xxxx xxxx xxxx (1)
correlation (DNB975119605,380): 11x.xxxx xxxx xxxx xxxx (1)

```

Reihe B statt H (**F29**)? Alle Titeldatensätze sind in Reihe H (Datei: H0708ti2.dis) enthalten.

DNB-K043 {DNB0980234662, DNB0980357268, DNB0981962009, DNB0981978711, DNB0982462174, DNB0983608946, DNB0983696756, DNB0984254048, DNB0984422730, DNB098480661X, DNB0985050268} Reihe A statt H.

Reihe A statt H (**F30**)? Alle Titeldatensätze sind in Reihe H (Datei: H0709ti2.dis) enthalten. Die automatische Bewertung ist:

```

correlation (DNB980234662,344): 110.xxxx xxxx xxxx xxxx (0.666667)
correlation (DNB980357268,346): 111.xxxx xxxx xxxx xxxx (1)
correlation (DNB981962009,345): 111.xxx  xxx  xxx  xxx  (1)
correlation (DNB981978711,332.6322): 111.111 1xx xxx xxx (1)
correlation (DNB982462174,343): 110.xxxx xxxx xxxx xxxx (0.666667)
correlation (DNB983608946,338.4766): 000.000 0xx xxxx xxxx (0)
correlation (DNB983696756,355.031): 000.000 xxx xxxx xxxx (0)
correlation (DNB984254048,370.113): 111.000 xxx xxx xxx (0.5)
correlation (DNB984422730,338.18): 110.00x xxx xxxx xxxx (0.4)
correlation (DNB98480661X,342.73085): 110.000 00x xxxx xxxx (0.25)
correlation (DNB985050268,338.476292): 110.000 000 xxxx xxxx (0.222222)

```

DNB-K044 {DNB0981010954, DNB0983624356} vergleicht mit zweiter DNB-Notation

Die verbesserte `vc_dcl`-Programmversion (Oktober 2008) ermittelt für DNB981010954:

```

number of ddc-classified title:      7892
id (title record in vc_DB!,schedno):DNB0981010954 (340.03,340)
DNB DDC notation (MAB2 field 700):  {340,330}
DDC notation (MAB2 field 705):      {340.03,330.03}
calculated cutoff value:            72
title:                               Woerterbuch der Rechts- und Wirtschaftssprache
- Franzoesisch-Deutsch, Deutsch-Franzoesisch
title (series):                      BeckLex
considered descriptor values:        |11| {<100b>-michel#doucet[2], <331>-
wirtschaftssprache[72], <331>-franzoesisch-deutsch[12], <331>-deutsch-
franzoesisch[16], <331>-woerterbuch[1271], <331>-rechts[643], <412@410>-<033A>-
beck@muenchen[688], <451b>-becklex[0], <540a>-3-406-55375-3[0], <912s>-recht[1727],
<917s>-wirtschaft[1821]}
matched descriptor values:           |4| {wirtschaftssprache, deutsch-franzoesisch,
michel#doucet, franzoesisch-deutsch}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):    |3| {330, 340, 440}
calculated1 ddc classes (sections):  |2| {330, 340}
calculated1 ddc classes (main):      |1| {300}
calculated2 ddc classes (subdiv):    {}

```



```

calculated2 ddc classes (sections): {330[1], 340[1]}
calculated2 ddc classes (divisions): {330[1], 340[1]}
calculated2 ddc classes (main):      {300[2]}
correlation (dnb_A0748_DNB0981010954,340): 11x.xxx xxx xxx (1) !

```

Abb. 21: Klassifizierungsergebnis von *in_dnb_DNB0981010954* mit *vc_dcl* (Oktober 2008)

Die intellektuelle Klassifizierung ist 1. *dno_like_i=340*, 2. *dno_like_i=330*, 1. *dno_i=340.03* (wird auf Haupttafelnotation *schedno=340* gekürzt, vgl. Kapitel 2.1.1, letzter Abschnitt) und 2. *dno_j=330.03*. *vc_dcl* bestimmt als DDC-Notationskandidaten *dno_a=330* und *dno_a=340*. Die automatische Bewertung erfolgt für 1. *dno_i=340*. Da *dno_a=340* in den DDC-Notationskandidaten vorkommt, ist $CN=1$. *vc_dcl* erkennt, daß die Deskriptorwerte in *vc_IDB* enthalten sind (zu erkennen an „!“ in Zeilen „id ...“ und „correlation ...“), die zu dem Titeldatensatz (*ppn_0388455861*) mit *dno=340.03* gehören.

DNB-K045 {DNB098393083X} vergleicht mit NSG

Was heisst: vergleicht mit NSG (**F31**)? Die intellektuelle Klassifizierung für den Titeldatensatz DNB98393083X ist *dno_like_i=370*, *dno_like_i=150* und *dno_j=371.10019*. In *calc2 (sections)* ist *dno_a=150* enthalten, die automatische Bewertung ist daher $CN=0$. Allerdings wird mit *vc_dcl / vc_IDB* (Dezember 2008) festgestellt, dass die Deskriptorwerte des / der Titeldatensatz als *ppn_0528369695* in *vc_IDB* mit *dno=150* und *dno=370* enthalten ist, daher ist $\max. CN=2/8=0.25$ erreichbar.

DNB-K046 {DNB0981365140} RB; 11 Seiten, keine DNB-Notation vorhanden

Was bedeutet der „Kommentar: RB; 11 Seiten“, was für die automatische Klassifizierung (**F32**)? Es wurde nur *dno_like_i=360* vergeben, in *calc2 (divisions)* ist *dno_a=360*, daher $CN=1$.

DNB-K047 {DNB0985147911} RB, 7 Blatt

Was bedeutet „RB, 7 Blatt“ für die automatische Klassifizierung (**F33**)? Einzige *dno_like_i=320* ist mit *vc_dcl / vc_IDB* (Dezember 2008) in *calc2 (sections)* enthalten, daher $CN=1$.

DNB-K048 {DNB0984384448} 333.72 ist die 2. Notation, Colibri wertet auch diese aus; more than 15 calculated ddc classes

Der Titeldatensatz DNB984384448 enthält *dno_like_i=360*, *dno_like_i=330*, *dno_i= 363.738747* und *dno_j=333.72*. Bei mehreren *dno_i*'s wird in *vc_dcl* (April 2008) nur letzte *dno_i* (s. Fußnote 26) berücksichtigt. Die Ausgabe ist bei mehr als 15 DDC-Klassen reduziert (s. DNB-K007). In *vc_dcl* (Oktober 2008) wird 1. *dno_i=363.738747* (*schedno=363.73874* („Treibhausgase“)) berücksichtigt. $CN=1$, da *schedno* von 1. *dno_i* in *calc2 (subdiv)* enthalten.

DNB-K049 {DNB0970135963} wertet aber SG als Notation, Reihe A statt H



Wg. „SG“ vgl. DNB-K030. DNB970135963 ist in Reihe H (Datei: H0710ti2.dis) enthalten. Da $dno_like_j=320$ und $dno_a=300$ ist $CN=0.5$.

DNB-K050 {DNB0971226458, DNB0971668957} wertet aber SG als Notation, Reihe A statt H, Hat DNB keine Notation gegeben, wählt Programm aus seinen möglichen Treffern keinen präferierten Treffer aus.

Wg. „SG“ vgl. DNB-K030. DNB971226458 und DNB971668957 sind in Reihe H (Datei: H0710ti2.dis) enthalten. Was bedeutet: „Hat DNB keine Notation gegeben, wählt Programm aus seinen möglichen Treffern keinen präferierten Treffer aus“ (**F34**)? In DNB971226458 ist $dno_like_j=340$ und in *calc2 (divisions)* ist $dno_a=340$ enthalten. Die automatische Bewertung ist „correlation (DNB971226458,340): 1 | x.xxx xxx xxx xxx (1)“. In DNB971668957 selber Sachverhalt.

DNB-K051 {DNB0953015688} Ac-Satz, vergleicht mit NSG

Was ist ein „Ac-Satz“ (**F35**)? Was heisst „vergleicht mit NSG“ (**F36**)? DNB953015688 enthält 1. $dno_like_j=320$ und 2. $dno_like_j=070$, in *calc2 (divisions)* ist $dno_a=300$. Deshalb ist $CN=1/2=0.5$.

DNB-K052 {DNB0970426135, DNB0971054037, DNB0971431353} Reihe A statt H, Hat DNB keine Notation gegeben, wählt Programm aus seinen möglichen Treffern keinen präferierten Treffer aus.

Vgl. DNB-K050. Alle drei Titeldatensätze sind in Reihe H (Datei: H0711ti2.dis) enthalten. Für DNB970426135 wurde nur $dno_like_j=340$ vergeben. In *calc2 (divisions)* ist $dno_a=340$ enthalten., weshalb $CN=1$ ist. Dto. für DNB971054037 und DNB971431353.

DNB-Excel-Datei-Kommentare zur Hauptklasse 400

DNB-K053 {DNB0984714243, DNB0984768718, DNB0984778039, DNB0984921443, DNB0940187507} More than 15 calculated ...

Vgl. DNB-K007.

DNB-K054 {DNB0984013725, DNB0985764368, DNB0986233765, DNB098623978X} Sf-Satz!

Was ist ein „Sf-Satz“ (**F37**), was bedeutet es für die automatische Klassifizierung? Die Titeldatensätze haben keinen Titel (keine MAB2-Felder 310, 331, 335, 341 und 451 => „title lines: {}“) und somit keinen Deskriptorwert, mit dem DDC-Notationskandidaten ermittelt werden könnten.



DNB-K055 {DNB0973618329} "Partizip" nicht matched descriptor value!

An diesem Titeldatensatz sollen exemplarisch die Schwierigkeiten einer automatischen Klassifizierung – im Unterschied zu einer (einfachen) intellektuellen Klassifizierung – aufgezeigt werden. Auf Anhieb ist (als Mensch) erkennbar, dass „Partizip“ der einschlägige Deskriptorwert ist, um den Titeldatensatz in die DDC-Hauptklasse 400 („Sprache“) [bzw. zusammen mit dem Deskriptorwert „Deutschen“ in die DDC-Hauptklasse 430 („Germanische Sprachen Deutsch“)], zu klassifizieren. Zwar kommt in *vc_IDB* der Deskriptorwert „partizip“ nur in vier unterschiedlichen DDC-Klassen (435, 437.02, 440 und 465.8) vor, jedoch führen die ersten drei der vier Deskriptorwerte „germanistik“, „typologische“, „danziger“ und „polnischen“ in *calc2 (subdiv)* gemeinsam zu der (anderen, als mit „4“ beginnenden) DDC-Klasse „833.914“ („Deutsche Erzählprosa--1945–1990“) (Abb. 22).

```

number of ddc-classified title:      13251
identifier (dno,schedno):           DNB0973618329 (490,490)
DNB DDC notation (MAB2 field 700):  {490,430}
calculated cutoff value:            70
title:                              ÂªDasª Partizip im Deutschen und Polnischen
title (remainder):                  eine typologische Studie
title (series):                      Danziger Beitræge zur Germanistik ; Bd. 18
considered descriptor values:       |14| {<100>-andrzej#feret[0], <331>-
deutschen[3096], <331>-polnischen[70], <331>-partizip[4], <335>-eine[3865], <335>-
typologische[23], <335>-studie[951], <412@410>-<033A>-lang@frankfurt[11], <451>-
germanistik[52], <451>-danziger[33], <451>-beitraege[1660], <540a>-3-631-53513-
9[0], <902g>-polnisch[77], <902g1>-deutsch[1517]}
matched descriptor values:          |4| {germanistik, typologische, danziger,
polnischen}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):   |2| {833.914, 914}
calculated1 ddc classes (sections): |2| {833, 914}
calculated1 ddc classes (main):     |2| {800, 900}
calculated2 ddc classes (subdiv):   {833.914[1]}
calculated2 ddc classes (sections): {833[1], 914[1]}
calculated2 ddc classes (divisions):{830[1], 910[1]}
calculated2 ddc classes (main):     {800[1], 900[1]}
correlation (DNB973618329,490): 00x.xxx xxx xxx xxx (0)

```

Abb. 22: Klassifizierungsergebnis von *in_dnb_DNB0973618329*

Werden die Deskriptorwerte des MAB2-Feldes 451 („Gesamttitle in Vorlageform“) gelöscht, wird ein anderer *cutoff_val_dyn* (119 statt 70) und andere Deskriptorwerte (die zum Klassifizierungsergebnis beitragen) ermittelt und *CN=1* erreicht (Abb. 23).

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0973618329 (490,490)
DNB DDC notation (MAB2 field 700):  {490,430}
calculated cutoff value:            119
title:                              ªDasª Partizip im Deutschen und Polnischen
title (remainder):                  eine typologische Studie
title (series):                      Danziger Beitræge zur Germanistik ; Bd. 18
considered descriptor values:       |11| {<100>-andrzej#feret[0], <331>-
deutschen[6541], <331>-polnischen[97], <331>-partizip[4], <335>-eine[8273], <335>-

```



```

typologische[24], <335>-studie[1336], <412@410>-<033A>-lang@frankfurt[11], <540a>-
3-631-53513-9[0], <902g>-polnisch[119], <902g1>-deutsch[3083]}
matched descriptor values:          |3| {typologische, polnisch, polnischen}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):    |1| {491.8}
calculated1 ddc classes (sections):  |1| {491}
calculated1 ddc classes (main):      |1| {400}
calculated2 ddc classes (subdiv):    |1| {491.8[1]}
calculated2 ddc classes (sections):  {491[1]}
calculated2 ddc classes (divisions): {490[1]}
calculated2 ddc classes (main):      {400[1]}
correlation (dnb_H0708_DNB0973618329,490): 11x.xxxx xxx xxx xxx (1)

```

Abb. 23: Klassifizierungsergebnis von *in_dnb_DNB0973618329* mit gelöschten Deskriptorwerten aus MAB2-Feld 451 („Gesamttitle in Vorlageform“)

Werden zusätzlich die Deskriptorwerte des MAB2-Feldes 335 („Zusätze zum Hauptsachtitel“) gelöscht, ändern sich wiederum die Deskriptorwerte, die für die Klassifizierung berücksichtigt werden. Der CN-Wert ist immer noch CN=1, allerdings erhöht sich die Anzahl der DDC-Notationskandidaten in *calc2 (subdiv)* (Abb. 24):

```

number of ddc-classified title:      2
identifier (dno,schedno):            DNB0973618329 (490,490)
DNB DDC notation (MAB2 field 700):   {490,430}
calculated cutoff value:             119
title:                               aDasa Partizip im Deutschen und Polnischen
title (remainder):                  eine typologische Studie
title (series):                      Danziger Beitrage zur Germanistik ; Bd. 18
considered descriptor values:        |8| {<100>-andrzej#feret[0], <331>-
deutschen[6541], <331>-polnischen[97], <331>-partizip[4], <412@410>-<033A>-
lang@frankfurt[11], <540a>-3-631-53513-9[0], <902g>-polnisch[119], <902g1>-
deutsch[3083]}
matched descriptor values:          |4| {<033A>-lang@frankfurt, polnisch,
polnischen, partizip}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |35| {010, 303.482430438, 306.4494386, 327.47,
330, 370, 398.9094, 413.028, 415, 430, 435, 490, 491.8, 491.8509, 491.852431,
491.855, 491.8554, 491.85592, 491.858, 491.8580231, 792.09438, 800, 830, 890,
891.553309358, 891.7090044, 891.8, 891.8509, 891.8509007, 891.85090073, 891.850911,
891.850932438, 891.851709, 943.8, 943.8025}
calculated1 ddc classes (sections):  |6| {413, 415, 430, 435, 490, 491}
calculated1 ddc classes (main):      |1| {400}
calculated2 ddc classes (subdiv):    |8| {413.028[1], 491.8509[1], 491.852431[1],
491.8554[1], 491.85592[1], 491.855[1], 491.858[1], 491.8[1]}
calculated2 ddc classes (sections):  {491[8]}
calculated2 ddc classes (divisions): {490[9]}
calculated2 ddc classes (main):      {400[13]}
correlation (dnb_H0708_DNB0973618329,490): 11x.xxxx xxx xxx xxx (1)

```

Abb. 24: Klassifizierungsergebnis von *in_dnb_DNB0973618329* mit gelöschten Deskriptorwerten aus den MAB2-Feldern 335 („Zusätze zum Hauptsachtitel“) und 451 („Gesamttitle in Vorlageform“)

Wird zusätzlich der Deskriptorwert „Verleger@Ort“ gelöscht, bleibt immer noch CN=1 (Abb. 25). Werden nur die Deskriptorwerte („polnisch“, „partizip“, „deutsch“) des Deskriptors „Schlagwort“ berücksichtigt, wird immerhin noch ein Klassifizierungsergebnis mit CN=0.5 (Abb. 26) von *vc_dcl*



geliefert. CN=0.5 wird auch durch den einzigen Deskriptorwert „partizip“ (Abb. 27) geliefert. An diesem Beispiel ist zu sehen, dass nicht unbedingt die (große) Anzahl an verfügbaren Deskriptorwerten für die Qualität der Klassifizierung ausschlaggebend ist und was für einen Balanceakt die automatische Klassifizierung darstellt.

```

number of ddc-classified title:      3
identifier (dno,schedno):           DNB0973618329 (490,490)
DNB DDC notation (MAB2 field 700):  {490,430}
calculated cutoff value:            119
title:                              aDasa Partizip im Deutschen und Polnischen
title (remainder):                  eine typologische Studie
title (series):                      Danziger Beitrage zur Germanistik ; Bd. 18
considered descriptor values:        |7| {<100>-andrzej#feret[0], <331>-
deutschen[6541], <331>-polnischen[97], <331>-partizip[4], <540a>-3-631-53513-9[0],
<902g>-polnisch[119], <902g1>-deutsch[3083]}
matched descriptor values:           |3| {polnisch, polnischen, partizip}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |34| {010, 303.482430438, 306.4494386, 327.47,
370, 398.9094, 413.028, 415, 430, 435, 490, 491.8, 491.8509, 491.852431, 491.855,
491.8554, 491.85592, 491.858, 491.8580231, 792.09438, 800, 830, 890, 891.553309358,
891.7090044, 891.8, 891.8509, 891.8509007, 891.85090073, 891.850911, 891.850932438,
891.851709, 943.8, 943.8025}
calculated1 ddc classes (sections):  |6| {413, 415, 430, 435, 490, 491}
calculated1 ddc classes (main):       |1| {400}
calculated2 ddc classes (subdiv):     |8| {413.028[1], 491.8509[1], 491.852431[1],
491.8554[1], 491.85592[1], 491.855[1], 491.858[1], 491.8[1]}
calculated2 ddc classes (sections):  {491[8]}
calculated2 ddc classes (divisions): {490[9]}
calculated2 ddc classes (main):       {400[13]}
correlation (dnb_H0708_DNB0973618329,490): 11x.xxxx xxxx xxxx xxxx (1)

```

Abb. 25: Klassifizierungsergebnis von *in_dnb_DNB0973618329* nur mit Deskriptorwerten des Deskriptors „Hauptsachtitel“ (MAB2-Feld 331) und „Schlagwort“ (MAB2-Feld 902)

```

number of ddc-classified title:      4
identifier (dno,schedno):           DNB0973618329 (490,490)
DNB DDC notation (MAB2 field 700):  {490,430}
calculated cutoff value:            119
title:                              aDasa Partizip im Deutschen und Polnischen
title (remainder):                  eine typologische Studie
title (series):                      Danziger Beitrage zur Germanistik ; Bd. 18
considered descriptor values:        |5| {<100>-andrzej#feret[0], <540a>-3-631-
53513-9[0], <902g>-polnisch[119], <902s1>-partizip[4], <902g1>-deutsch[3083]}
matched descriptor values:           |2| {polnisch, partizip}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |1| {435}
calculated1 ddc classes (sections):  |1| {435}
calculated1 ddc classes (main):       |1| {400}
calculated2 ddc classes (subdiv):     |0| {}
calculated2 ddc classes (sections):  {435[1]}
calculated2 ddc classes (divisions): {430[1]}
calculated2 ddc classes (main):       {400[1]}
correlation (dnb_H0708_DNB0973618329,490): 10x.xxx xxx xxx xxx (0.5)

```

Abb. 26: Klassifizierungsergebnis von *in_dnb_DNB0973618329* ohne Deskriptorwerte des Deskriptors „Hauptsachtitel“, sondern nur mit Deskriptorwerten des Deskriptors „Schlagwort“



```

number of ddc-classified title:      5
identifier (dno,schedno):           DNB0973618329 (490,490)
DNB DDC notation (MAB2 field 700):  {490,430}
calculated cutoff value:            4
title:                              aDasa Partizip im Deutschen und Polnischen
title (remainder):                  eine typologische Studie
title (series):                      Danziger Beitræege zur Germanistik ; Bd. 18
considered descriptor values:       |3| {<100>-andrzej#feret[0], <540a>-3-631-
53513-9[0], <902s1>-partizip[4]}
matched descriptor values:          |1| {partizip}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv):   |4| {435, 437.02, 440, 465.8}
calculated1 ddc classes (sections): |4| {435, 437, 440, 465}
calculated1 ddc classes (main):     |1| {400}
calculated2 ddc classes (subdiv):   |2| {437.02[1], 465.8[1]}
calculated2 ddc classes (sections): {435[1], 437[1], 440[1], 465[1]}
calculated2 ddc classes (divisions):{430[2]}
calculated2 ddc classes (main):     {400[4]}
correlation (dnb_H0708_DNB0973618329,490): 10x.xxx xxx xxx xxx (0.5)

```

Abb. 27: Klassifizierungsergebnis von *in_dnb_DNB0973618329* nur mit dem Deskriptorwert „partizip“

DNB-K056 {DNB0983979332} "Semantik" nicht matched descriptor value!

Der Deskriptorwert „semantik“ des Deskriptors „Hauptsachtitel“ kommt in 274 unterschiedlichen DDC-Klassen vor („<331>-semantik[274]“), trägt jedoch wegen des großen Häufigkeitswertes nicht zum Klassifizierungsergebnis bei. Berücksichtigt werden wegen des *cutoff_val_dyn=32* Deskriptorwerte mit Häufigkeitswerten kleiner gleich 32: „verstehens“, „metaphern“, „innovativer“ und „emergente“ (Abb. 28).

```

number of ddc-classified title:      14218
identifier (dno,schedno):           DNB0983979332 (401.43,401.43)
DNB DDC notation (MAB2 field 700)   {400}
calculated cutoff value:            32
title:                              Emergenz als Phaenomen der Semantik am Beispiel
des Verstehens innovativer Metaphern
title (remainder):                  emergente konzeptuelle Merkmale an der
Schnittstelle von Semantik und Pragmatik
considered descriptor values:       |13| {<100>-helge#skirl[0], <331>-
semantik[274], <331>-beispiel[991], <331>-verstehens[24], <331>-innovativer[9],
<331>-metaphern[25], <331>-emergenz[7], <331>-phaenomen[119], <335>-
schnittstelle[62], <335>-pragmatik[99], <335>-emergente[32], <335>-
konzeptuelle[32], <335>-merkmale[36]}
matched descriptor values:          |4| {verstehens, metaphern, innovativer,
emergente}
max. match value of matched descriptor values: |4|
calculated1 ddc classes (subdiv):   |1| {370}
calculated1 ddc classes (sections): |1| {370}
calculated1 ddc classes (main):     |1| {300}
calculated2 ddc classes (subdiv):   {}
calculated2 ddc classes (sections): {370[1]}
calculated2 ddc classes (divisions):{370[1]}
calculated2 ddc classes (main):     {300[1]}
correlation (DNB983979332,401.43): 000.00x xxx xxx xxx (0)

```

Abb. 28: Klassifizierungsergebnis von *in_dnb_DNB0983979332*



Der Titeldatensatz DNB983979332 enthält keine Deskriptorwerte des Deskriptors „Schlagwort“. Wäre der Deskriptorwert „kognitive semantik“ vergeben worden, lieferte *vc_dcl* den Wert *CN=1* (Abb. 29).

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0983979332 (401.43,401.43)
DNB DDC notation (MAB2 field 700):  {400}
DDC notation (MAB2 field 705):      {401.43}
calculated cutoff value:            39
title:                               Emergenz als Phaenomen der Semantik am Beispiel
des Verstehens innovativer Metaphern
title (remainder):                  emergente konzeptuelle Merkmale an der
Schnittstelle von Semantik und Pragmatik
considered descriptor values:        |13| {<100>-helge#skirl[0], <331>-
semantik[370], <331>-beispiel[1637], <331>-verstehens[24], <331>-metaphern[27],
<331>-emergenz[7], <331>-phaenomen[135], <335>-schnittstelle[66], <335>-
pragmatik[122], <335>-emergente[32], <335>-konzeptuelle[39], <335>-merkmale[40],
<902s>-kognitive semantik[10]}
matched descriptor values:           |6| {konzeptuelle, verstehens, kognitive
semantik, metaphern, emergenz, emergente}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):    |5| {150, 370, 401, 401.43, 430}
calculated1 ddc classes (sections):  |2| {401, 430}
calculated1 ddc classes (main):      |1| {400}
calculated2 ddc classes (subdiv):    |1| {401.43[1]}
calculated2 ddc classes (sections):  {401[2]}
calculated2 ddc classes (divisions): {400[2]}
calculated2 ddc classes (main):      {400[3]}
correlation (dnb_H0708_DNB0983979332,401.43): 111.11x xxx xxx (1)

```

Abb. 29: Klassifizierungsergebnis von *in_dnb_DNB0983979332* mit Deskriptorwert „kognitive semantik“

DNB-K057 {DNB0974225894} DNB-HSG hier falsch

Für den Titeldatensatz DNB974225894 mit dem Titel „Was der Volksmund in einem Sprichwort verpackt ... moderne Aspekte des Sprichwortgebrauchs anhand von Beispielen aus dem Internet. Sprichwoerterforschung ; Bd. 24“ wurden *dno_like_i=430* („Deutsch“) und *dno_like_i=390* („Ethnologie“) vergeben. Ist die in *class2 (subdiv)* ermittelte *dno_a= 398.9* („Sprichwörter“) vertretbar (**F38**)? In *vc_dcl / vc_IDB* (Dezember 2008) wird in *calc2 (subdiv)* *dno_a=398.931* („Sprichwörter--Deutsch“) ermittelt. Wegen *dno_like_i*'s ist *CN=0*.

DNB-K058 {DNB0983977127} NSG 370 fehlt in DNB-Daten

Der Titeldatensatz DNB983977127 mit dem Titel „Kreatives Schreiben im vielsprachigen Deutschunterricht“ wurde mit *dno_like_i=430* („Deutsch“) und mit *dno_i=430.712* („Germanische Sprachen Deutsch--Sekundarbildung“) klassifiziert. *vc_dcl* ermittelt *dno_a=808.0431*, da die drei Deskriptorwerte „kreatives“, „deutschunterricht“ und „schreiben“ des Deskriptors „Hauptsachtitel“ alle gemeinsam in der DDC-Klasse „808.0431“ („Deutsch--Rhetorik“) vorkommen. Zusätzlich



vergebene Deskriptorwerte „kreatives schreiben“, „schriftlicher ausdrück“ liefern CN=1 (Abb. 30):

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB0983977127 (430.712,430)
DNB DDC notation (MAB2 field 700): {430}
DDC notation (MAB2 field 705): {430.712}
calculated cutoff value: 150
title: Kreatives Schreiben im vielsprachigen
Deutschunterricht
considered descriptor values: |7| {<100>-ursula#brock[0], <331>-
vielsprachigen[1], <331>-deutschunterricht[150], <331>-kreatives[231], <331>-
schreiben[345], <412@410>-<033A>-finalsatz@nuernberg[0], <540a>-3-935930-50-X[0],
<902s1>-kreatives schreiben[80], <902s2>-schriftlicher ausdrück[17]}
matched descriptor values: |3| {deutschunterricht, schriftlicher ausdrück,
kreatives schreiben}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv): |4| {302.2244, 370, 400, 430}
calculated1 ddc classes (sections): |4| {302, 370, 400, 430}
calculated1 ddc classes (main): |2| {300, 400}
calculated2 ddc classes (subdiv): |1| {302.2244[1]}
calculated2 ddc classes (sections): {302[1], 370[1], 400[1], 430[1]}
calculated2 ddc classes (divisions): {300[1], 370[1], 400[1], 430[1]}
calculated2 ddc classes (main): {300[2], 400[2]}
correlation (dnb_H0709_DNB0983977127,430): 11x.xxx xxx xxx xxx (1)

```

Abb. 30: Klassifizierungsergebnis von *in_dnb_DNB0983977127* mit Deskriptorwerten „kreatives schreiben“ und „schriftlicher ausdrück“

DNB-K059 {DNB0985891726} Springer-Online aus Dordrecht

Für den Titeldatensatz DNB985891726 mit dem Titel „Argumentative indicators in discourse – a pragma dialectical study“ wurde *dno_like_i=400* („Sprache, Linguistik“) vergeben. *vc_dcl* ermittelt in *calc2 (sections) dno_a=168* („Argument und Überzeugung“) und *dno_a=501* („Philosophie und Theorie“), weshalb CN=0 ist. Sind *dno_a*'s vertretbar (**F39**)? In Version *vc_dcl / vc_IDB* (Dezember 2008) wird nur *dno_a=168* ermittelt.

DNB-K060 {DNB0982966741} Language ist indexiert; unklar, warum keine Notation aus HSG 400

Was bedeutet: „Language ist indexiert“ (**F40**)? Der Titeldatensatz DNB982966741 mit dem Titel „Wulguru – a salvage study of a north-eastern Australian language from Townsville“ enthält 14 Deskriptorwerte der Deskriptoren „I. Verfasser“, „Hauptsachtitel“, „Zusätze zum Hauptsachtitel“, „Verleger@Ort“, „Gesamttitle“, „ISBN“ und „geographisch-ethnographisches Schlagwort“: {<100>-mark#donohue[4], <331>-wulguru[0], <335>-north-eastern[64], <335>-australian[2775], <335>-language[3940], <335>-townsville[40], <335>-salvage[261], <335>-study[9307], <412@410>-<033A>-lincom europa@muenchen[114], <451>-world[11377], <451>-materials[3448], <451>-languages[1223], <540a>-3-89586-327-0[0], <902g>-wulgurukaba[0]}. DNB982966741 wurde mit *dno_like_i=490* („Andere Sprachen“) und *dno_i=499.15* („Australische Aborigines-Sprachen, ...“) intellektuell klassifiziert. *vc_dcl* (April 2008) ermittelt in *calc2 (sections) dno_a=301* („Soziologie und



Anthropologie“). *vc_dcl* (Oktober 2008) ermittelt mit *vc_IDB* (November 2008, alle *dno*'s, nicht nur *schedno*'s, s. a. Kapitel 4.1, Programmänderung VC3.) *CN=1*:

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0982966741 (499.15,499.15)
DNB DDC notation (MAB2 field 700):  {490}
DDC notation (MAB2 field 705):      {499.15}
calculated cutoff value:            31539
title:                               Wulguru
title (remainder):                  a salvage study of a north-eastern Australian
language from Townsville
title (series):                      Languages of the world : Materials ; 463
considered descriptor values:        |14| {<100>-mark#donohue[4], <331>-wulguru[0],
<335>-north-eastern[80], <335>-australian[5011], <335>-language[9162], <335>-
townsville[45], <335>-salvage[316], <335>-study[26634], <412@410>-<033A>-lincom
europa@muenchen[126], <451>-world[31539], <451>-materials[6878], <451>-
languages[2893], <540a>-3-89586-327-0[0], <902g>-wulgurukaba[0]}
matched descriptor values:           |10| {world, australian, study, language,
north-eastern, languages, salvage, townsville, materials, <033A>-lincom
europa@muenchen}
max. match value of matched descriptor values: |7|
calculated1 ddc classes (subdiv):    |13| {080, 330, 400, 410, 415, 499.15, 621.395,
649.1, 658.562, 811.54, 820.9, 823.914, 978}
calculated1 ddc classes (sections):  |4| {400, 410, 415, 499}
calculated1 ddc classes (main):      |1| {400}
calculated2 ddc classes (subdiv):    |1| {499.15[1]}
calculated2 ddc classes (sections):  {400[1], 410[1], 415[1], 499[1]}
calculated2 ddc classes (divisions): {410[2]}
calculated2 ddc classes (main):      {400[4]}
correlation (dnb_A0748_DNB0982966741,499.15): 111.11x xxx xxx (1)

```

Abb. 31: Klassifizierungsergebnis von *in_dnb_DNB0982966741* mit *vc_dcl* (Oktober) und *vc_IDB* (ohne Begrenzung auf Haupttafelnotationen)

DNB-K061 {DNB0983718563} Hier 5420 herangezogen!! Colibri zieht offenbar jeweils die letzte DNB-Notation zum Vergleich heran!

Was bedeutet: „Hier 5420 herangezogen!!“ (**F41**)? Zur „letzten DNB-Notation“ s. Fußnote 26. Der Titeldatensatz DNB983718563 wurde mit *dno_like_i=430*, *dno_like_i=150*, *dno_like_i=440*, *dno_j=430.19*, *dno_j=445.62* und *dno_j=445.63* intellektuell klassifiziert. In *vc_dcl* (April 2008) ist *CN=0.33* wegen Vergleich mit „letzter DNB-Notation“. In *vc_dcl* (Oktober) ist *CN=1*, allerdings (sind die Deskriptorwerte des /) ist der Titeldatensatz(es) als *ppn_0527092223* mit *dno=430* in *vc_IDB* enthalten; daher keine besondere *vc_dcl*-Leistung.

DNB-Excel-Datei-Kommentare zur Hauptklasse 500

DNB-K062 {DNB0984333096, DNB0984745033, DNB0985475307, DNB0985606304, DNB0985705701} more than 15



Vgl. DNB-K007. Bei [DNB984333096](#) ist nur der Deskriptorwert „ansaeetze“ (in 542 DDC-Klassen) verfügbar. Wegen „{<100>-cornelia#kasper[0], <331>-zellkulturtechnik[0], <331>-ansaeetze[542]}“ ist *CN*=0. Die Hinzufügung des einschlägigen Deskriptorwertes „zellkultur“ des Deskriptors „Sachschlagwort“ liefert *CN*=1 (Abb. 32).

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB0984333096 (571.638,571.638)
DNB DDC notation (MAB2 field 700): {570}
DDC notation (MAB2 field 705): {571.638}
calculated cutoff value: 81
title: Neue Ansaetze in der Zellkulturtechnik
considered descriptor values: |4| {<100>-cornelia#kasper[0], <331>-
zellkulturtechnik[0], <331>-ansaeetze[656], <902s>-zellkultur[81]}
matched descriptor values: |1| {zellkultur}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv): |81| {338.16, 571, 571.32, 571.538, 571.53812,
571.5382, 571.6, 571.6011, 571.615774, 571.629, 571.63431, 571.638, 571.6381,
571.6381276, 571.63819, 571.6382, 571.638293, 571.835, 571.9648, 572.565, 573.8536,
574.0724, 574.192508, 574.8, 574.8028, 574.87, 574.87072, 574.872, 574.876,
574.8761, 576.028, 576.078, 576.15, 576.3, 578, 579.2, 581.0724, 581.8028, 581.81,
581.82028, 582.08, 591.0724, 591.870724, 596.0188, 596.033, 599.0188, 599.087,
599.0870724, 610.28, 611.0181, 611.0181028, 611.0187, 611.0187072, 611.0188,
611.01894, 612.028, 612.112, 612.80724, 615.19, 615.1901, 615.32, 615.36, 615.907,
616.0277, 616.02774, 616.07, 616.0756, 616.50072, 616.8, 616.9920072, 616.994,
616.99407, 628.161, 631.5233, 631.53, 632.3, 633.738, 660.6, 660.6078, 660.63, 664}
calculated1 ddc classes (sections): |12| {571, 572, 573, 574, 576, 578, 579, 581,
582, 591, 596, 599}
calculated1 ddc classes (main): |1| {500}
calculated2 ddc classes (subdiv): |43| {571.32[1], 571.53812[1], 571.5382[1],
571.538[1], 571.6011[1], 571.615774[1], 571.629[1], 571.63431[1], 571.63819[1],
571.6381[1], 571.638293[1], 571.6382[1], 571.638[1], 571.6[1], 571.835[1],
571.9648[1], 572.565[1], 573.8536[1], 574.0724[1], 574.192508[1], 574.8028[1],
574.87072[1], 574.872[1], 574.8761[1], 574.876[1], 574.87[1], 574.8[1], 576.028[1],
576.078[1], 576.15[1], 576.3[1], 579.2[1], 581.0724[1], 581.8028[1], 581.81[1],
581.82028[1], 582.08[1], 591.0724[1], 591.870724[1], 596.0188[1], 596.033[1],
599.0188[1], 599.087[1]}
calculated2 ddc classes (sections): {571[18]}
calculated2 ddc classes (divisions): {570[35]}
calculated2 ddc classes (main): {500[47]}
correlation (H0708_DNB0984333096,571.638): 111.111 xxx xxx xxx (1)

```

Abb. 32: Klassifizierungsergebnis von *in_dnb_DNB0984333096* mit Deskriptorwert „zellkultur“

Ähnlich verhält es sich bei [DNB984745033](#): der einzige Deskriptorwert, der für *vc_dcl* in Betracht kommt, ist „entwicklung“ (kommt in 5.077 DDC-Klassen vor). Wird der Deskriptorwert „analytische chemie“ des Deskriptors „Sachschlagwort“ hinzugefügt, wird *CN*=1 (Abb. 33) durch *vc_dcl* ermittelt.

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB0984745033 (543,543)
DNB DDC notation (MAB2 field 700): {540}
DDC notation (MAB2 field 705): {543}
calculated cutoff value: 107
title: Entwicklung eines IR-Kohlenmonoxidsensors
considered descriptor values: |4| {<100>-holger#geuking[0], <331>-
entwicklung[5077], <331>-ir-kohlenmonoxidsensors[0], <902s>-analytische

```



```

chemie[107]}
matched descriptor values:          |1| {analytische chemie}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv):  |107| {519.5, 519.5024541, 519.535, 535.842,
540, 540.9224, 541.226, 541.280212, 541.38, 542, 542.8, 542.85, 543, 543.0011354,
543.0014, 543.00141, 543.0015195, 543.00182, 543.0021, 543.00218, 543.0028,
543.002855369, 543.003, 543.00685, 543.0072, 543.00727, 543.0078, 543.009, 543.01,
543.0103, 543.015195, 543.02, 543.03, 543.0515, 543.0685, 543.07, 543.0711,
543.072, 543.08134, 543.085, 543.0852, 543.0858, 543.08583, 543.08583077,
543.08584, 543.086, 543.087, 543.0873, 543.0877, 543.08770285, 543.089, 543.0892,
543.0893, 543.0894, 543.0896, 543.1, 543.19, 543.2, 543.5, 543.52, 544.63, 545,
546.078, 546.226, 546.383, 546.3916, 546.41, 546.7126, 547, 547.1226, 547.30858,
547.308585, 547.59046, 547.7046, 547.78046, 549.125, 551.9, 572.36, 572.6536,
572.7, 572.85, 572.86, 574.19243046, 574.192456, 576.024541, 577.14, 610.28,
615.19, 615.1901, 615.952920287, 620.11299, 620.19292, 621.381, 628.161,
628.501543, 628.50287, 628.52, 631.4, 632.951, 660.2815, 660.283, 660.6, 660.63,
665.5, 668.55, 702.88, 751}
calculated1 ddc classes (sections): |16| {519, 535, 540, 541, 542, 543, 544, 545,
546, 547, 549, 551, 572, 574, 576, 577}
calculated1 ddc classes (main):     |1| {500}
calculated2 ddc classes (subdiv):  |74| {519.535[1], 519.5[1], 535.842[1],
540.9224[1], 541.226[1], 541.280212[1], 541.38[1], 542.85[1], 542.8[1],
543.00141[1], 543.0014[1], 543.00182[1], 543.00218[1], 543.0021[1], 543.0028[1],
543.003[1], 543.00685[1], 543.00727[1], 543.0072[1], 543.0078[1], 543.009[1],
543.0103[1], 543.015195[1], 543.01[1], 543.02[1], 543.03[1], 543.0515[1],
543.0685[1], 543.0711[1], 543.072[1], 543.07[1], 543.08134[1], 543.0852[1],
543.08583[1], 543.08584[1], 543.0858[1], 543.085[1], 543.086[1], 543.0873[1],
543.0877[1], 543.087[1], 543.0892[1], 543.0893[1], 543.0894[1], 543.0896[1],
543.089[1], 543.19[1], 543.1[1], 543.2[1], 543.52[1], 543.5[1], 544.63[1],
546.078[1], 546.226[1], 546.383[1], 546.3916[1], 546.41[1], 546.7126[1],
547.1226[1], 547.308585[1], 547.30858[1], 547.59046[1], 547.7046[1], 547.78046[1],
549.125[1], 551.9[1], 572.36[1], 572.6536[1], 572.7[1], 572.85[1], 572.86[1],
574.192456[1], 576.024541[1], 577.14[1]}
calculated2 ddc classes (sections): {543[48]}
calculated2 ddc classes (divisions): {540[72]}
calculated2 ddc classes (main):     {500[86]}
correlation (H0708_DNB0984745033,543): 111.xxx xxx xxx xxx (1)

```

Abb. 33: Klassifizierungsergebnis von *in_dnb_DNB0984745033* mit Deskriptorwert „analytische chemie“

Beim Titeldatensatz *DNB985475307* stehen in *vc_dcl / vc_IDB* (April 2008) die Deskriptorwerte „collider“, „constraints“, „masses“ (*cutoff_val_dyn=376*) zur Verfügung, in *vc_dcl / vc_IDB* (Oktober 2008) nur „collider“ und „masses“ (*cutoff_val_dyn=371*), was immerhin zu *CN=0.5* führt (Abb. 34).

```

number of ddc-classified title:      1
identifier (dno,schedno):            DNB0985475307 (539.721,539.721)
DNB DDC notation (MAB2 field 700):  {530}
DDC notation (MAB2 field 705):      {539.721}
calculated cutoff value:             371
title:                               Constraints on neutralino masses and mixings
from cosmology and collider physics
considered descriptor values:        |8| {<100>-ulrich#langenfeld[0], <331>-
masses[371], <331>-mixings[0], <331>-cosmology[591], <331>-collider[22], <331>-
physics[2816], <331>-constraints[484], <331>-neutralino[0]}
matched descriptor values:           |2| {collider, masses}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |1| {539}
calculated1 ddc classes (sections):  |1| {539}

```



```

calculated1 ddc classes (main):      |1| {500}
calculated2 ddc classes (subdiv):    |0| {}
calculated2 ddc classes (sections):  {539[1]}
calculated2 ddc classes (divisions): {530[1]}
calculated2 ddc classes (main):      {500[1]}
correlation (dnb_H0710_DNB0985475307,539.721): 111.XXX xxx xxx xxx (0.5)

```

Abb. 34: Klassifizierungsergebnis von *in_dnb_DNB0985475307* mit *vc_dcl* / *vc_IDB* (Oktober 2008)

Der Titeldatensatz [DNB985606304](#) mit den Deskriptorwerten „{notes, beijing, september, exercise, october, china, lecture}“ liefert mit einer *vc_dcl*-Programmerweiterung (Dezember 2008: Erweiterung um MAB2-Feld 370: „Weitere Sachtitel“) den CN-Wert=I (Abb. 35):

```

number of ddc-classified title:      1
identifier (dno,schedno):            DNB0985606304 (551.22,551.22)
DNB DDC notation (MAB2 field 700):   {550}
DDC notation (MAB2 field 705):       {551.22}
calculated cutoff value:             13702
title:                               Lecture and exercise notes
title (remainder):                  Beijing, China, 26 September to 30 October
title (other):                      Seismology, seismic hazard assessment and risk
mitigation
considered descriptor values:        |15| {<100b>-peter#bormann[0], <331a>-
notes[7237], <331a>-lecture[2579], <331a>-exercise[1336], <335>-september[9576],
<335>-october[8062], <335>-beijing[830], <335>-china[13702], <370a>-
assessment[6314], <370a>-risk[3026], <370a>-mitigation[270], <370a>-
seismology[188], <370a>-seismic[452], <370a>-hazard[467], <412@410>-<033A>-
geoforschungszentrum@potsdam[2]}
matched descriptor values:           |12| {mitigation, seismic, notes, seismology,
assessment, september, exercise, october, china, lecture, hazard, risk}
max. match value of matched descriptor values: |12|
calculated1 ddc classes (subdiv):    |2| {551.22, 624.1762}
calculated1 ddc classes (sections):  |2| {551, 624}
calculated1 ddc classes (main):      |2| {500, 600}
calculated2 ddc classes (subdiv):    |2| {551.22[1], 624.1762[1]}
calculated2 ddc classes (sections):  {551[1], 624[1]}
calculated2 ddc classes (divisions): {550[1], 620[1]}
calculated2 ddc classes (main):      {500[1], 600[1]}
correlation (B0746_DNB0985606304,551.22): 111.11x xxx xxx xxx (1)

```

Abb. 35: Klassifizierungsergebnis von *in_dnb_DNB0985606304* mit *vc_dcl*-Erweiterung (Dezember 2008) um MAB2-Feld 370: „Weitere Sachtitel“

Der Titeldatensatz [DNB985705701](#) mit dem Titel „On the universality of peptide computing“ ist intellektuell mit *dno_like_i=570* („Biowissenschaften, Biologie“) und *dno_i=572.650285* klassifiziert (Abb. 36).

```

572.650285 <dno_DNB985705701_to_analyze; length: 10>
5----- Science <hatzen>
57----- Life sciences; biology <hatzen>
572----- Biochemistry <hat>
572.6----- Proteins <hat>
572.65----- Components of proteins <hat>
572.65----- Components of proteins <nfa1->572.5-572.8>
---.--02-- Miscellany <hat:T1--02>
---.--028- Auxiliary techniques and procedures; apparatus, equipment, materials

```



```
<hat:T1--028>
---.--0285 Data processing Computer applications <hat:T1--0285>
```

Abb. 36: *vc_daygram* der *dno_i=572.650285*

Ist diese DDC-Notation korrekt (**F42**)? Laut [<http://portal.acm.org/citation.cfm?id=1342101>] handelt es sich um eine Turingmaschinen-Simulation: „We present a simulation of Turing machines by peptide---antibody interactions. ... We also give conditions on the peptide computing model so that it can be simulated by a Turing machine.“ *vc_dcl* ermittelt wegen der Deskriptorwerte „peptide“, „preprint“, „universality“ in *calc2* (*sections*) *dno_a=005* („Computerprogrammierung, Computerprogramme, Daten“) und *dno_a=016* („Bibliografien und Kataloge von Werken über einzelne Themen oder in einzelnen Fachgebieten“).

DNB-K063 {DNB0985765887, DNB0985592400} Nichtsammelgebiet

Nichtsammelgebiet, was heißt das für die Klassifizierung (**F43**)?

DNB985765887 mit dem Titel „The genesis of general relativity“ ist intellektuell mit *dno_like_j=530* („Physik“) klassifiziert. *vc_dcl* ermittelt in *calc2* (*subdiv*) *dno_a=530.092* („Physiker“), *dno_a=530.11* („Relativitätstheorie“) und *dno_a=530.1* („Theorien und mathematische Physik“). *CN=1*.

DNB985592400 mit dem Titel “Thomas Harriot's Artis analyticae praxis“ enthält *dno_like_j=510* („Mathematik“). *vc_dcl* ermittelt *dno_a=510*, also ist *CN=1*. Da 13 Deskriptorwerte „{english, commentary, mathematics, physical, history, harriot's, thomas#harriot, translation, series, analyticae, artis, sources, praxis}“ zur automatischen Klassifizierung beitragen, liegt der Verdacht nahe, dass diese (also der Titeldatensatz) in *vc_IDB* enthalten sind. Dies ist der Fall: mit Hilfe des Titels “Thomas Harriot's Artis analyticae praxis“ ist der Titeldatensatz unter der *ppn_0519519094* (Pica+-Kategorie <003@>) und *DNB981671926* (Pica+-Kategorie <007G>) ungleich *DNB985592400*! mit *dno=510* identifizierbar. Die ISBN in *ppn_0519519094* ist 00387495118 (Pica+-Kategorie <004A>) ist ungleich der ISBN 978-0-387-49512-5 (MAB2 Feld 540a) in *DNB985592400*!

DNB-K064 {DNB0969956975, DNB0971586594, DNB097184173X} Reihe a 2004

Was heißt „Reihe a 2004“ (**F44**)? Alle drei Titeldatensätze sind in der Datei *H0711ti2.dis* enthalten.

DNB-K065 {DNB0985652993} *CN=1* ?

Für den Titeldatensatz *DNB985652993* mit *dno_like_j=530* („Physik“) und *dno_i=530.11* („Relativitätstheorie“) wird *CN=1* von *vc_dcl* ermittelt, da *dno_a=530.11* in *calc2* (*subdiv*) enthalten. *CN=1* ist in diesem Fall nicht aussagekräftig, da *dno_a=530.11* eine von 283 (!) ermittelten DDC-Notationskandidaten ist. Die *vc_dcl*-Programmversion (Dezember 2008) enthält als einzigen Deskriptorwert „<412@410>-<033A>-nora@berlin[7]“, damit ist der *cutoff_val_dyn=7*. Durch Löschung des Deskriptorwertes „nora@berlin“ und Hinzufügung des Deskriptorwertes „einstein“ des



Deskriptors „Sachschlagwort“ wird weiterhin $CN=1$ erreicht, wobei $dno_a=530.11$ in *calc2 (subdiv)* nun (nur noch) einer von neun DDC-Notationskandidaten ist. In *calc2 (sections)* wird sogar nur ein $dno_a=530$ ermittelt (Abb. 37). In diesem Fall ist ein besseres Ergebnis also durch eine *QCUT*- und *QTIT*-Verbesserung erreichbar. Es ist zu prüfen, ob durch eine Erweiterung / Veränderung der heuristischen Regeln für solche Fälle – wenn nur der Deskriptorwert des Deskriptors „Verleger@Ort“ für *vc_dcl* verfügbar ist – möglich ist, ohne dass andere Klassifizierungsergebnisse verschlechtert werden.

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0985652993 (530.11,530.11)
DNB DDC notation (MAB2 field 700):  {530}
DDC notation (MAB2 field 705):      {530.11}
calculated cutoff value:            1671
title:                              aDiea Einsteinlegende
considered descriptor values:        |6| {<100>-uwe#koehler[0], <331>-
einsteinlegende[0], <412@410>-<033A>-nora@berlin[7], <540a>-3-86557-114-X[0],
<902s>-relativitaetstheorie[534], <902s11>-kritik[1671], <902s2>-einstein[504]}
matched descriptor values:           |3| {relativitaetstheorie, einstein, kritik}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):    |19| {115, 121, 128, 128.2, 155.7, 190, 193,
501, 510.1, 510.76, 530, 530.01, 530.092, 530.1, 530.11, 530.12, 551.46, 576.8,
650}
calculated1 ddc classes (sections):  |5| {501, 510, 530, 551, 576}
calculated1 ddc classes (main):      |1| {500}
calculated2 ddc classes (subdiv):    |9| {510.1[1], 510.76[1], 530.01[1],
530.092[1], 530.11[1], 530.12[1], 530.1[1], 551.46[1], 576.8[1]}
calculated2 ddc classes (sections):  {530[6]}
calculated2 ddc classes (divisions): {530[6]}
calculated2 ddc classes (main):      {500[11]}
correlation (dnb_A0748_DNB0985652993,530.11): 111.11x xxx xxx xxx (1)

```

Abb. 37: Klassifizierungsergebnis von *in_dnb_DNB0985652993* mit gelöschtem Deskriptorwert „nora@berlin“ und hinzugefügtem Deskriptorwert „einstein“

DNB-K066 {DNB0983772452} cn-wert 1?

Ja, $CN=1$, vgl. DNB-K007. Titeldatensatz enthält nur Deskriptorwerte der Deskriptoren „Hauptsachtitel“ und „Personennamen“. Nur der Deskriptorwert „speciation“ kommt für *vc_dcl* in Betracht.

DNB-K067 {DNB0984205594} cn-wert?

Vgl. DNB-K-007. Der Titeldatensatz DNB984205594 wurde mit $dno_like_j=540$ („Chemie“) und $dno_j=541.345$ („Kolloidchemie“) intellektuell klassifiziert. $dno_a=541.341$ ist als einer von 89 DDC-Notationskandidaten in *calc2 (subdiv)* enthalten, deshalb ist $CN=5/6=0.83$.



DNB-K068 {DNB0985415541} dno=2.DNB-DDCNotation

Vgl. Fußnote 26.

DNB-Excel-Datei-Kommentare zur Hauptklasse 600

DNB-K069 {DNB0974947962, DNB0975230638} Gute C-Notation keine DNB-Notation

Der Titeldatensatz DNB974947962 mit dem Titel „Trellis based joint data-channel estimation for differentially encoded SISO and MIMO systems“ ist mit *dno_like_i=620* („Ingenieurwissenschaften“) und *dno_like_i=004* („Informatik“) klassifiziert. Aufgrund der aussagekräftigen, in *vc_IDB* enthaltenen Deskriptorwerte des Deskriptors „Sachschlagwort“: „{blinde identifikation <informationstheorie>, intersymbolstoerung, trelliscodierte modulation}“ (*vc_dcl*, April 2008) bzw. „{intersymbolstoerung, uebertragungskanal, trelliscodierte modulation}“ und „mimo“ des Deskriptors „Hauptsachtitel“ (*vc_dcl*, Oktober 2008) wird in *calc2 (subdiv)* als *dno_a=621.3822* („Signal processing“) ermittelt. *CN=2/2=1*.

Der Titeldatensatz DNB975230638 enthält *dno_like_i=620* („Ingenieurwissenschaften“) und *dno_like_i=004* („Informatik“). Aufgrund sehr guter (=einschlägiger) Deskriptorwerte des Deskriptors „Sachschlagwort“: „<902s2>-greifmanipulator[2], <902s1>-taktiler sensor[2], <907s>-roboterarm[2]“ (geringe Häufigkeiten, *cutoff_val_dyn=2!*) werden *dno_a=629.892* („Roboter“) und *dno_a=629.8933* („Manipulatoren“) in *calc2 (subdiv)* ermittelt.

DNB-K070 {DNB0986049239} Sehr viele falsche Treffer bei *calculated2*.

Bei dem Titeldatensatz DNB986049239 mit *dno_like_i=610* ist die Lage verzwick. Die Anzahl von 12 DDC-Notationskandidaten („viele falsche Treffer“) in *calc2 (subdiv)*: „{610.28[1], 615.19[1], 616.0754[1], 620.118[1], 621.3678[1], 621.36[1], 621.381[1], 621.3[1], 621.4022[1], 628.5[1], 660.6[1], 681.2[1]}“ kommt aufgrund der grossen Anzahl (11) der Deskriptorwerte „{mathematics, physics, applications, medical, nato, series, science, chemistry, imaging, yves#lemoigne, <033A>-springer@dordrecht}“ zustande, von denen je 10 Deskriptorwerte zu den DDC-Klassen beitragen. *CN=2/2=1*, da die ersten beiden Stellen „61“ in *calc2 (subdiv)* auftreten. In Version *vc_dcl / vc_IDB* (Dezember 2008) stellt sich heraus, dass der Titeldatensatz (bzw. dessen Deskriptorwerte) in *vc_IDB* mit *ppn_0548385459* (Pica+-Kategorie <003@>) und *GBV0548385459* (Identifikationsnummer der ersterfassenden Institution, Pica+-Kategorie <007G>) und als *ppn_0519809939* (Pica+-Kategorie <003@>) und *DNB0981616879* (Pica+-Kategorie <007G>) mit *dno=616.0754* enthalten ist; jedoch ist *CN=0*. Ursache hierfür ist der dynamisch ermittelte *cutoff_val_dyn=19085* (eine der heuristischen Regeln von *vc_dcl* ermittelt diesen hohen Wert). Wird der Wert statisch im Bereich *1631 <= cutoff_val_stat <= 2816* gesetzt, wird in *calc2 (subdiv)* die DDC-Notation *dno_a=616.0754* ermittelt.



Es muss experimentiert werden, ob eine Verfeinerung der heuristischen Funktion (QCUT-Verbesserung) möglich ist, ohne dass sich die anderen Klassifizierungsergebnisse im Durchschnitt verschlechtern (vgl. DNB-K065).

DNB-K071 {DNB0983744475, DNB098518406X, DNB098597446X} Keine C-Notation; Keine Colibrinotation vorhanden

Kein einziger Deskriptorwert der Titeldatensätze DNB983744475, DNB98518406X und DNB98597446X ist in *vc_IDB* enthalten, deshalb kann in diesen Fällen kein DDC-Notationskandidat ermittelt werden.

DNB-K072 {DNB0986026271, DNB0983686424, DNB0985113197, DNB0985301546, DNB098472138X} keine C-Notation bei *calculated2 ddc classes (subdiv)*; Kein Wert in *calculated2 ddc classes (subdiv)*

Ja, bei diesen Titeldatensätzen werden – aus unterschiedlichen Gründen – von *vc_dcl* (April 2008) keine spezifischeren DDC-Notationskandidaten als aus den tausend Klassen (der dritten Ebene) ermittelt.

DNB986026271 mit dem Titel „Maxim Zetkin – 1883 - 1965 ; Arzt, Gesundheitspolitiker und Wissenschaftler“ ist mit *dno_like_i=610* („Medizin, Gesundheit“) und *dno_i=610.92* („Personen“) klassifiziert; *vc_dcl* (April 2008) ermittelt für *calc2 (subdiv)* = {} und für *calc2 (sections)* = {016[1], 346[1]}, daher ist *CN=0*. *vc_dcl / vc_IDB* (Dezember 2008) liefert *CN=1*, da *dno_a=610.92* eine von 39 DDC-Notationskandidaten in *calc2 (subdiv)* ist. In *calc2 (sections)* wird *dno_a=615* („Pharmakologie und Therapeutik“) und in *calc2 (divisions)* wird *dno_a=610* („Medizin und Gesundheit“) ermittelt.

Der Titeldatensatz DNB983686424 mit dem Titel „Evaluierung der Genauigkeit eines Norland-XR26-DXA-Systems im Vergleich zu einem GE-Lunar-DPX-IQ unter Verwendung eines modifizierten Variable-composition-Phantoms“ ist mit *dno_like_i=630* („Landwirtschaft, Veterinaermedizin“), *dno_like_i=610* („Medizin, Gesundheit“) und *dno_i=636.089671075* (*schedno=636.08967*) („Bewegungsapparat--Krankheiten--Tiere--Tiermedizin“) klassifiziert. Ist *dno_i* korrekt (**F45**)? Denn nach [<http://edoc.ub.uni-muenchen.de/6605/>] geht es um: „Inhalt dieser Arbeit sind die Fettgehaltsmessungen zweier Röntgenabsorptionsmessgeräte verschiedener Hersteller unter Verwendung eines „Variable Composition Phantoms“ (VCP) in der Kombination mit einem Wirbelsäulenphantom. ... Das Ziel der Arbeit war die Eröffnung der Möglichkeit des Vergleiches von Fettgehaltsmessungen zweier DXA-Geräte unterschiedlicher Hersteller. ...“ ist. Als Schlagwörter (keywords) wurden „Röntgenabsorptiometrie“, „Phantom“, „Fettgehaltsmessung“, „Norland“ und „Lunar“ und als DDC-Notation 600 („Technik und Technologie“) vergeben. *vc_dcl* ermittelt in *calc2 (sections)* *dno_a=650* („Management und unterstützende Tätigkeiten“), deshalb ist *CN=1/8=0.125*. Mit *dno_i=600* wäre *CN=1*.



Der Titeldatensatz [DNB985113197](#) mit dem Titel „Praevalenz und Risikofaktoren fuer eine transfusionsbedingte Infektion mit GB-Virus C bei erwachsenen Patienten mit Herz-OP im Kindesalter“ enthält nur Deskriptorwerte des Deskriptors „Hauptsachtitel“. Intellektuell klassifiziert wurde mit *dno_like_i=610* („Medizin, Gesundheit“) und *dno_i=616.3* („Krankheiten des Verdauungssystems“). Ist *dno_i* korrekt (**F46**)? In [[mediatum2.ub.tum.de/doc/622781/document.pdf](#)], S. 1 und 2 geht es um: „Ziel der hier vorgestellten Studie war eine Beschreibung der GBV-C Prävalenz und möglicher Risikofaktoren der Übertragung in einer polytransfundierte Risikopopulation im Vergleich mit einer Kontrollgruppe aus der Normalpopulation. Es wurden Blut- und Serumproben von 414 Patienten mit angeborenen Herzfehlern untersucht ... Dem Patientenkollektiv wurde eine Kontrollgruppe mit 487 alters- und geschlechtsgemachten Kontrollpersonen gegenübergestellt. ...“ Mit den Deskriptorwerten „{erwachsenen, praevalenz, kindesalter, risikofaktoren, infektion, patienten}“ ermittelt *vc_dcl* in *calc2 (sections) dno_a=610* („Medizin und Gesundheit“), deshalb ist $CN=2/4=0.5$.

Der Titeldatensatz [DNB985301546](#) mit dem Titel: „Belastung von Rettungshunden waehrend einer dreitaegigen Truemmersuche auf einem Katastrophenuebungsgelaende“ wurde mit *dno_like_i=630* („Landwirtschaft, Veterinaermedizin“), *dno_i=636.73089* (*schedno=636.73*: „Gebrauchs- und Hütehunde“) intellektuell klassifiziert (Abb. 38).

```
636.73089 <dno_DNB985301546_to_analyze; length: 9>
6----- Technology <hatzen>
63----- Agriculture <hatzen>
636----- Animal husbandry <hat>
636.7---- Dogs <hat>
636.73--- Working and herding dogs <hat>
---.--08- History and description with respect to kinds of persons <T1--08>
---.--089 Ethnic and national groups <T1--089>
```

Abb. 38: *vc_daygram* der *dno_i=636.73089*

Wieso wird an 636.73 die DDC-Hilfstafel-Notation T1--089 („Ethnische und nationale Gruppen“) angefügt (**F47**)? Die mit *vc_dcl* ermittelten *dno_a*'s beginnen alle mit „3“, daher $CN=0$. Das unbefriedigende Ergebnis liegt daran, dass nur Deskriptorwerte des Deskriptors „Hauptsachtitel“ verfügbar sind und diese z. T. nicht in *vc_IDB* enthalten sind: „|7| {<100b>-stephanie#wilhelm[0], <331>-waehrend[273], <331>-dreitaegigen[1], <331>-truemmersuche[0], <331>-katastrophenuebungsgelaende[0], <331>-belastung[73], <331>-rettungshunden[0]}“. Von diesen Deskriptorwerten sind nur „{belastung, waehrend}“ gemeinsam in den mit „3“ beginnenden DDC-Klassen enthalten. Wird der Deskriptorwert „waehrend“ eliminiert und die einschlägigen Deskriptorwerte „search dogs“ und „rescue dogs“ des Deskriptors „Sachschlagwort“ hinzugefügt, ist *schedno=636.73* in *calc2 (subdiv)* enthalten, dann ist $CN=1$ (Abb. 39):

```
number of ddc-classified title: 1
identifier (dno,schedno): DNB0985301546 (636.73089,636.73)
DNB DDC notation (MAB2 field 700): {630}
DDC notation (MAB2 field 705): {636.73089}
calculated cutoff value: 81
```



```

title: Belastung von Rettungshunden waehrend einer
dreitaegigen Truemmersuche auf einem Katastrophenuebungsgelaende
considered descriptor values: |8| {<100b>-stephanie#wilhelm[0], <331>-
dreitaegigen[1], <331>-truemmersuche[0], <331>-katastrophenuebungsgelaende[0],
<331>-belastung[81], <331>-rettungshunden[0], <902s>-rescue dogs[12], <902s>-search
dogs[8]}
matched descriptor values: |2| {search dogs, rescue dogs}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |6| {636, 636.70886, 636.73, 636.7536, 813.54,
813.6}
calculated1 ddc classes (sections): |1| {636}
calculated1 ddc classes (main): |1| {600}
calculated2 ddc classes (subdiv): |3| {636.70886[1], 636.73[1], 636.7536[1]}
calculated2 ddc classes (sections): {636[4]}
calculated2 ddc classes (divisions): {630[4]}
calculated2 ddc classes (main): {600[4]}
correlation (dnb_H0711_DNB0985301546,636.73): 111.11x xxx xxx (1)

```

Abb. 39: Klassifizierungsergebnis von *in_dnb_DNB0985301546* mit verbessertem QTIT und QELIM

Der Titeldatensatz DNB98472138X mit dem Titel „Anforderungen an den hydrothermischen Kraftwerksverbund unter dem Einfluss regenerativer Energien am Beispiel der Windkraft“ wurde mit *dno_like_i=620* („Ingenieurwissenschaften“) und *dno_i= 621.042* („Energietechnik“) klassifiziert. *vc_dc* ermittelt nicht spezifischer als *dno_a=620* („Ingenieurwissenschaften und zugeordnete Tätigkeiten“), daher ist CN=0.33. Werden die beiden Deskriptorwerte „wind energy“ und „wind power“ zum Titeldatensatz hinzugefügt, kann CN=0.5 erreicht werden (Abb. 40), eine weitere Verbesserung ist vermutlich nur durch weitere, ähnliche Titeldatensätze erreichbar (Q/DB-Verbesserung).

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB098472138X (621.042,621.042)
DNB DDC notation (MAB2 field 700): {620}
DDC notation (MAB2 field 705): {621.042}
calculated cutoff value: 79
title: Anforderungen an den hydrothermischen
Kraftwerksverbund unter dem Einfluss regenerativer Energien am Beispiel der
Windkraft
considered descriptor values: |14| {<100>-rene#kuehne[0], <331>-
hydrothermischen[0], <331>-kraftwerksverbund[0], <331>-unter[3036], <331>-
einfluss[772], <331>-regenerativer[4], <331>-energien[18], <331>-beispiel[1637],
<331>-windkraft[12], <331>-anforderungen[82], <902s1>-energietechnik[48], <902s2>-
thermodynamik[201], <902s2>-wind energy[8], <902s2>-wind power[79]}
matched descriptor values: |6| {wind energy, energien, regenerativer, wind
power, energietechnik, windkraft}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv): |6| {333.79, 333.792, 333.794, 620, 621.312136,
621.45}
calculated1 ddc classes (sections): |3| {333, 620, 621}
calculated1 ddc classes (main): |2| {300, 600}
calculated2 ddc classes (subdiv): |5| {333.792[1], 333.794[1], 333.79[1],
621.312136[1], 621.45[1]}
calculated2 ddc classes (sections): {333[3]}
calculated2 ddc classes (divisions): {330[3], 620[3]}
calculated2 ddc classes (main): {300[3], 600[3]}
correlation (dnb_H0708_DNB098472138X,621.042): 111.000 xxx xxx xxx (0.5)

```

Abb. 40: Klassifizierungsergebnis von *in_dnb_DNB098472138X* mit verbessertem QTIT



DNB-K073 {DNB0985733403, DNB0985733918, DNB0985735074} Af-Satz

Was bedeutet „Af-Satz“, was für die automatische Klassifizierung (F48)?

Der Titeldatensatz DNB985733403 enthält nur *dno_like_i=650*. *vc_dcl* (April 2008) ermittelt (nur) mit „3“ beginnende DDC-Notationskandidaten. Bei Hinzufügung des Deskriptorwertes „marketing management“ wird *CN=I* erreicht (Abb. 41). Dasselbe gilt für den Titeldatensatz DNB985733918.

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0985733403 (650,650)
DNB DDC notation (MAB2 field 700):  {650}
calculated cutoff value:             340
title:                               Marketing channels and value networks
title (remainder):                  for the textbook "Marketing management" by
Philip Kotler and Kevin Lane Keller
considered descriptor values:        |13| {<331>-value[2644], <331>-networks[1666],
<331>-marketing[5704], <331>-channels[340], <335>-management[21526], <335>-
philip[971], <335>-kotler[3], <335>-kevin[184], <335>-lane[849], <335>-keller[256],
<335>-textbook[3149], <540a>-978-3-89828-919-1[0], <902s>-marketing management[7]}
matched descriptor values:           |3| {channels, marketing management, kotler}
max. match value of matched descriptor values: |4|
calculated1 ddc classes (subdiv):    |1| {658.8}
calculated1 ddc classes (sections):  |1| {658}
calculated1 ddc classes (main):      |1| {600}
calculated2 ddc classes (subdiv):    |1| {658.8[1]}
calculated2 ddc classes (sections):  {658[1]}
calculated2 ddc classes (divisions): {650[1]}
calculated2 ddc classes (main):      {600[1]}
correlation (dnb_B0748_DNB0985733403,650): 1lx.xxx xxx xxx xxx (1)

```

Abb. 41: Klassifizierungsergebnis von *in_dnb_ DNB0985733403* mit verbessertem QTIT

Auch der Titeldatensatz DNB985735074 enthält nur *dno_like_i=650*. *vc_dcl* (April 2008) ermittelt (nur) mit „3“ beginnende DDC-Notationskandidaten. Mit *vc_dcl / vc_IDB* (Dezember 2008) wird ohne Titeldatensatzanreicherung *dno_a=658.8* in *calc2 (subdiv)* vorgeschlagen (Abb. 42).

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0985735074 (650,650)
DNB DDC notation (MAB2 field 700):  {650}
calculated cutoff value:             56599
title:                               Planning of marketing strategies. - 1.
Differentiation, positioning and product development
title (remainder):                  study guide for the textbook "Marketing
management" by Philip Kotler and Kevin Lane Keller
considered descriptor values:        |17| {<331>-strategies.[0], <331>-
differentiation[7], <331>-positioning[220], <331>-product[1374], <331>-
development[22566], <331>-planning[11589], <331>-marketing[5704], <335>-
textbook[3149], <335>-management[21526], <335>-philip[971], <335>-kotler[3], <335>-
kevin[184], <335>-lane[849], <335>-keller[256], <335>-study[26634], <335>-
guide[56599], <540a>-978-3-89828-911-5[0]}
matched descriptor values:           |11| {marketing, product, study, positioning,
management, textbook, planning, development, guide, kotler, philip}
max. match value of matched descriptor values: |12|

```



```

calculated1 ddc classes (subdiv): |1| {658.8}
calculated1 ddc classes (sections): |1| {658}
calculated1 ddc classes (main): |1| {600}
calculated2 ddc classes (subdiv): |1| {658.8[1]}
calculated2 ddc classes (sections): {658[1]}
calculated2 ddc classes (divisions): {650[1]}
calculated2 ddc classes (main): {600[1]}
correlation (dnb_B0748_DNB0985735074,650): 11x.xxx xxx xxx xxx (1)

```

Abb. 42: Klassifizierungsergebnis von *in_dnb_DNB985735074* mit *vc_dcl / vc_IDB* (Dezember 2008)

DNB-K074 {DNB0971851417, DNB0971901775} keine DNB Notation

Bei DNB971851417 mit *dno_like_i=660* ist *dno_a=660*, daher *CN=1*. Alle vier Deskriptorwerte „{polypropylen, wasserdampf, verfahrenstechnik, kunststoffolie}“ tragen gemeinsam zum Ergebnis bei. Dasselbe gilt für DNB971901775 mit *dno_like_i=620=dno_a*, *CN=1*. Die beiden (spezifischen) Deskriptorwerte des Deskriptors „Hauptsachtitel“ bzw. „Sachschlagwort“ „<331>-waermedaemmschichtsystemen[1]“ bzw. „<907sl>-haftvermittler[1]“ tragen zum Ergebnis bei (*cutoff_val_dyn=2!*).

DNB-K075 {DNB0976896443, DNB0984278680, DNB0983474826, DNB0985343184, DNB0984980059, DNB0985043725} Mehr als 15 C-DDC bei *calculated2*; mehr als 15 *calculated2 ddc classes (subdiv)*; *calculated2 ddc classes (subdiv)* more than 15 *calculated ddc classes*

Vgl. K007. Die Titeldatensätze enthalten nur Deskriptorwerte des Deskriptors „Hauptsachtitel“ und nicht in *vc_IDB* vorkommende Deskriptorwerte der Deskriptoren „Personenname“ und / oder *Verleger@Ort*. In *vc_IDB* sind bei den Titeldatensätzen (nur) folgende Deskriptorwerte des Deskriptors „Hauptsachtitel“ vorhanden (*matched descriptor values*): DNB976896443: „medical“ und „card“ (*CN=0*); DNB984278680: „pharmakotherapie“ (*CN=1*); DNB983474826: „funktionelle“ (*CN=1*); DNB985343184: „natuerliche“ und „mensch-maschine-interaktion“ (*CN=0*); DNB984980059: „verwandte“ und „substanzen“ (*CN=0*); DNB985043725: „endometrium“, „peripartale“, „rindes“ und „toll-like-rezeptoren“ (*CN=1*).

DNB-Excel-Datei-Kommentare zur Hauptklasse 700

DNB-K076 {DNB0980488028, DNB0983878773, DNB0983933707, DNB0984491740, DNB0985652969, DNB0979748666, DNB0986008214, DNB0982708726, DNB0984216685, DNB0985824115, DNB0985270101, DNB0985478349, DNB0982695446, DNB0984363556, DNB0985366672, ZDB2170173-8, DNB0983825661, DNB0985516836, DNB0985960051, DNB0986065641, DNB0983677662, DNB098444940X, DNB0985660910, DNB0985790385, DNB0984524991, DNB0984535101, DNB0974539694, DNB0981508103, DNB0982129653, DNB0984405097, DNB0985136375, DNB0971757801, DNB0985254130, DNB0985610034} *calculated2 ddc classes (subdiv)* {}



Vgl. Fußnote 17. [DNB980488028](#) ist mit der nicht in einer Haupttafel enthaltenen *dno_j*=779.4421092 klassifiziert. Die erste DDC-Haupttafelnotation (*schedno*), wenn *dno_j* von rechts nach links gekürzt wird, ist „779“. Da *schedno*=779 zu *class2* (*sections*) gehört, muß *class2* (*subdiv*) leer sein. Analoges gilt für die Titeldatensätze: [DNB983878773](#) (*dno_j*=709.43155090511, *schedno*=709); [DNB983933707](#) (*dno_j*=795, *schedno*=795), [DNB984491740](#) (*dno_j*=779.28092, *schedno*=779); [DNB986008214](#) (*dno_j*=728.0222, *schedno*=728); [DNB986065641](#) (*dno_j*=708.34164, *schedno*=708); [DNB985660910](#) (*dno_j*=750.74, *schedno*=750); [DNB971757801](#) (*dno_j*=780, *schedno*=780) und [DNB985610034](#) (*dno_j*=716, *schedno*=716); [DNB985790385](#) (700,700). Bei den anderen DNB-K076-Titeldatensätzen wird aus unterschiedlichen Gründen keine spezifischere als (höchstens) zu *class2* (*sections*) gehörende DDC-Notation ermittelt. (Kurz)-Charakterisierung / -Analyse: [DNB985652969](#) (*dno_j*=796.071, *dno_a*=370, *CN*=0); [DNB979748666](#) (*dno_j*=720.473097471, als *ppn_0387007849* mit *dno_j*=917.4710443 in *vc_DB!*. *vc_dcl* (Dezember 2008) ermittelt *dno_a*=917.4710443⁶¹ in *calc1* (*subdiv*). *CN*=0, da *dno* von *ppn_0387007849* ungleich *dno_j* ist). [DNB982708726](#) (Verbesserung durch *QELIM* möglich, da nur die (nicht sinnvollen) Deskriptorwerte „28.“ und „16.“ zur automatischen Klassifizierung beitragen. Nach Löschung dieser (und zusätzlich „11.“ Und „19.“) wird von *vc_dcl* / *vc_IDB* (Dezember 2008) *dno_a*=738.2094 (*schedno*=738.209) in *calc2* (*subdiv*) ermittelt; daher *CN*=6/6=1); [DNB984216685](#) (*dno_j*=721.04203, *schedno*=721.042 („Gebäude nach Form, Gebäude mit Atrien“), *dno_a*=720 („Architektur“), daher *CN*=2/6=0.33. *CN* mit *vc_IDB* (Dezember 2008) nicht weiter verbesserbar, da [DNB984216685](#) als *ppn_0529969548* mit *dno*=720 in *vc_IDB* enthalten); [DNB985824115](#) (*QTIT* muß verbessert werden: „{<100b>-maarten#hart[1], <331>-mozart[642], <331>-ich[2136], <412@410>-<033A>-piper@muenchen[174], <451>-serie[382], <540a>-3-492-25058-0[0]}“, je einer der beiden Deskriptorwerte „maarten#hart“ und „<033A>-piper@muenchen“ tragen zu den mit „3“ („Sozialwissenschaften“) beginnenden DDC-Notationskandidaten bei); [DNB985270101](#) (*dno_j*=720.949454 (Abb. 43), *QTIT* muß verbessert werden, ein für *vc_dcl* verbleibender Deskriptorwert (s. u.) ist für die automatische Klassifizierung zuwenig; hier die in Betracht kommenden Deskriptorwerte: „{<100b>-esther#wetli[0], <331a>-daerstetten[1], <331a>-bauinventar[0], <331a>-gemeinde[298], <335>-perimeter[25], <335>-teibereich[0], <335>-gemaess[32], <412@410>-<033A>-denkmalpflege kantons bern@bern[0]}“. Nach Löschung von „gemaess“ (*QELIM*-Verbesserung) werden von *vc_dcl* / *vc_IDB* (Dezember 2008) „nur“ mit dem Deskriptorwert „perimeter“ (der beiden Deskriptorwerte („perimeter“ und „daerstetten“) in *calc2* (*subdiv*) mit „5“ beginnende DDC-Notationskandidaten ermittelt: „{511.8[1], 515.2433[1], 516.13[1], 516.15[1], 557.1241[1], 557.124[1]}“, „daerstetten“ kommt in *vc_IDB* nur in der DDC-Klasse 550 („Geowissenschaften“) vor.);

```
720.949454 <dno_DNB985270101_to_analyze; length: 10>
7----- Arts & recreation <hatzen>
72----- Architecture <hatzen>
720----- Architecture <hat>
720.9----- Historical, geographic, persons treatment <hat>
--0.9----- Historical, geographic, persons treatment <T1--09>
```

⁶¹ Bei Gelegenheit muß geklärt werden, wieso *dno_a*=917.4710443 nicht in *calc2* (*subdiv*) enthalten ist. Vgl. DNB-K086



```

--0.94---- Treatment by specific continents, countries, localities;
             extraterrestrial worlds <T1--094>
---.-4---- Europe      Western Europe <ba4r2span:T1--093-T1--099:T2--4>
---.-49--- Other parts of Europe <ba4r2span:T1--093-T1--099:T2--49>
---.-494-- Switzerland <ba4r2span:T1--093-T1--099:T2--494>
---.-4945- Mittelland cantons <ba4r2span:T1--093-T1--099:T2--4945>
---.-49454 Bern <ba4r2span:T1--093-T1--099:T2--49454>

```

Abb. 43: *vc_daygram* der *dno_i*=720.949454

DNB985478349 (Bei *QTIT*-Verbesserung durch Hinzufügung des Deskriptorwertes „lotteries“ wird *CN*=1 erzielt); DNB982695446 (*dno_i*=745.59412074435613 (Abb. 44), *vc_dcl* (Dezember 2008) bestimmt *dno_a*=390 („Bräuche, Etikette, Folklore“), deshalb *CN*=0. Für diesen Titeldatensatz muß entweder *QIDB* oder *QTIT* verbessert werden, in *vc_IDB* ist kein Titeldatensatz zu *dno_i* vorhanden, es gibt lediglich einige englische Deskriptorwerte von Titeldatensätzen zu *dno_i*=745.59412. Mit einer *QTIT*-Verbesserung durch die Deskriptorwerte „handicrafts“, „creche“ und „christmas decorations“ läßt sich *CN*=1 erreichen (Abb. 45));

```

745.59412074435613 <dnb_DNB982695446_to_analyze; length: 18>
7----- Arts & recreation <hatzen>
74----- Drawing & decorative arts <hatzen>
745----- Decorative arts <hat>
745.5----- Handicrafts <hat>
745.59----- Making specific objects <hat>
745.594----- Decorative objects <hat>
745.5941----- Objects for special occasions <hat>
745.59412----- Christmas <hat>
---.-07----- Education, research, related topics <T1--07>
---.-074----- Museums, collections, exhibits <T1--074>
---.-4----- Europe      Western Europe <na4r7:T2--4>
---.-43----- Central Europe      Germany <na4r7:T2--43>
---.-435----- Northwestern Germany <na4r7:T2--435>
---.-4356----- MÄ¼nster, Arnsberg, Detmold districts <na4r7:T2--4356>
---.-43561----- MÄ¼nster district (MÄ¼nster Regierungsbezirk) <na4r7:T2--43561>

```

Abb. 44: *vc_daygram* der *dno_i*=745.59412074435613 (die letzte „3“ stammt von *T2--435613* („Kreis Warendorf“) und ist bislang nicht in *WebDewey* bzw. der engl. *DDC-Wissensbasis vc_KB* enthalten.)

DNB984363556 (*dno_like_i*=750, *dno_i*=759.3609045, *schedno*=759.3, *dno_a*=750 und *CN*=2/4=0.5. Als *ppn_0531685055* (Pica+-Kategorie <003@>) bzw. *OBV0AC04882563* (Pica+-Kategorie <006L>) sind die Deskriptorwerte des Titeldatensatzes *DNB984363556* in *vc_IDB* mit *dno*=750 (Pica+-Kategorie <045F>) enthalten, daher wird *dno_a*=750 (und nicht spezifischer) von *vc_dcl* ermittelt); DNB985366672 (Titel: *Moderne in Leipzig – Architektur und Staedtebau 1918 bis 1933*, *dno_like_i*=720, *dno_i*=720.943212209042022, *schedno*=720.9, *vc_dcl* schlägt mit “9” („Geschichte, Geografie und Hilfswissenschaften“) beginnende *dno_a*’s vor. *QTIT* sollte verbessert werden, da keine Deskriptorwerte des Deskriptors „Schlagwort“ vorhanden; ev. *QCUT*-Verbesserung möglich); ZDB2170173-8 (Titel: „Gemaelde oesterreichischer Meister des 19. und 20. Jahrhunderts“, *dno_like_i*=750, *dno_i*=759.3609034, *schedno*=759.3. Durch Verbesserung von *QELIM*, hier Eliminierung der Deskriptorwerte „19.“ und „20.“ (vgl. *DNB982708726*) wird *CN*=1 von *vc_dcl*



ermittelt); [DNB983825661](#) (Titel: „Strassen Bruecken – Ingenieur-Bau-Kunst in Deutschland; Buch zur Ausstellung. Bridges - engineer construction art in Germany, *dno_i*=725.98 („Brücken, Tunnel, Gräben“). *CN*=0, da *dno_a*=943 („Mitteleuropa Deutschland“) mit *vc_dcl* (April 2008) und *dno_a*=620 („Ingenieurwissenschaften und zugeordnete Tätigkeiten“) mit *vc_dcl* / *vc_IDB* (Dezember 2008) vorgeschlagen wird, wobei *dno_a*=620 mit II (!) Deskriptorwerten: „{strassen, art, buch, bridges, germany, engineer, ausstellung, deutschland, construction, bruecken, strassenbruecken}“ bestimmt wird. Diese Deskriptorwerte haben wenig mit der „Architektur von öffentlichen Bauwerken“ zu tun; dieser Aspekt müßte mit Schlagworten ausgedrückt werden (*QTIT*-Verbesserung)); [DNB985516836](#) (Titel: „Stefanie Hoellering – Retrospektive, Bilder 1974 – 2000 – anlaesslich der Ausstellungsreihe: Stefanie Hoellering - Retrospektive, Bilder 1974 – 2000 ; Rathausgalerie Muenchen ... Centro Culturale S. Chiara Casalmaggiore“, *dno_i*=759.309045, *schedno*=759.3 („Deutsche Malerei, . . .“), *dno_a*=730 („Plastische Künste Bildhauerkunst“), *CN*=1/4=0.25);

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB0982695446 (745.59412074435613,745.59412)
DNB DDC notation (MAB2 field 700): {740}
DDC notation (MAB2 field 705): {745.59412074435613}
calculated cutoff value: 118
title: aDasa Weihnachtsgeheimnis
title (remainder): Katalog ; Katalog der 66. Krippenausstellung
vom 19. November 2006 bis zum 3. Januar 2007 im Museum Heimathaus Muensterland und
im Krippenmuseum, Telgte, mit dem besonderen Thema "Das Weihnachtsgeheimnis"
title (series): Sonderausstellung des Museums Heimathaus
Muensterland ; 220
considered descriptor values: |21| {<100b>-thomas#ostendorf[0], <331>-
weihnachtsgeheimnis[0], <335>-museum[6211], <335>-thema[372], <335>-heimathaus[0],
<335>-66.[0], <335>-muensterland[35], <335>-krippenausstellung[0], <335>-19.[16],
<335>-november[7680], <335>-krippenmuseum[0], <335>-telgte[10], <335>-januar[853],
<335>-katalog[1080], <412@410>-<033A>-museum heimathaus muensterland
krippenmuseum@telgte[0], <451>-sonderausstellung[138], <451>-museums[1812], <540a>-
3-927072-25-7[0], <902s>-handicrafts[118], <902s>-creche[20], <902s>-christmas
decorations[106]}
matched descriptor values: |3| {christmas decorations, creche,
handicrafts}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv): |4| {394.2663, 680, 745.5, 745.59412}
calculated1 ddc classes (sections): |1| {745}
calculated1 ddc classes (main): |1| {700}
calculated2 ddc classes (subdiv): |2| {745.59412[1], 745.5[1]}
calculated2 ddc classes (sections): {745[2]}
calculated2 ddc classes (divisions): {740[2]}
calculated2 ddc classes (main): {700[2]}
correlation (dnb_B0746_DNB0982695446,745.59412): 111.111 11x xxx xxx (1)

```

Abb. 45: Klassifizierungsergebnis von *in_dnb_DNB0982695446* mit *vc_dcl* (Dezember 2008) und verbessertem *QTIT* durch die Deskriptorwerte „handicrafts“, „creche“ und „christmas decorations“

[DNB985960051](#) (Titel: „I. K. H. Diane Herzogin von Wuerttemberg, Prinzessin von Frankreich - plastische Werke – anlaesslich der Ausstellungen vom 5. April bis 10. Juni 2007 bei der EnBW Energie



Baden-Wuerttemberg AG in Karlsruhe und vom 10. Juli bis 16. August 2007 bei der Sparkasse Pforzheim Calw in Pforzheim“, *dno_like_i=730*, *dno_i=730.92*. *vc_dcl* (April 2008) bestimmt DDC-Notationskandidaten, die mit „5“ beginnen. Mit *vc_dcl / vc_IDB* (Dezember 2008) wird erkannt, dass der Titeldatensatz (dessen Deskriptorwerte) DNB985960051 in *vc_IDB* als *ppn_0547494599* mit *dno=730* (Pica+-Kategorie <045F>) und *dno_like_i=730* (<045G>) enthalten sind. Trotz *QELIM*-Verbesserung (Eliminierung von „10.“ und „16.“) ist *CN=0*, was hier am dynamisch ermittelten *cutoff_val_dyn=25* liegt, dadurch werden mit „9“ beginnende DDC-Notationskandidaten bestimmt. In diesem Fall muß über eine *QCUT*-Verbesserung (24 Deskriptorwerte!) nachgedacht werden. Mit *cutoff_val_stat=138* wird *CN=3/5=0.6*. Dies ist der höchste, erreichbare *CN*-Wert, da die Deskriptorwerte des Titeldatensatzes DNB985960051 in *vc_IDB* „nur“ mit *dno_like_i=730* enthalten sind; DNB983677662 (*dno_i=708.315507443155*, *schedno=708*, mit *ppn_0527084689* in *vc_IDB* mit *dno=700* (Pica+-Kategorie <045F>) und *dno_like_i=700* (<045G>) enthalten, *CN=0.67*); DNB98444940X (Titel: „Indikatoren zur Siedlungsentwicklung – Flaechen gewinnen: Orientierungshilfen fuer Kommunen“, *dno_like_i=710* („Landschaftsgestaltung, Raumplanung“), *dno_i=711.58*, *dno_a=650*, *CN=0*. Es gibt keinen Deskriptorwert des Deskriptors „Schlagwort“. Hier muß *QTIT* und auch *QIDB* verbessert werden); DNB984524991 (Titel: „Papier als Werkstoff in der Bildenden Kunst – eine Bestandsaufnahme der Moderne und die gestalterischen Moeglichkeiten fuer den Kunstunterricht“, *dno_like_i=700* („Kuenste, Bildende Kunst allgemein“), *dno_like_i=370* („Erziehung, Schul- und Bildungswesen“), *dno_i=702.80712*, *schedno=702.8* („Techniken, Verfahren, Geräte, Ausstattung, Materialien“), alle fünf Deskriptorwerte „{kunstunterricht, bestandsaufnahme, papier, bildenden, werkstoff}“ tragen zu *dno_a=700* bei, daher *CN=2/4=0.5*. *dno_i=702.80712* („Techniken, Verfahren, Geräte, Ausstattung, Materialien--Sekundarbildung“, s. Abb. 46) ist in *vc_IDB* (Dezember 2008) nicht enthalten und steht daher nicht als DDC-Notationskandidat zur Verfügung. In *vc_IDB* ist nur die DDC-Klasse mit DDC-Notation 702.8 und ihre Oberklassen enthalten.);

```
702.80712 <dnb_DNB984524991_to_analyze; length: 9>
7----- Arts & recreation <hatzen>
70----- Arts <hatzen>
702----- Miscellany of fine and decorative arts <hat>
702.8---- Techniques, procedures, apparatus, equipment, materials <hat>
---.-07-- Education, research, related topics <T1--07>
---.-071- Education <T1--071>
---.-0712 Secondary education <T1--0712>
```

Abb. 46: *vc_daygram* der *dno_i=702.80712*

DNB984535101 (Titel: „Foerderung, Erfolg und Karrieren im Spitzenradsport – eine Ranglistenanalyse von 1980 bis 2004“, *dno_like_i=796* („Sport“), *dno_i=796.62* („Radrennsport“). Wegen der Deskriptorwerte „{foerderung, karrieren, erfolg}“, werden mit „3“ beginnende *dno_a*'s ermittelt. Hier sollten *QTIT* und *QIDB* verbessert werden.); DNB974539694 (Titel: „Grabdenkmaeler nordeuropaeischer Fuerstenhaeuser im Zeitalter der Renaissance ; 1550 – 1650 – Bau + Kunst ; Bd. 9“. *dno_like_i=720*, *dno_like_i=730*, *dno_like_i=943*, *dno_like_i=230*, *dno_i=726.8094309031*, *schedno=726.8*, alle vier Deskriptorwerte „{<033A>-ludwig@kiel, grabdenkmaeler, fuerstengrab,



grabmal}“ führen zu $dno_a=700$, somit $CN=1/4=0.25$); [DNB981508103](#) (Titel: „Helga Philipp – eine Monografie – Angewandte Kulturwissenschaften Wien; Bd. 6“, $dno_like_i=700$ („Kuenste, Bildende Kunst allgemein“), $dno_i=709.2$ („Personen“), $dno_a=709$, $CN=3/4=0.75$.); [DNB982129653](#) (Titel: ‚Die Kulturzeitschrift "magnum" – photographische Befunde der Moderne‘, $dno_like_i=770$ („Fotografie, Computerkunst“), $dno_like_i=070$ („Nachrichtenmedien, Journalismus, Verlagswesen“), $dno_like_i=300$ („Sozialwissenschaften, Soziologie“), $dno_i=778.538070$, $schedno=778.53807$ („Bildreportage“). *vc_dcl* (April 2008) ermittelt mit „9“ beginnende dno_a 's, weshalb $CN=0$. In *vc_dcl* / *vc_IDB*-Version (Dezember 2008) werden mit „{bildreportage, kulturzeitschrift, <033A>-jonas-verl.@marburg, kulturtheorie}“ in *calc2* (sections) die DDC-Notationskandidaten „{390[1], 700[1]}“ bestimmt, daher $CN=1/8=0.125$); [DNB984405097](#) (Titel: „Wie viel Abstraktion ertraegt die Lernwirksamkeit? – Diskussion der Vermittlung einer modellgeleiteten Ausbildungsbotschaft an Sporthochschulen in der Schweiz im Spannungsfeld zwischen Reduktion und Komplexitaet“, $dno_like_i=796$ („Sport“), $dno_like_i=370$ („Erziehung, Schul- und Bildungswesen“), $dno_i=796.0711$, $schedno=796.071$ („Ausbildung“). Trotz einiger Deskriptorwerte „{vermittlung, abstraktion, komplexitaet, viel, reduktion}“ ist $CN=0$. Auch nach Löschung des Deskriptorwertes „viel“ bleibt $CN=0$. Alle fünf Deskriptorwerte „{spannungsfeld, vermittlung, abstraktion, komplexitaet, reduktion}“ tragen zu DDC-Notationskandidaten bei, die mit „1“ („Philosophie, Parapsychologie und Okkultismus, Psychologie“) beginnen. Eine Verbesserung muß durch *QTIT* und *QIDB* erfolgen); [DNB985136375](#) („Titel: Paedagogische Professionalitaet und stellvertretende Problembearbeitung – ausgelegt durch Beispiele aus Schulsport und Sportstudium. Sport, Medien, Gesellschaft ; Bd. 4“, $dno_like_i=796$, $dno_like_i=370$, $dno_i=796.0711$, $schedno=796.071$. *vc_dcl* ermittelt in *calc2* (sections) mit den Deskriptorwerten „{<033A>-sportverl. strauss@koeln, lehramtsstudium, sportstudium, paedagogische faehigkeit, schulsport}“ $dno_a=796$, daher $CN=3/6=0.5$); [DNB985254130](#) (Titel: „Lehrplaene zwichen Offenheit und Verbindlichkeit – zur Standardisierung paedagogischer Qualitaet im Sportunterricht“, $dno_like_i=796$, $dno_like_i=370$, $dno_i=796.071$. Alle vier Deskriptorwerte „{sportunterricht, lehrplaene, verbindlichkeit, standardisierung}“ führen mit *vc_dcl* (April 2008) gemeinsam zu $dno_a=344$ („Arbeitsrecht, Sozialrecht, Bildungsrecht, Kulturrecht“), $CN=0$. Mit *vc_dcl* / *vc_IDB* (Dezember 2008) wird mit je zwei der fünf Deskriptorwerte „{sportunterricht, lehrplaene, verbindlichkeit, standardisierung, offenheit}“ in *calc2* (sections) „{128[1], 230[1], 370[1], 420[1], 796[1]}“ ermittelt, daher $CN=3/6=0.5$).

DNB-K077 {DNB0983714525, DNB0985622733, DNB0983153086, DNB0985824115} calculated2 ddc classes (subdiv) more than 15 calculated ddc classes

Vgl. K007.

[DNB-Excel-Datei-Kommentare zur Hauptklasse 800](#)

DNB-K078 {DNB0984907068} Bei Einzelschriftstellern gute Trefferrate



Der Titeldatensatz DNB984907068 mit dem Titel: „Quest and conquest in the fiction of David Lodge“ ist mit *dno_like_i*=820 („Englische Literatur“) und *dno_i*= 823.914 („Englische Erzählprosa--1945–1999, . . .“) intellektuell klassifiziert. Alle fünf Deskriptorwerte „{quest, lodge, fiction, conquest, david}“ tragen zu *dno_a*= 823.914 – eine von mehreren in *calc2 (subdiv)* von *vc_dcl* ermittelten DDC-Notationskandidaten – bei. In *calc2 (sections)* wird von *vc_dcl dno_a*=823 mit der Häufigkeit 5 ermittelt. *CP*=111.111 xxx xxx xxx und *CN*=1.

DNB-K079 {DNB098477825X} Kafka nicht indexiert

Der mit *dno_like_i*=830 („Deutsche Literatur) und *dno_i*=838.91203, *schedno*=838 („Deutsche vermischte Schriften“) klassifizierte Titeldatensatz DNB98477825X mit dem Titel „Hier muss ich mich festhalten ... – die Tagebuecher von Franz Kafka Elektronische Ressource :.ein literarisches Laboratorium 1909-1923“ enthält nur Deskriptorwerte der Deskriptoren „Name der I. sonstigen beteiligten Person“ und „Hauptsachtitel“: „{<100b>-andrea#rother[1], <331>-mich[379], <331>-festhalten[1], <331>-hier[394], <331>-muss[183], <335>-franz[831], <335>-kafka[377], <335>-elektronische[106], <335>-ressource[109], <335>-:ein[0], <335>-literarisches[61], <335>-laboratorium[118], <335>-1909-1923[2], <335>-tagebuecher[390]}“. Gemeinsam führen alle drei Deskriptorwerte „{elektronische, literarisches, ressource}“ in *calc2 (sections)* zu *dno_a*=016 („Bibliografien und Kataloge von Werken über einzelne Themen oder in einzelnen Fachgebieten“), daher *CN*=0. Wird der Deskriptorwert „kafka franz“ zum Titeldatensatz hinzugefügt (*QTIT*-Verbesserung) , wird von *vc_dcl dno_a*=833 („Deutsche Erzählprosa“) bestimmt und *CN*=0.6.

DNB-K080 {DNB0982104707, DNB098214296X, DNB0983874581, DNB0984195327, DNB098424834X, DNB0979438810, DNB0983429022, DNB0984858784, DNB0982551061, DNB0983252076, DNB0984014152, DNB0984123865, DNB0984250441, DNB0985992913, DNB0947922717, DNB0985763426, DNB0983543208, DNB098508720X, DNB0985186135, DNB0985275898, DNB0985276061, DNB0985425598, DNB0985717505, DNB0985717548, DNB0985794070, DNB098582283X, DNB098606467X, DNB0985100419, DNB0985475331, DNB098549235X, DNB0985653779, DNB0985870249, DNB0981014828, DNB0983557799, DNB098394587X, DNB0984954201, DNB0984975098, DNB0984975160, DNB0984988106, DNB0985471778, DNB0985797150, DNB0985982357, DNB0985985313} Belletristik

Die folgenden DNB-Titeldatensätze enthalten (nur) eine 1- oder 2-stellige *dno_like_i*. Kurzcharakterisierung: DNB985276061, DNB985425598: *dno_like_i*=800 („Literatur, Rhetorik, Literaturwissenschaft“), *CN*=1. DNB983252076: *dno_like_i*=810 („Englische Literatur Amerikas“), *CN*=1. DNB982104707, DNB984123865: *dno_like_i*=820 („Englische Literatur“), *CN*=1. DNB985763426, DNB983874581, DNB984195327, DNB979438810, DNB983429022, DNB982551061, DNB984014152, DNB984250441, DNB985992913, DNB985275898,



DNB985475331, DNB984954201, DNB985982357, DNB985985313: *dno_like_i*=830 („Deutsche Literatur“), *CN*=1. DNB985717505: *dno_like_i*=830 („Deutsche Literatur“), *CN*=0.5. DNB983543208, DNB984858784, DNB947922717, DNB98508720X, DNB985717548, DNB985794070, DNB98582283X, DNB985100419, DNB98549235X, DNB985653779, DNB985870249, DNB981014828, DNB983557799, DNB98394587X, DNB984975098, DNB984975160, DNB984988106, DNB985471778, DNB985797150: *dno_like_i*=830 („Deutsche Literatur“), *CN*=0. DNB98606467X: *dno_like_i*=840 („Franzoesische Literatur“), *CN*=1. DNB98424834X, DNB985186135: *dno_like_i*=840 („Franzoesische Literatur“), *CN*=0.5. DNB98214296X: *dno_like_i*=890 („Literatur in anderen Sprachen“), *CN*=1.

DNB-K081 {DNB0973560940, DNB097362986X, DNB097366942X, DNB097390349X, DNB0974247898, DNB0974948039} Titel von 2005

DNB973560940 mit dem Titel „Fanatics, hypocrites, Christians – Katholiken als stereotype Romanfiguren bei Richardson, Lewis, Radcliffe und Maturin ; Vorformen, Darstellung und Funktion“ enthält nur *dno_like_i*=820 („Englische Literatur“). Wie entscheidend das Ergebnis der automatischen Klassifizierung u. a. von der heuristischen Funktion *cutoff_val* (, die wiederum von der Häufigkeitsverteilung der Deskriptorwerte in *vc_IDB* abhängt), ist hier zu sehen: In Version *vc_dcl / vc_IDB / cutoff_val_dyn=11* (April 2008) ist *CN*=1 (Abb. 47), in *vc_dcl / vc_IDB / cutoff_val_dyn=22* (Oktober 2008) ist *CN*=0 (Abb. 48) und in *vc_dcl / vc_IDB / cutoff_val_stat=200* (Dezember 2008) ist *CN*=1 (Abb. 49). D. h. für solche Fälle (grosse Anzahl von Deskriptorwerten, hier: 24) muß *QCUT* verbessert werden.

```
number of ddc-classified title:      13209
identifier (dno,schedno):           DNB0973560940 (820,820)
DNB DDC notation (MAB2 field 700)  {820}
calculated cutoff value:           11
title:                              Fanatics, hypocrites, Christians
title (remainder):                 Katholiken als stereotype Romanfiguren bei
Richardson, Lewis, Radcliffe und Maturin ; Vor
formen, Darstellung und Funktion
title (series):                     Schriftenreihe Studien zur Anglistik und
Amerikanistik ; Bd. 6 considered descriptor values: |24| {<100>-
peter#prokisch[0], <331>-fanatics[22], <331>-hypocrites[12], <331>-christians[702],
<335>-romanfiguren[1], <335>-richardson[162], <335>-lewis[959], <335>-
radcliffe[84], <335>-maturin[7], <335>-vorformen[16], <335>-darstellung[854],
<335>-katholiken[28], <335>-funktion[468], <335>-stereotype[113], <412@410>-<033A>-
kovac@hamburg[21], <451>-anglistik[71], <451>-amerikanistik[15], <451>-
schriftenreihe[122], <540a>-3-8300-1821-5[0], <902s4>-romangestalt[30], <902s1>-
katholizismus <motiv>[11], <902s2>-stereotyp <motiv>[7], <907s>-katholizismus
<motiv>[11], <907s>-stereotyp <motiv>[7]}
matched descriptor values:          |2| {katholizismus <motiv>, stereotyp <motiv>}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):   |18| =====> more than 15 calculated ddc
classes
calculated1 ddc classes (sections): |10| {809, 812, 813, 820, 822, 823, 828, 830,
841, 863}
calculated1 ddc classes (main):     |1| {800}
```



```

calculated2 ddc classes (subdiv): {809.9164[1], 812.52[1], 813.509[1], 813.54[1],
820.9353[1], 820.9382[1], 820.9[1], 822.33[1], 823.809[1], 823.912[1],
823.91409[1], 828.809[1], 830.9[1], 863.64[1]}
calculated2 ddc classes (sections): {820[3], 823[3]}
calculated2 ddc classes (divisions): {820[8]}
calculated2 ddc classes (main): {800[15]}
correlation (DNB973560940,820): 11x.xxxx xxx xxx (1)

```

Abb. 47: Klassifizierungsergebnis von *in_dnb_DNB973560940* mit *vc_dcl /vc_IDB / cutoff_val_dyn=11* (April 2008)

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB0973560940 (820,820)
DNB DDC notation (MAB2 field 700): {820}
calculated cutoff value: 22
title: Fanatics, hypocrites, Christians
title (remainder): Katholiken als stereotype Romanfiguren bei
Richardson, Lewis, Radcliffe und Maturin ; Vorformen, Darstellung und Funktion
title (series): Schriftenreihe Studien zur Anglistik und
Amerikanistik ; Bd. 6
considered descriptor values: |22| {<100>-peter#prokisch[0], <331>-
fanatics[23], <331>-hypocrites[12], <331>-christians[1150], <335>-romanfiguren[1],
<335>-richardson[196], <335>-lewis[1362], <335>-radcliffe[121], <335>-maturin[8],
<335>-vorformen[16], <335>-darstellung[1318], <335>-katholiken[44], <335>-
funktion[612], <335>-stereotype[126], <412@410>-<033A>-kovac@hamburg[21], <451>-
anglistik[124], <451>-amerikanistik[22], <451>-schriftenreihe[151], <540a>-3-8300-
1821-5[0], <912s4>-romangestalt[32], <912s1>-katholizismus[472], <912s2>-
stereotyp[90]}
matched descriptor values: |4| {<033A>-kovac@hamburg, maturin,
amerikanistik, vorformen}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |3| {340, 370, 823.914}
calculated1 ddc classes (sections): |2| {340, 370}
calculated1 ddc classes (main): |1| {300}
calculated2 ddc classes (subdiv): |0| {}
calculated2 ddc classes (sections): {340[1], 370[1]}
calculated2 ddc classes (divisions): {340[1], 370[1]}
calculated2 ddc classes (main): {300[2]}
correlation (H0708_DNB0973560940,820): 00x.xxxx xxx xxx (0)

```

Abb. 48: Klassifizierungsergebnis von *in_dnb_DNB0973560940* mit *vc_dcl /vc_IDB / cutoff_val_dyn=22* (Oktober 2008)

```

number of ddc-classified title: 2
identifier (dno,schedno): DNB0973560940 (820,820)
DNB DDC notation (MAB2 field 700): {820}
calculated cutoff value: 200
title: Fanatics, hypocrites, Christians
title (remainder): Katholiken als stereotype Romanfiguren bei
Richardson, Lewis, Radcliffe und Maturin ; Vorformen, Darstellung und Funktion
title (series): Schriftenreihe Studien zur Anglistik und
Amerikanistik ; Bd. 6
considered descriptor values: |22| {<100>-peter#prokisch[0], <331>-
fanatics[23], <331>-hypocrites[12], <331>-christians[1150], <335>-romanfiguren[1],
<335>-richardson[196], <335>-lewis[1362], <335>-radcliffe[121], <335>-maturin[8],
<335>-vorformen[16], <335>-darstellung[1318], <335>-katholiken[44], <335>-
funktion[612], <335>-stereotype[126], <412@410>-<033A>-kovac@hamburg[21], <451>-
anglistik[124], <451>-amerikanistik[22], <451>-schriftenreihe[151], <540a>-3-8300-
1821-5[0], <912s4>-romangestalt[32], <912s1>-katholizismus[472], <912s2>-
stereotyp[90]}

```



```

matched descriptor values: |10| {stereotyp, richardson, stereotype,
maturin, amerikanistik, romangestalt, anglistik, fanatics, vorformen, radcliffe}
max. match value of matched descriptor values: |5|
calculated1 ddc classes (subdiv): |3| {813.54, 820.9, 823.914}
calculated1 ddc classes (sections): |3| {813, 820, 823}
calculated1 ddc classes (main): |1| {800}
calculated2 ddc classes (subdiv): |3| {813.54[1], 820.9[1], 823.914[1]}
calculated2 ddc classes (sections): {813[1], 820[1], 823[1]}
calculated2 ddc classes (divisions): {820[2]}
calculated2 ddc classes (main): {800[3]}
correlation (H0708_DNB0973560940,820): 1lx.xxx xxx xxx xxx (1)

```

Abb. 49: Klassifizierungsergebnis von *in_dnb_DNB0973560940* mit *vc_dcl / vc_IDB / cutoff_val_stat=200* (Dezember 2008)

DNB97362986X mit dem Titel „Beschreibungs-Kunst – zur intermedialen Poetik angloamerikanischer Ikontexte 1880 – 2000. Theorie und Geschichte der Literatur und der schoenen Kuenste ; Bd. 110“ ist mit *dno_like_i=810* („Englische Literatur Amerikas“), *dno_like_i=820* („Englische Literatur“) und *dno_like_i=700* („Kuenste, Bildende Kunst allgemein“) klassifiziert. *vc_dcl* ermittelt in *calc2 (subdiv)*: „{700.1[1], 791.4309[1], 833.912[1]}“, weshalb *CN=0.5*. In *calc2 (divisions)* liefert *vc_dcl* als DDC-Notationskandidaten: „{700[1], 790[1], 820[1], 830[1]}“, womit zwei der drei *dno_like_i* getroffen werden. Hilfreich ist hier der intellektuell vergebene Deskriptorwert „ekphrasis“ des Deskriptors „Sachschlagwort“.

DNB97366942X: keine besondere *vc_dcl*-Leistung, da Titeldatensatz (dessen Deskriptorwerte) als *ppn_0479462526* mit *dno=830.993629772* in *vc_IDB* enthalten.

Bei DNB97390349X mit dem Titel „Die Konstituierung des literarischen Feldes in Deutschland nach 1871 – Berliner Moderne, Literaturkritik und die Anfaenge der Literatursoziologie. Studien und Texte zur Sozialgeschichte der Literatur ; Bd. 101“ wurde *dno_like_i=830* („Deutsche Literatur“) und *dno_like_i=300* („Sozialwissenschaften, Soziologie“) vergeben. In *calc2 (subdiv)* wird als DDC-Notationskandidat *dno_a=830.93552* bestimmt, daher *CN=1*. Eine Prüfung ergibt, dass der Titeldatensatz (dessen Deskriptorwerte) als *ppn_0393697754* mit *dno_i=830.93552* („Deutsche Literatur--Geschichte, Darstellung, kritische Bewertung von Werken aus mehr als einer Gattung--Soziologie und Anthropologie“) in *vc_IDB* enthalten sind.

Auch DNB974247898 ist als *ppn_0483312479* mit *dno_like_i=830* („Deutsche Literatur“) in *vc_IDB* enthalten, *vc_dcl* ermittelt *dno_a=830*, daher *CN=1*.

DNB974948039 enthält *dno_like_i=830*. *vc_dcl* schlägt in *calc2 (sections)* *dno_a=830* vor, also ist *CN=1*.

DNB-K082 {DNB097083554X, DNB0971186774, DNB0971268649, DNB0971446385, DNB0971583765, DNB0971608326, DNB0971668493, DNB0969408390, DNB0970621345, DNB0971475008, DNB0971678545, DNB0971850240, DNB0971975272, DNB0972031421} Titel von 2004



Kurz-Charakterisierung: DNB97083554X (*dno_like_i*=820, CN=1); DNB971186774 (*dno_like_i*=890, CN=0.5); DNB971268649 (*dno_like_i*=830, *dno_like_j*=390, CN=1, als ppn_0388118806 mit *dno*=830.93826413 in *vc_IDB!*); DNB971446385 (*dno_like_j*=850, CN=0.5); DNB971583765 (*dno_like_i*=800, CN=1, als ppn_0389985333 mit *dno_i*=809.1=*dno_a* in *vc_IDB!*); DNB971608326 (*dno_like_i*=830, CN=1); DNB971668493 (*dno_like_i*=820, CN=0); DNB969408390 (*dno_like_i*=830, *dno_like_j*=943, CN=1); DNB970621345 (*dno_like_i*=830, CN=0 mit *vc_dcl* (April 2008), CN=1 mit *vc_dcl* / *vc_IDB* (Dezember 2008), wenn *QELIM* durch Löschung der Deskriptorwerte „20.“ und „19.“ verbessert wird); DNB971475008 (*dno_like_i*=870, CN=1, als ppn_0389590541 mit *dno*=872.04 in *vc_IDB!*); DNB971678545 (*dno_like_i*=830, CN=1, als ppn_0390668702 mit *dno*=830 in *vc_IDB!*); DNB971850240 (*dno_like_i*=860, *dno_like_j*=070, CN=1); DNB971975272 (*dno_like_j*=830, CN=1, als ppn_0392581515 mit *dno*=830 in *vc_DB!*); DNB972031421 (*dno_like_i*=830, CN=1, als ppn_0392709341 mit *dno*=831.6 in *vc_IDB!*).

DNB-Excel-Datei-Kommentare zur Hauptklasse 900

DNB-K083 {DNB0975205447, DNB071096015} C-Notation ist korrekt

Der Titeldatensatz DNB975205447 mit dem Titel „Die Bronze- und fruehe Eisenzeit in der ehemaligen Neumark – Untersuchungen zu Forschungsentwicklung und Fundbild bis 1945. Universitaetsforschungen zur praehistorischen Archaeologie ; Bd. 119“ enthält *dno_like_i*=930 („Alte Geschichte, Archaeologie“). *vc_dcl* ermittelt mit den vier Deskriptorwerten „{eisenzeit, praehistorischen, <033A>-habelt@bonn, bronzezeit}“ als DDC-Notationskandidaten in *calc2* (*subdiv*) *dno_a*=930.15 („Kupfer- und Bronzezeit“), *dno_a*=936.3 („Germanische Gebiete bis 481“) und *dno_a*=947.01 („Anfänge bis 862“). Somit ist CN=1.

Der Titeldatensatz DNB071096015 mit dem Titel „Taten und Tugenden Traians – Herrschaftsdarstellung im Principat. Hermes : Einzelschriften ; H. 91“ ist mit *dno_like_i*=930 („Alte Geschichte, Archaeologie“) klassifiziert und als ppn_0387730699 mit *dno*=937.07092 in *vc_DB* enthalten, daher CN=1 (keine besondere *vc_dcl*-Leistung).

DNB-K084 {DNB0982833326} C-Notation ist nicht exakt

Der Titeldatensatz DNB982833326 mit dem Titel ‚Voelkermord oder "tragische" Kriegereignisse? – der historische Hintergrund der Massenvernichtung von Armenien 1915 n. Chr.‘ ist mit *dno_like_i*=940 („Geschichte Europas“) klassifiziert. *vc_dcl* (April 2008) ermittelt in *calc2* (*subdiv*) als *dno_a*=940.5318 („Holocaust“, nicht zutreffend), jedoch *dno_a*=940 in *calc2* (*sections*), daher CN=1. In der Version *vc_dcl* / *vc_IDB* (Dezember 2008) ist aufgrund anderer Häufigkeitsverhältnisse in *vc_IDB* (Abb. 50) CN=0; der (u. a. zu DDC-Klasse 830 gehörende) in *vc_IDB* enthaltene Deskriptorwert



„anton-heinz#schmidt“ ist Ursache hierfür. Für solche Fälle sollte *QCUT* um eine Regel erweitert werden (*QCUT*-Verbesserung):

H4. wenn der Personennamenname `<per_for#per_fam>` in `vc_IDB`, dann `cutoff_val`=nächster Häufigkeitswert aus der Häufigkeitswerteliste `S_list` (s. Fußnote 28).

Das Ergebnis für `DNB982833326` mit H4.-Regel ist in Abb. 51 wiedergegeben.

Durch die zusätzlichen Deskriptorwerte „armenische frage“ und „massaker“ des Deskriptors „Sachschlagwort“ kann `dno_a=956.620154` („Geschichte Asiens des Fernen Ostens--Naher Osten (Mittlerer Osten)--Nordöstliche Türkei--1774–1918“) ermittelt werden (Abb. 52). Ist `dno_a` vertretbar (**F49**)?

```

number of ddc-classified title:      1
identifizier (dno,schedno):         DNB0982833326 (940,940)
DNB DDC notation (MAB2 field 700):  {940}
calculated cutoff value:            161
title:                              Voelkermord oder "tragische" Kriegsereignisse?
title (remainder):                  der historische Hintergrund der
Massenvernichtung von Armenien 1915 n. Chr.
considered descriptor values:       |10| {<100>-anton-heinz#schmidt[2], <331>-
kriegsereignisse[3], <331>-voelkermord[230], <331>-tragische[34], <335>-
massenvernichtung[2], <335>-armenien[273], <335>-chr.[1], <335>-historische[1212],
<335>-hintergrund[161], <412@410>-<033A>-i¿%a.-h.i¿% schmidt@aigen-voglhuh[0]}
matched descriptor values:          |3| {anton-heinz#schmidt, tragische,
hintergrund}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):   |6| {809, 822.33, 830, 831.2, 832.914,
940.5318}
calculated1 ddc classes (sections): |5| {809, 822, 830, 831, 832}
calculated1 ddc classes (main):      |1| {800}
calculated2 ddc classes (subdiv):    |3| {822.33[1], 831.2[1], 832.914[1]}
calculated2 ddc classes (sections):  {809[1], 822[1], 830[1], 831[1], 832[1]}
calculated2 ddc classes (divisions): {830[3]}
calculated2 ddc classes (main):      {800[5]}
correlation (B0745_DNB0982833326,940): 00x.xxx xxx xxx xxx (0)

```

Abb. 50: Klassifizierungsergebnis von `in_dnb_DNB0982833326` mit `vc_dcl / vc_IDB` (Dezember 2008)

```

number of ddc-classified title:      2
identifizier (dno,schedno):         DNB0982833326 (940,940)
DNB DDC notation (MAB2 field 700):  {940}
calculated cutoff value:            280
title:                              Voelkermord oder "tragische" Kriegsereignisse?
title (remainder):                  der historische Hintergrund der
Massenvernichtung von Armenien 1915 n. Chr.
considered descriptor values:       |9| {<331>-kriegsereignisse[3], <331>-
voelkermord[230], <331>-tragische[34], <335>-massenvernichtung[2], <335>-
armenien[273], <335>-chr.[1], <335>-historische[1212], <335>-hintergrund[161],
<412@410>-<033A>-i¿%a.-h.i¿% schmidt@aigen-voglhuh[0]}
matched descriptor values:          |4| {armenien, tragische, voelkermord,
hintergrund}
max. match value of matched descriptor values: |3|

```



```

calculated1 ddc classes (subdiv): |5| {320, 910, 940.5318, 943, 956}
calculated1 ddc classes (sections): |4| {910, 940, 943, 956}
calculated1 ddc classes (main): |1| {900}
calculated2 ddc classes (subdiv): |1| {940.5318[1]}
calculated2 ddc classes (sections): {910[1], 940[1], 943[1], 956[1]}
calculated2 ddc classes (divisions): {940[2]}
calculated2 ddc classes (main): {900[4]}
correlation (B0745_DNB0982833326,940): 11x.xxx xxx xxx (1)

```

Abb. 51: Klassifizierungsergebnis von *in_dnb_DNB0982833326* mit *vc_dcl / vc_IDB / cutoff_val_stat=280* (Dezember 2008)

```

number of ddc-classified title: 3
identifier (dno,schedno): DNB0982833326 (940,940)
DNB DDC notation (MAB2 field 700): {940}
calculated cutoff value: 46
title: Voelkermord oder "tragische" Kriegereignisse?
title (remainder): der historische Hintergrund der
Massenvernichtung von Armenien 1915 n. Chr.
considered descriptor values: |12| {<100>-anton-heinz#schmidt[2], <331>-
kriegereignisse[3], <331>-voelkermord[230], <331>-tragische[34], <335>-
massenvernichtung[2], <335>-armenien[273], <335>-chr.[1], <335>-historische[1212],
<335>-hintergrund[161], <412@410>-<033A>-i.%a.-h.i.% schmidt@aigen-voglhub[0],
<902s1>-armenische frage[6], <902s2>-massaker[45]}
matched descriptor values: |2| {armenische frage, massaker}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |1| {956.620154}
calculated1 ddc classes (sections): |1| {956}
calculated1 ddc classes (main): |1| {900}
calculated2 ddc classes (subdiv): |1| {956.620154[1]}
calculated2 ddc classes (sections): {956[1]}
calculated2 ddc classes (divisions): {950[1]}
calculated2 ddc classes (main): {900[1]}
correlation (B0745_DNB0982833326,940): 10x.xxx xxx xxx (0.5)

```

Abb. 52: Klassifizierungsergebnis von *in_dnb_DNB0982833326* mit *vc_dcl / vc_IDB / cutoff_val_stat=46* und den Deskriptorwerten „armenische frage“ und „massaker“ (Dezember 2008)

DNB-K085 {DNB0982833024, DNB0985187093} gefundene C-Notation ist ziemlich exakt

Der Titeldatensatz [DNB982833024](#) mit dem Titel „Das Raetsel einer Sprache im antiken Kleinasien – Fund einer karisch-griechischen Staatsurkunde in der Tuerkei“ ist mit *dno_like_i=930* („Alte Geschichte, Archaeologie“) klassifiziert. *CN=1*, da in *calc2 (divisions)* *dno_a=930*. Ferner bestimmt *vc_dcl* die DDC-Notationskandidaten in *calc2 (subdiv)*: *dno_a=936.3* („Germanische Gebiete bis 481“) und *dno_a=939.2* („Westliches Kleinasien bis 640“).

Der Titeldatensatz [DNB985187093](#) mit dem Titel „Sichelschnitt“ mit Panzern und Luftwaffe – Hitlers Westfeldzug zerstoert Europas Maechtegleichgewicht“ ist mit *dno_like_i=940* („Geschichte Europas“) klassifiziert, *CN=1*. Ausserdem wird in *calc2 (subdiv)* der DDC-Notationskandidat *dno_a=940.5421* („Europaescher Kriegsschauplatz“) geliefert.

DNB-K086 {DNB0972844619, DNB0973158727} C-Notation ist annaehrend richtig



Der Titeldatensatz DNB972844619 mit dem Titel „Wegbereiter der Shoah – die Waffen-SS, der Kommandostab Reichsfuehrer-SS und die Judenvernichtung 1939 – 1945. Veroeffentlichungen der Forschungsstelle Ludwigsburg der Universitaet Stuttgart ; Bd. 4“ ist mit *dno_like_i=940* („Geschichte Europas“) und *dno_like_i=320* („Politik“) klassifiziert. Da in *calc2 (divisions)* *dno_a=940*, ist *CN=1*. Allerdings ist dies – wie bei einigen Titeldatensätzen (s. Kapitel 4) – keine besondere *vc_dcl*-Leistung, da DNB972844619 als *ppn_0474702111* mit *dno=940.54050947* in *vc_DB* enthalten ist. In *vc_dcl* (April 2008) werden DDC-Notationen auf DDC-Haupttafeln-Notationen gekürzt (vgl. Fußnote 17), daher liefert *vc_dcl* (April 2008) in *calc2 (subdiv)* „nur“ *dno_a=940.5405* („Unterdrückungsmaßnahmen und Grausamkeiten“). Die Version *vc_dcl / vc_IDB* (Dezember 2008) liefert die vollständige *dno_a=940.54050947* in *calc1 (subdiv)*. Wieso diese *dno_a* nicht in *calc2 (subdiv)* enthalten ist, muss geklärt werden (vgl. Fußnote 61).

Auch der Titeldatensatz DNB973158727 mit dem Titel ‚Diktatur und Revolution – Reformation und Bauernkrieg in der Geschichtsschreibung des "Dritten Reiches" und der DDR. Quellen und Forschungen zur Agrargeschichte; Bd. 50‘ und *dno_like_i=943* („Geschichte Deutschlands“) sowie *dno_like_i=230* („Theologie, Christentum“) ist in *vc_IDB* als *ppn_0475505425* (Pica+-Kategorie <003@>) und *GBV0475505425* (Pica+-Kategorie <007G>), mit *dno=943.03072043* enthalten, daher ist *CN=1* wiederum keine spezielle Klassifizierungsleistung von *vc_dcl*.

DNB-K087 {DNB0970335105, DNB0968804969} C-Notation durchaus brauchbar

Analog wie unter DNB-K086 beschrieben. Der Titeldatensatz DNB970335105 ist unter der *ppn_0380101920* mit *dno=909.0492402* in *vc_IDB* enthalten. *CN=0.5*, wegen intellektueller Klassifizierung *dno_like_i=940*, *dno_like_i=290* und *dno_like_i=340*.

Der Titeldatensatz DNB968804969 mit dem Titel „Hegels Schueler C. L. Michelet: Recht und Geschichte jenseits der Schulteilung. Philosophische Schriften ; Bd. 54“ ist mit *dno_like_i=900* („Geschichte“) und *dno_like_i=340* („Recht“) klassifiziert. Mit den fünf Deskriptorwerten „{philosophische, rechtsphilosophie, schueler, hegels, geschichtsphilosophie}“ kommt *vc_dcl* in *calc2 (sections)* zu den DDC-Notationen *dno_a=100* („Philosophie, Parapsychologie und Okkultismus, Psychologie“) und *dno_a=193* („Deutschland und Österreich“), daher *CN=0*.

DNB-K088 {ZDB2391538-9, ZDB2291033-5, ZDB2303051-3, DNB0971862397, DNB0972013156, DNB0974665762, DNB0974992003, DNB097149357X} C-Notation abwegig

ZDB2391538-9 mit dem Titel „Schriften der Stadt Rheinberg zur Geschichte und Heimatkunde“ ist mit *dno_i= 914.3* („Geografie Mitteleuropas und Reisen in Mitteleuropa Geografie Deutschlands und Reisen in Deutschland“) intellektuell klassifiziert. In *WebDewey* ist „914.3“ keine Haupttafel-Notation, daher ist *schedno=914* („Geografie Europas und Reisen in Europa“). *vc_dcl* bestimmt mit „9“



beginnende DDC-Notationskandidaten, in *calc2 (subdiv)* ist u. a. auch *dno_a=917.2*, daher ist $CN=2/3=0.6$. Mit Version *vc_dcl / vc_IDB* (Dezember 2008) ist $CN=1$ (Abb. 53).

```

number of ddc-classified title:      1
identifier (dno,schedno):           ZDB2391538-9 (914.3,914)
DNB DDC notation (MAB2 field 700):  {914.3,z-001}
calculated cutoff value:            20632
title:                              Schriften der Stadt Rheinberg zur Geschichte
und Heimatkunde
considered descriptor values:        |6| {<331>-rheinberg[1], <331>-
geschichte[20632], <331>-heimatkunde[5835], <331>-schriften[1075], <331>-
stadt[2939], <412@410>-<033A>-kath. kirchengemeinde st. peter@rheinberg[0]}
matched descriptor values:           |4| {heimatkunde, schriften, stadt, geschichte}
max. match value of matched descriptor values: |4|
calculated1 ddc classes (subdiv):    |41| {081, 100, 193, 200, 229.1, 253, 270.1,
270.2092, 270.6, 282, 290, 300, 305, 320, 330, 340, 370, 380, 500, 700, 709, 720,
780, 796, 800, 813.54, 830, 833.914, 838.91209, 890, 910, 914.3, 929.20973, 930,
940, 943, 943.41, 943.512, 943.595, 946, 951}
calculated1 ddc classes (sections):  |8| {910, 914, 929, 930, 940, 943, 946, 951}
calculated1 ddc classes (main):      |1| {900}
calculated2 ddc classes (subdiv):    |5| {914.3[1], 929.20973[1], 943.41[1],
943.512[1], 943.595[1]}
calculated2 ddc classes (sections):  {943[4]}
calculated2 ddc classes (divisions): {940[6]}
calculated2 ddc classes (main):      {900[11]}
correlation (B0746_ZDB2391538-9,914): 111.xxx xxx xxx xxx (1)

```

Abb. 53: Klassifizierungsergebnis von *in_dnb_ZDB2391538-9* mit *vc_dcl / vc_IDB* (Dezember 2008)

Der Titeldatensatz ZDB2291033-5 mit dem Titel „Quellen zur Genealogie im Landkreis Saarlouis und angrenzenden Gebieten“ ist mit *dno_like_i=920* („Biografie, Genealogie, Heraldik“) und *dno_like_i=943* („Geschichte Deutschlands“) klassifiziert. *vc_dcl* (April 2008) bestimmt in *calc2 (sections)* *dno_a=016* („Bibliografien und Kataloge von Werken über einzelne Themen oder in einzelnen Fachgebieten“), daher $CN=0$. *vc_dcl / vc_IDB* (Dezember 2008) liefert *dno_a=016.92934* in *calc2 (subdiv)*, weshalb wieder $CN=0$ (Abb. 54). Ist *dno_a* (Abb. 55) wirklich „abwegig“ (F50)?

```

number of ddc-classified title:      1
identifier (dno,schedno):           ZDB2291033-5 (920,920)
DNB DDC notation (MAB2 field 700):  {920,943,z-000}
calculated cutoff value:            169
title:                              Quellen zur Genealogie im Landkreis Saarlouis
und angrenzenden Gebieten
considered descriptor values:        |7| {<331>-landkreis[169], <331>-saarlouis[5],
<331>-angrenzenden[36], <331>-gebieten[66], <331>-quellen[574], <331>-
genealogie[426], <412@410>-<033A>-kreisarchiv@saarlouis[1]}
matched descriptor values:           |5| {landkreis, gebieten, angrenzenden,
saarlouis, <033A>-kreisarchiv@saarlouis}
max. match value of matched descriptor values: |4|
calculated1 ddc classes (subdiv):    |1| {016.92934}
calculated1 ddc classes (sections):  |1| {016}
calculated1 ddc classes (main):      |1| {000}
calculated2 ddc classes (subdiv):    |1| {016.92934[1]}
calculated2 ddc classes (sections):  {016[1]}
calculated2 ddc classes (divisions): {010[1]}
calculated2 ddc classes (main):      {000[1]}

```



```
correlation (B0746_ZDB2291033-5,920): 00x.xxx xxx xxx xxx (0)
```

Abb. 54: Klassifizierungsergebnis von *in_dnb_ZDB2291033-5* mit *vc_dcl* / *vc_IDB* (Dezember 2008)

```
016.92934 <ul2_to_analyze; length: 9>
0----- Computer science, information & general works <hatzen>
01----- Bibliographies <hatzen>
016----- Bibliographies and catalogs of works on specific subjects or in specific
           disciplines <hat>
---.9---- History & geography <na2r7:9>
---.92--- Biography & genealogy <na2r7:92>
---.929-- Genealogy, names, insignia <na2r7:929>
---.9293- Genealogical sources <na2r7:929.3>
---.92934 Treatment by specific continents, countries, localities <na2r7:929.34>
```

Abb. 55: *vc_daygram* der *dno_a=016.92934*

Der Titeldatensatz ZDB2303051-3 mit Titel „Jahrheft von Schlieren ...“ enthält *dno_like_i=940* („Geschichte Europas“) und *dno_like_i=910* („Geografie, Reisen“). Nur der (mehrdeutige)⁶² Deskriptorwert „schlieren“ trägt zum Klassifizierungsergebnis (*CN=0*) bei. Ein besseres Ergebnis ist ev. durch *QTIT*-Verbesserung (exakt die Zeichenkette „schlieren“ kommt zwar nur in acht DDC-Klassen vor, allerdings nicht in den mit „9“ beginnenden DDC-Klassen (Abb. 56); „heimatkunde“ kommt in 5.835 DDC-Klassen vor und sollte mit einem aussagekräftigeren Deskriptorwert ersetzt werden) oder durch *QIDB*-Verbesserung möglich.

```
<033A>-neue bruecke@schlieren-zuerichY274.3
<033A>-verl. zuercher zeitung@schlierenY830.9494
<033A>-verlagsanstalt bruecke g@schlieren-zuerichY282.494
optik schlierenmethodeY535.2
schlieren methods opticsY535.2 535.33 535.4 535
schlieren methods opticsY621.08 621.369 621.36
schlieren photographyY533.2028 535.3
schlierenmethodeY535.2
schlierenY535.2 535.33 535.3
schlierenY620.10640287 620 621.08 629.134353 629.13452
```

Abb. 56: Auszug aus *vc_IDB* (Dezember 2008), Deskriptorwerte mit Zeichenkette „schlieren“

Der Titeldatensatz DNB971862397 mit dem Titel „Die Annales Quedlinburgenses – Monumenta Germaniae Historica : Scriptorum : 7, Scriptorum rerum Germanicarum in usum scholarum separatim editi ; 72“ ist mit *dno_like_i=943* („Geschichte Deutschlands“) klassifiziert. *vc_dcl* (April 2008) ermittelt wegen der Deskriptorwerte „{germaniae, editi, usum, quedinburg, scholarum}“ mit „4“ beginnende DDC-Notationskandidaten, deshalb *CN=0*. Durch Löschung der Deskriptorwerte „editi“ und „usum“ (*QELIM*-Verbesserung) kann *CN=1* mit Version *vc_dcl* / *vc_IDB* (Dezember 2008) erreicht werden (Abb. 57). In *calc2* (*subdiv*) wird *dno_a=943.02* („843–1519“) ermittelt.

```
number of ddc-classified title: 1
identifier (dno,schedno): DNB0971862397 (943,943)
DNB DDC notation (MAB2 field 700): {943}
calculated cutoff value: 105
```

⁶² schlieren: 1. Stadt im Kanton Zürich, Schweiz und 2. optisches Phänomen



```

title:                               ^Die^ Annales Quedlinburgenses
title (other):                       ^Die^ Annales Quedlinburgenses
title (series):                      Monumenta Germaniae Historica : Scriptorum : 7,
Scriptores rerum Germanicarum in usum scholarum separatim editi ; 72
considered descriptor values:        |14| {<100>-martina#giese[0], <331>-
annales[1230], <331>-quedlinburgenses[0], <412@410>-<033A>-hahn@hannover[99],
<451>-scriptores[119], <451>-rerum[1321], <451>-germanicarum[0], <451>-
scholarum[41], <451>-separatim[1], <451>-monumenta[105], <451>-germaniae[72],
<451>-historica[1092], <540a>-3-7752-5472-2[0], <907g>-quedlinburg[19]}
matched descriptor values:           |3| {germaniae, monumenta, <033A>-
hahn@hannover}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):    |1| {943.02}
calculated1 ddc classes (sections):  |1| {943}
calculated1 ddc classes (main):      |1| {900}
calculated2 ddc classes (subdiv):    |1| {943.02[1]}
calculated2 ddc classes (sections):  {943[1]}
calculated2 ddc classes (divisions): {940[1]}
calculated2 ddc classes (main):      {900[1]}
correlation (B0746_DNB0971862397,943): 111.xxx xxx xxx xxx (1)

```

Abb. 57: Klassifizierungsergebnis von *in_dnb_DNB0971862397* ohne „editi“ und „usum“ (Dez. 2008)

Der Titeldatensatz [DNB972013156](#) mit dem Titel „André François-Poncet als Botschafter in Berlin – 1931 – 1938. Pariser historische Studien ; Bd. 64“ ist mit *dno_like_i=940* und *dno_like_i=943* klassifiziert und – wie sich herausstellt – als *ppn_0393116514* bzw. *GBV0393116514* (Pica+-Kategorie <007G>) mit *dno=327.2029* (Abb. 58) enthalten, daher wird in *calc2 (sections)* *dno_a=327* ermittelt und in *calc2 (subdiv)* ist *dno_a=327.2029* als einer von 23 DDC-Notationskandidaten enthalten.

```

327.2029 <ul3_to_analyze; length: 8>
3----- Social sciences <hatzen>
32----- Political science <hatzen>
327----- International relations <hat>
327.2--- Diplomacy <hat>
---.-02- Miscellany <T1--02>
---.-029 Commercial miscellany <T1--029>

```

Abb. 58: *vc_daygram* der *dno=327.2029*

Der Titeldatensatz [DNB0974665762](#) mit dem Titel „Der niederlaendische Aufstand gegen Spanien 1568 – 1609 – eine kommunikationswissenschaftliche Analyse. Studien zur Geschichte und Kultur Nordwesteuropas ; Bd. 12“ enthält *dno_like_i=940*. *vc_dcl* (April 2008) bestimmt in *calc2 (divisions)* die DDC-Notationskandidaten „{010[1], 270[1], 340[1], 740[1], 940[1]}“, daher *CN=1*, jedoch auch in *calc2 (subdiv)* *dno_a=741.9*. In der Version *vc_dcl / vc_IDB* (Dezember 2008) ist auch *CN=1*, das Klassifizierungsergebnis ist jedoch eindeutiger (Abb. 59).

```

number of ddc-classified title:      1
identifier (dno,schedno):            DNB0974665762 (940,940)
DNB DDC notation (MAB2 field 700):  {940}
calculated cutoff value:             20632
title:                               ^Der^ niederlaendische Aufstand gegen Spanien
1568 - 1609
title (remainder):                  eine kommunikationswissenschaftliche Analyse
title (series):                     Studien zur Geschichte und Kultur

```



```

Nordwesteuropas ; Bd. 12
considered descriptor values:      |16| {<100>-dirk#maczkiewitz[0], <331>-
gegen[1241], <331>-spanien[1939], <331>-niederlaendische[43], <331>-aufstand[247],
<335>-eine[8273], <335>-kommunikationswissenschaftliche[4], <335>-analyse[3803],
<412@410>-<033A>-waxmann@muenster[1], <451>-kultur[4856], <451>-nordwesteuropas[0],
<451>-geschichte[20632], <540a>-3-8309-1521-7[0], <902g>-niederlande[1029],
<902s1>-kommunikation[1417], <902s2>-achtzigjaehriger krieg[4]}
matched descriptor values:        |10| {kultur, eine, gegen, kommunikation,
niederlande, spanien, analyse, aufstand, niederlaendische, geschichte}
max. match value of matched descriptor values: |10|
calculated1 ddc classes (subdiv): |1| {940}
calculated1 ddc classes (sections): |1| {940}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):   |0| {}
calculated2 ddc classes (sections): {940[1]}
calculated2 ddc classes (divisions): {940[1]}
calculated2 ddc classes (main):     {900[1]}
correlation (H0708_DNB0974665762,940): 1lx.xxx xxx xxx xxx (1)

```

Abb. 59: Klassifizierungsergebnis von *in_dnb_DNB0974665762* mit *vc_dcl / vc_IDB* (Dezember 2008)

Ähnliches gilt für DNB0974992003, das *vc_dcl*-Ergebnis kann von *CN=0.5* (Version: April 2008) auf *CN=1* (Version: Dezember 2008) verbessert werden (Abb. 60). In *calc2 (subdiv)* wird zusätzlich *dno_a=955.03* („Zeit der persischen Dynastien, 1499–1794“) ermittelt.

```

number of ddc-classified title:    1
identifier (dno,schedno):          DNB0974992003 (950,950)
DNB DDC notation (MAB2 field 700): {950}
calculated cutoff value:           202
title:                             Politik und Handel zwischen Ming und Timuriden
title (remainder):                 China, Iran und Zentralasien im
Spaetmittelalter
title (series):                     Iran - Turan ; Bd. 7
considered descriptor values:       |13| {<100>-ralph#kauz[0], <331>-ming[2679],
<331>-timuriden[3], <331>-politik[7087], <331>-handel[955], <335>-
zentralasien[202], <335>-spaedmittelalter[152], <335>-china[13702], <335>-
iran[1577], <412@410>-<033A>-reichert@wiesbaden[165], <451>-turan[9], <540a>-3-
89500-388-3[0], <902s>-mingdynastie[20]}
matched descriptor values:         |5| {zentralasien, <033A>-reichert@wiesbaden,
timuriden, spaetmittelalter, turan}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):   |3| {930, 943, 955.03}
calculated1 ddc classes (sections): |3| {930, 943, 955}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):   |1| {955.03[1]}
calculated2 ddc classes (sections): {930[1], 943[1], 955[1]}
calculated2 ddc classes (divisions): {930[1], 940[1], 950[1]}
calculated2 ddc classes (main):     {900[3]}
correlation (H0708_DNB0974992003,950): 1lx.xxx xxx xxx xxx (1)

```

Abb. 60: Klassifizierungsergebnis von *in_dnb_DNB0974992003* mit *vc_dcl / vc_IDB* (Dezember 2008)

Der Titeldatensatz DNB097149357X mit dem Titel „Marudu – the destruction and reconstruction of a coastal state in Borneo. Austronesia ; Vol. 5“ ist mit *dno_like_i=950* („Geschichte Asiens“), *dno_j=959.53* und *schedno=959.5* („Malaysia, Brunei, Singapur“) klassifiziert. *vc_dcl* ermittelt mit den beiden Deskriptorwerten „<033A>-abera@hamburg“ und „borneo“ den DDC-Notationskandidaten *dno_a=320.9*, weshalb *CN=0* ist. Zwar sind Deskriptorwerte des Deskriptors Schlagwort vergeben



(„<902gl>-bandau kota marudu <region>[0]“ und „<907g>-bandau kota marudu <region>[0]“), sie befinden sich jedoch nicht in *vc_IDB* (wie an dem in eckiger Klammer stehenden Häufigkeitswert „0“ zu erkennen ist). In „<...>“ stehende Deskriptorwerte wie „<region>“ werden in der neuen Programmversion *vc_dcl* (Dezember) gelöscht. Aber auch die Deskriptorwerte „austronesia“, „bandau kota marudu“, „kota marudu“ oder „marudu“ sind nicht in *vc_IDB* enthalten. Für ein besseres Klassifizierungsergebnis müsste *QTIT* und/oder *QIDB* verbessert werden.

DNB-K089 {ZDB1388266-1, ZDB2288182-7, DNB0960245766, DNB0982811683} völlig abwegige C-Notation

Ursache: Bei folgenden Titeldatensätzen wurden keine Deskriptorwerte des Deskriptors „Schlagwort“ vergeben und die Deskriptorwerte, die für die Klassifizierung übrigbleiben („considered descriptor values“) / zur Klassifizierung beitragen („matched descriptor values“) sind unzureichend: ZDB1388266-1: „{grundwissenschaften, einzeldarstellungen}“; ZDB2288182-7: „{berlinische}“; DNB0960245766: {schottische, fuenf, baenden} (vgl. auch DNB-K094). Beim Titeldatensatz DNB982811683 mit dem Titel „Pekings Probleme mit der kommunistischen Vergangenheit – gravierende Luecken in der Geschichte der Volksrepublik China“ und *dno_like_i*=950 („Geschichte Asiens“) tragen je zwei Deskriptorwerte zu den mit „3“ beginnenden DDC-Notationskandidaten bei. In *vc_dcl* / *vc_IDB* (Dezember 2008) werden in *calc2* (*subdiv*) die DDC-Notationskandidaten: „{320.95105[1], 327.51[1]}“ ermittelt (Abb. 61 - 63), daher nach wie vor *CN*=0, aber ev. vertretbar (**F51**)?

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0982811683 (950,950)
DNB DDC notation (MAB2 field 700):  {950}
calculated cutoff value:            256
title:                              Pekings Probleme mit der kommunistischen
Vergangenheit
title (remainder):                 gravierende Luecken in der Geschichte der
Volksrepublik China
considered descriptor values:        |11| {<100>-anton-heinz#schmidt[2], <331>-
kommunistischen[883], <331>-vergangenheit[631], <331>-pekings[4], <331>-
probleme[2176], <335>-geschichte[20632], <335>-volksrepublik[256], <335>-
china[13702], <335>-gravierende[0], <335>-luecken[8], <412@410>-<033A>-aa.-h.a
schmidt@aigen-voglhuh[0]}
matched descriptor values:           |4| {anton-heinz#schmidt, volksrepublik,
pekings, luecken}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |4| {320.95105, 327.51, 341, 830}
calculated1 ddc classes (sections):  |3| {320, 327, 341}
calculated1 ddc classes (main):      |1| {300}
calculated2 ddc classes (subdiv):    |2| {320.95105[1], 327.51[1]}
calculated2 ddc classes (sections):  {320[1], 327[1], 341[1]}
calculated2 ddc classes (divisions): {320[2]}
calculated2 ddc classes (main):      {300[3]}
correlation (dnb_B0745_DNB0982811683,950): 00x.xxx xxx xxx xxx (0)

```

Abb. 61: Klassifizierungsergebnis von *in_dnb_DNB0982811683* mit *vc_dcl* / *vc_IDB* (Dezember 2008)



```

320.95105 <ul4_to_analyze; length: 9>
3----- Social sciences <hatzen>
32----- Political science <hatzen>
320----- Political science (Politics and government) <hat>
320.9---- Political situation and conditions <hat>
320.951-- China--politics, . . . <hatien>
---.-5--- Asia    Orient    Far East <na4r7span:320.901-320.99:T2--5>
---.-51-- China and adjacent areas <na4r7span:320.901-320.99:T2--51>
---.-05  Serial publications <T1--05>

```

Abb. 62: *vc_daygram* der *dno_a*=320.95105

```

327.51 <ul5_to_analyze; length: 6>
3----- Social sciences <hatzen>
32---- Political science <hatzen>
327--- International relations <hat>
327.5- Foreign relations of specific continents and localities, foreign relations
of specific countries <hatspan:327.3-327.9:327.5>
---.5- Asia    Orient    Far East <na4r7span:327.3-327.9:T2--5>
---.51 China and adjacent areas <na4r7span:327.3-327.9:T2--51>

```

Abb. 63: *vc_daygram* der *dno_a*=327.51

DNB-K090 {ZDB504201-x, DNB0982811195, DNB0982812027, DNB0982814844, DNB0982814925, DNB0982833539, DNB0981531172, ZDB1111645-6} C-Notation falsch

Der Titeldatensatz ZDB504201-x mit dem Titel „Berliner geographische Arbeiten“ und *dno_i*=914.3 („Geografie Mitteleuropas und Reisen in Mitteleuropa Geografie Deutschlands und Reisen in Deutschland“) wird von *vc_dcl* / *vc_IDB* (Dezember 2008) in *dno_a*=540 („Chemie und zugeordnete Wissenschaften“) klassifiziert (Abb. 64), da (ungünstigerweise) alle drei Deskriptorwerte „<033A>-akademie-verl.@berlin“, „geographische“ und „arbeiten“ gemeinsam in dieser DDC-Klasse vorkommen. Laut [http://www.geographie.hu-berlin.de/forschung/berliner_geographische_arbeiten] decken die „Berliner geographische Arbeiten ein breites Spektrum geographischer Interessenfelder ab. Als Sammelbände von Beiträgen und als Monographien zu Themen der physischen Geographie, der Human- und Sozialgeographie, der Wirtschaftsgeographie, der angewandten Geographie und der Fachdidaktik geben die Berliner Geographischen Arbeiten Einblick in die wissenschaftliche Arbeit des Instituts.“ Trotz breitem Spektrums der „Berliner geographische Arbeiten“ ist *dno_a*=540 eine Fehlklassifizierung, die durch zusätzliche Deskriptorwerte z. B. aus dem Beschreibungstext („sozialgeographie“, „humangeographie“ und „wirtschaftsgeographie“) auf z. B. *dno_a*=910 (CN=0.6) korrigiert werden kann (Abb. 65).

```

number of ddc-classified title:      1
identifier (dno,schedno):            ZDB504201-x (914.3,914)
DNB DDC notation (MAB2 field 700):  {914.3,z-766}
calculated cutoff value:            346
title:                               Berliner geographische Arbeiten
considered descriptor values:        |4| {<331>-geographische[276], <331>-
arbeiten[346], <412@410>-<033A>-akademie-verl.@berlin[90], <542a>-1430-4775[0]}
matched descriptor values:          |3| {<033A>-akademie-verl.@berlin,
geographische, arbeiten}
max. match value of matched descriptor values: |3|
calculated ddc classes (subdiv):     |1| {540}

```



```

calculated1 ddc classes (sections): |1| {540}
calculated1 ddc classes (main): |1| {500}
calculated2 ddc classes (subdiv): |0| {}
calculated2 ddc classes (sections): {540[1]}
calculated2 ddc classes (divisions): {540[1]}
calculated2 ddc classes (main): {500[1]}
correlation (dnb_B0745_ZDB504201-x,914): 000.xxx xxx xxx xxx (0)

```

Abb. 64: Klassifizierungsergebnis von *in_dnb_ZDB504201-x* mit *vc_dcl* / *vc_IDB* (Dezember 2008)

```

number of ddc-classified title: 2
identifier (dno,schedno): ZDB504201-x (914.3,914)
DNB DDC notation (MAB2 field 700): {914.3,z-766}
calculated cutoff value: 346
title: Berliner geographische Arbeiten
considered descriptor values: |7| {<331>-geographische[276], <331>-
arbeiten[346], <412@410>-<033A>-akademie-verl.@berlin[90], <542a>-1430-4775[0],
<902s1>-sozialgeographie[135], <902s2>-humangeographie[2], <902s3>-
wirtschaftsgeographie[265]}
matched descriptor values: |5| {wirtschaftsgeographie, geographische,
humangeographie, sozialgeographie, arbeiten}
max. match value of matched descriptor values: |5|
calculated1 ddc classes (subdiv): |1| {910}
calculated1 ddc classes (sections): |1| {910}
calculated1 ddc classes (main): |1| {900}
calculated2 ddc classes (subdiv): |0| {}
calculated2 ddc classes (sections): {910[1]}
calculated2 ddc classes (divisions): {910[1]}
calculated2 ddc classes (main): {900[1]}
correlation (dnb_B0745_ZDB504201-x,914): 110.xxx xxx xxx xxx (0.666667)

```

Abb. 65: Klassifizierungsergebnis von *in_dnb_ZDB504201-x* mit *vc_dcl* / *vc_IDB* und *QTIT*-Verbesserung mit den Deskriptorwerten „sozialgeographie“, „humangeographie“ und „wirtschaftsgeographie“ (Dezember 2008)

Wie bei den Titeldatensätzen in DNB-K089 wurden hier keine Deskriptorwerte des Deskriptors „Schlagwort“ vergeben und / oder die zur Klassifizierung verwendbaren Deskriptorwerte sind unzureichend (*QTIT*- / *QIDB*-Verbesserung notwendig): DNB982812027: {triest, anton-heinz#schmidt} (*dno_like_i*=940, *dno_a*=830, *CN*=0); DNB982814844: {anton-heinz#schmidt, schlesien} (*dno_like_i*=940, *dno_a*=830, *CN*=0); ZDB1111645-6: {<033A>-rheticus-ges.@feldkirch, vorarlberg} (*dno_like_i*=940, *dno_like_j*=910, *dno_j*=943.645005, *schedno*=943.64, *dno_a*=324, *CN*=0); DNB982814925: {anton-heinz#schmidt, chr.} (*dno_like_i*=943, *dno_a*=935, *CN*=0.34). Der Deskriptorwert „chr.“ muß selbstverständlich eliminiert werden, was zur Folge hat, dass nur noch der Deskriptorwert „anton-heinz#schmidt“ des Deskriptors Personennamen übrig bleibt, der in *vc_IDB* (nur) in den DDC-Klassen „304.80943909021“ und „830“ vorkommt; DNB982811195: *vc_dcl* (April 2008): {jahrestag, entstalinisierung} (*dno_like_i*=940, *dno_a*=320.943, *CN*=0); *vc_dcl* / *vc_IDB* (Dezember 2008) {jahrestag, anton-heinz#schmidt, ansehen, entstalinisierung} (*dno_like_i*=940, *dno_a*=940, *CN*=1); DNB982833539: {anton-heinz#schmidt, keine, originale} (*dno_like_i*=940, *CN*=0); die Hinzufügung des Deskriptorwertes „buendnis“ erzielt *CN*=1 (*vc_dcl* / *vc_IDB*, Dezember 2008)).

Von den in Betracht kommenden Deskriptorwerte „{20., marggraf, pirna, loanda}“ des



Titel datensatzes DNB981531172 (*dno_a*=910, *CN*=0) wird nach Löschung des Deskriptorwertes „20.“ (*QELIM*-Verbesserung) mit *vc_dcl* / *vc_IDB* (Dezember 2008) *CN*=1 erzielt.

DNB-K091 {ZDB2299456-7} C-Notation ist auch achtstellig, aber am Ende falsch

Der Titeldatensatz ZDB2299456-7 mit dem Titel „Zeitschrift fuer mitteldeutsche Familiengeschichte“ ist mit *dno_like_j*=920, *dno_j*=929.20943 und *schedno*=929.2 klassifiziert. *vc_dcl* (April 2008) liefert (Abb. 66):

```

number of ddc-classified title:      12478
identifier (dno,schedno):           ZDB2299456-7 (929.20943,929.2)
DNB DDC notation (MAB2 field 700)  {920,100,305,z681}
calculated cutoff value:            32
title:                              Zeitschrift fuer mitteldeutsche
Familiengeschichte
title (remainder):                  ZMFG
considered descriptor values:       |5| {<331>-familiengeschichte[32], <331>-
mitteldeutsche[5], <335>-zmfmg[0], <412@410>-<033A>-amf@leipzig[0], <542zEUR>-
8.75[4]}
matched descriptor values:          |3| {8.75, mitteldeutsche, familiengeschichte}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv):   |41| =====> more than 15 calculated ddc
classes
calculated1 ddc classes (sections): |7| {927, 929, 940, 943, 945, 956, 968}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):   {929.20973[1], 929.2[1], 929.3[1], 929.42[1],
929.6[1], 929.7[1], 940.5318[1], 943.004924[1], 943.1[1], 943.6[1], 956.94[1],
968.04[1]}
calculated2 ddc classes (sections): {929[7]}
calculated2 ddc classes (divisions):{940[9]}
calculated2 ddc classes (main):     {900[19]}
correlation (ZDB2299456-7,929.2):  111.1xx xxx xxx (1)

```

Abb. 66: Klassifizierungsergebnis von *in_dnb_ZDB2299456-7* mit *vc_dcl* (April 2008)

Nach Löschung des Deskriptorwertes „8.75“ des Deskriptors „542z“ (= „keine ISSN, aber Einbandart und/oder Preis“) wird weiterhin *CN*=1 erzielt (Abb. 67) und *dno_j* ist in *calc2* (*subdiv*) enthalten.

```

number of ddc-classified title:      1
identifier (dno,schedno):           ZDB2299456-7 (929.20943,929.2)
DNB DDC notation (MAB2 field 700):  {920,z-100,z-305,z-z681}
DDC notation (MAB2 field 705):     {929.20943}
calculated cutoff value:            42
title:                              Zeitschrift fuer mitteldeutsche
Familiengeschichte
title (remainder):                  ZMFG
title (other):                      ZMFG
considered descriptor values:       |4| {<331>-familiengeschichte[42], <331>-
mitteldeutsche[6], <335>-zmfmg[0], <412@410>-<033A>-amf@leipzig[0]}
matched descriptor values:          |2| {mitteldeutsche, familiengeschichte}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv):   |48| {016.9293089924, 130, 230.09224322,
284.2092, 304.6, 305.8924043416409034, 305.896392, 305.899220492, 306.8740943,

```



```

333.0943, 339.209433, 370.196, 540, 610, 780.922, 789.71, 830, 830.8, 830.82,
832.3, 833.91, 927.8, 929, 929.2089924043, 929.20899240436, 929.20899240438,
929.20943, 929.209436, 929.20947, 929.20973, 929.30899240435514, 929.3089924043593,
929.34342, 929.42089924, 929.60943412, 929.73, 940, 940.5318, 940.5318092,
940.53180922, 943, 943.004924, 943.182, 943.3, 943.61304410923924, 945.31040922,
956.94004924043092, 968.04092}
calculated1 ddc classes (sections): | 7 | {927, 929, 940, 943, 945, 956, 968}
calculated1 ddc classes (main): | 1 | {900}
calculated2 ddc classes (subdiv): | 12 | {927.8[1], 929.209436[1], 929.20943[1],
929.20947[1], 929.20973[1], 929.34342[1], 929.73[1], 940.5318[1], 943.004924[1],
943.182[1], 943.3[1], 968.04092[1]}
calculated2 ddc classes (sections): {929[14]}
calculated2 ddc classes (divisions): {920[15]}
calculated2 ddc classes (main): {900[26]}
correlation (dnb_B0748_ZDB2299456-7,929.2): 111.1xxx xxx xxx (1)

```

Abb. 67: Klassifizierungsergebnis von *in_dnb_ZDB2299456-7* (Dez. 2008) und Löschung von „8.75“

DNB-K092 {DNB098558470X} C-Notation auch vierstellig, aber am Ende falsch (Nicht völlig, aber viel zu allgemein)

Der Titeldatensatz DNB098558470X mit dem Titel “Domesticating space – construction, community, and cosmology in the late prehistoric Near East. Studies in early near eastern production, subsistence, and environment ; Vol. 12” ist mit *dno_like_i=930* (“Alte Geschichte, Archäologie”) und *dno_i=939.4* (“Naher Osten bis 640”) intellektuell klassifiziert. *vc_dcl* schlägt mithilfe der Deskriptorwerte „{production, east, community, early, near, late, cosmology, subsistence, construction, eastern, space, environment, prehistoric}” in *calc2 (divisions)* u. a. *dno_a=930* vor, daher *CN=2/4=0.5*. In *calc2 (subdiv)* wird *dno_a=304.2* (“Humanökologie“) und *dno_a=930.1* („Archäologie“) ermittelt. Ob irgendwann vier- (und mehr)stellige DDC-Notationen im großen Stil für bibliographische Titeldatensätze zufriedenstellend automatisch ermittelt werden können, ist z. Zt. eine offene Frage.

DNB-K093 {DNB0980606411} synthetische Notation

Der Titeldatensatz DNB980606411 mit dem Titel „Konventioneller Krieg oder NS-Weltanschauungskrieg? – Kriegführung und Partisanenbekaempfung in Frankreich 1943/44. Quellen und Darstellungen zur Zeitgeschichte ; Bd. 69.“ ist mit *dno_like_i=940* („Geschichte Europas“), *dno_j=940.54214* und *schedno=940.5421* („Europäischer Kriegsschauplatz“) intellektuell klassifiziert. *vc_dcl* ermittelt in *calc2 (subdiv)* *dno_a=940.5421*, daher ist *CN=7/7=1* (Vergleich mit *schedno*, s. Fußnote 17).

DNB-K094 {DNB0960245766, ZDB2303051-3, ZDB2390603-0} Freitext-Schlagwörter wie Geschichte werden gar nicht erkannt und beim Matchen daher nicht berücksichtigt

Beim Titeldatensatz DNB960245766 mit dem Titel „Schottische Geschichte – in fuenf Baenden“ kommen in *vc_dcl* (April 2008) folgende Deskriptorwerte in Betracht: „{<100>-hagen#seehase[0], <331>-schottische[23], <331>-geschichte[7575], <335>-fuenf[306], <335>-baenden[121],



<412@410>-<033A>-koenig@greiz/thueringen[0], <902g>-schottland[527]}“. Das bedeutet, dass z. B. der Deskriptorwert des Deskriptors „Personenname“ „hagen#seehase“ und der Deskriptorwert „<412@410>-<033A>-koenig@greiz/thueringen“ des Deskriptors „Verleger@Ort“ keinmal und der Deskriptorwert „geschichte“ des Deskriptors „Hauptsachtitel“ in 7.575 *vc_IDB*-DDC-Klassen vorkommt. Die heuristische Funktion ist *cutoff_val_dyn*=306, daher wird der Deskriptorwert „geschichte“ nicht weiter berücksichtigt (vgl. Kapitel 2.1.6). Alle drei Deskriptorwerte „{schottische, fuenf, baenden}“ tragen in *calc2 (sections)* zu *dno_a*=346 („Privatrecht“) und *dno_a*=780 („Musik“) bei. Intellektuell ist mit *dno_like_i*=940 („Geschichte Europas“) und *dno_i*=941.1 („Schottland“) klassifiziert, daher *CN*=0. Durch Löschung der (irrelevanten) Deskriptorwerte „fuenf“ und „baenden“ (*QELIM*-Verbesserung) und durch Hinzufügung z. B. der einschlägigen Deskriptorwerte „bannockburn“ und „stuarths“ (*QTIT*-Verbesserung) kann mit *vc_dcl / vc_IDB* (Dezember 2008) *CN*=1 erreicht werden (Abb. 68).

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0960245766 (941.1,941.1)
DNB DDC notation (MAB2 field 700):  {940}
DDC notation (MAB2 field 705):      {941.1}
calculated cutoff value:            55
title:                              Schottische Geschichte
title (remainder):                 in fuenf Baenden
considered descriptor values:       |7| {<100>-hagen#seehase[0], <331>-
schottische[23], <331>-geschichte[20632], <412@410>-<033A>-
koenig@greiz/thueringen[0], <902g>-schottland[844], <902g>-bannockburn[11], <902g>-
stuarths[55]}
matched descriptor values:          |3| {bannockburn, schottische, stuarths}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):   |1| {941.1}
calculated1 ddc classes (sections): |1| {941}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):   |1| {941.1[1]}
calculated2 ddc classes (sections): {941[1]}
calculated2 ddc classes (divisions):{940[1]}
calculated2 ddc classes (main):     {900[1]}
correlation (dnb_A0748_DNB0960245766,941.1): 111.1xx xxx xxx xxx (1)

```

Abb. 68: Klassifizierungsergebnis von *in_dnb_DNB0960245766* mit *vc_dcl / vc_IDB* (Dezember 2008) / *QELIM*- und *QTIT*-Verbesserung (Löschung der Deskriptorwerte „fuenf“ und „baenden“ und Hinzufügung der Deskriptorwerte „bannockburn“ und „stuarths“)

Zu Titeldatensatz [ZDB2303051-3](#) siehe DNB-K088.

[ZDB2390603-0](#) hat den Titel „Katalog der Ausstellung im Museum Heimathaus Muensterland in Telgte“ und ist mit *dno_i*= 914.3 („Geografie Mitteleuropas und Reisen in Mitteleuropa Geografie Deutschlands und Reisen in Deutschland“)⁶³ und *schedno*=914 („Geografie Europas und Reisen in Europa“) klassifiziert. *vc_dcl* (April 2008) bestimmt u. a. *dno_a*=940 („Geschichte Europas Geschichte Westeuropas“), weshalb *CN*=1/3=0.3 ist. *vc_dcl / vc_IDB* (Dezember 2008)

⁶³ Die DDC-Klasse „914.3“ kommt in der amerikanischen Ausgabe nicht vor, vgl. auch DNB-K088.



ermittelt u. a. *dno_a*= 936.3 („Germanische Gebiete bis 481“), CN bleibt CN=0.3.

DNB-K095 {ZDB2392134-1, ZDB2253436-2, ZDB1111645-6} viel zu unspezifischer Titel

Vgl. DNB-K089. QTIT- / QIDB-Verbesserung notwendig: ZDB2392134-1: {lebenswelten}; ZDB2253436-2: {<033A>-univ. toronto press@toronto}.

Beim Titeldatensatz ZDB1111645-6 mit dem Titel „Rheticus - Vierteljahresschrift der Rheticus-Gesellschaft“ schlägt *vc_dcl* / *vc_IDB* (Dezember 2008) *dno_a*=324.24364507509041 (Abb. 69) in *calc* (*subdiv*) (vgl. Fußnote 61) vor (Abb. 70), da die beiden Deskriptorwerte „<033A>-rheticus-ges.@feldkirch“ (Abb. 71) und „vorarlberg“ (Abb. 72) gemeinsam in *dno_a* vorkommen.

```

324.24364507509041 <ul6_to_analyze; length: 18>
3----- Social sciences <hatzen>
32----- Political science <hatzen>
324----- The political process <hat>
324.2----- Political parties <hat>
324.24----- Parties in specific countries in modern world <hatspan:324.24-
324.29:324.24>
324.243----- Parties of Germany <hat>
---.-4----- Europe Western Europe <ba4r2span:324.24-324.29:T2--4>
---.-43----- Central Europe Germany <ba4r2span:324.24-324.29:T2--43>
---.-436----- Austria and Liechtenstein <ba4r2span:324.24-324.29:T2--436>
---.-4364----- Western Austria, and Liechtenstein <ba4r2span:324.24-324.29:T2--
4364>
---.-43645----- Vorarlberg <ba4r2span:324.24-324.29:T2--43645>
---.-07----- Recent parties identified primarily by position on right-to-left
spectrum <ba4r2spanrenspan:324.24-324.29+03-07>
---.-075----- Communist parties <ba4r2span:324.24-324.29+075>
---.-09----- Historical, geographic, persons treatment <T1--09>
---.-0904- 20th century, 1900-1999 <T1--0904>
---.-09041 1900-1919 <T1--09041>

```

Abb. 69: *vc_daygram* der *dno_a*=324.24364507509041

```

number of ddc-classified title: 1
identifier (dno,schedno): ZDB1111645-6 (943.645005,943.64)
DNB DDC notation (MAB2 field 700): {940,910,z-370}
DDC notation (MAB2 field 705): {943.645005}
calculated cutoff value: 89
title: Rheticus
title (remainder): Vierteljahresschrift der Rheticus-Gesellschaft
considered descriptor values: |6| {<331>-rheticus[3], <335>-
vierteljahresschrift[2], <335>-rheticus-gesellschaft[0], <412@410>-<033A>-rheticus-
ges.@feldkirch[3], <902g>-vorarlberg[89], <902s11>-heimatkunde[5835]}
matched descriptor values: |2| {<033A>-rheticus-ges.@feldkirch,
vorarlberg}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |1| {324.24364507509041}
calculated1 ddc classes (sections): |1| {324}
calculated1 ddc classes (main): |1| {300}
calculated2 ddc classes (subdiv): |0| {}
calculated2 ddc classes (sections): {324[1]}

```



```
calculated2 ddc classes (divisions):{320[1]}
calculated2 ddc classes (main):      {300[1]}
correlation (B0748_ZDB1111645-6,943.64): 000.XXx xxxx xxxx (0)
```

Abb. 70: Klassifizierungsergebnis von *in_dnb_ZDB1111645-6* mit *vc_dcl / vc_IDB* (Dezember 2008)

```
<033A>-rheticus-ges.@feldkirch u.a.<9f>943.6450532086914
<033A>-rheticus-ges.@feldkirch<9f>304.8094364509034 324.24364507509041
<033A>-rheticus-ges.@feldkirch<9f>943.66
```

Abb. 71: Auszug aus der invertierten *DDC-Datenbasis vc_IDB* („<033A>-rheticus-ges.@feldkirch“)

```
vorarlberg <franzoesische zone><9f>940.531440943642
vorarlberg alpen geologie fuehrer<9f>554.3645
vorarlberg austria<9f>301.32943645 304.8094364509034 305.8924043645 320.94364509041
324.24364507509041 328.4364505 330.94364205 333.7160943645 344.36450828
348.43645028
vorarlberg austria<9f>813.54
vorarlberg austria<9f>914.36450453 914.3645045 914.364504 920 940.31424
940.41343642 940.5314409436 940 943.4600492401734 943.64500222 943.645004943500222
943.6450440222 943.6450510202 943.645051 943.6450522087 943.645
vorarlberg franzoesische besatzung<9f>940.531440943642
vorarlberg kunstdenkmaeler<9f>914.364504
vorarlberg literatur<9f>838.9140809943645
vorarlberg o.c. literatur literaturgeschichte<9f>838.9140809943645
vorarlberg region<9f>016.5543645
vorarlberg region<9f>595.780943645
vorarlberg; ges 3759a 1945-1946<9f>940.531440943642
vorarlberg; ges bombenkrieg<9f>940.534364
vorarlberg; kun<9f>914.364504
vorarlberg<9f>016.43746 016.5543645
vorarlberg<9f>133.430943645
vorarlberg<9f>301.32943645 305.8924043645 305 320.94364509041 322.10943609033
324.243603809043 324.24364507509041 324.364508 328.43645 330.94364205 331.31094364
331.880943645 333.7160943645 338.10943645 343.436450786902632 343.4364507869
344.3605527 344.360577 344.3607 344.364071090 344.364503786902632 344.3645037869
345.4360527 345.436077 347.43640109 347.436 350 362.10943621 363.69094363 370.943
376.943645 398.209494
vorarlberg<9f>430 437.46 437.645 437.64
vorarlberg<9f>500 512.02 530.1201512 549.943642 551.6943642 551.770943 554.3645
vorarlberg<9f>609.43642
vorarlberg<9f>700 720.94364509045 720.943645 720 770 783.0260943645 784.8
786.5194364 796
vorarlberg<9f>813.54 830.80943645 838.9140809943645
vorarlberg<9f>910 914.3462 914.363 914.36420453 914.36450022222 914.36450453
914.3645045 914.364504 920 929.34342 940.31424 940.41343642 940.531440943642
940.531440943645 940.531610820947 940.5343645 940.534364 940 943.4600492401734
943.47004924 943.6420522 943.64500222 943.645004924 943.645004943500222
943.6450440222 943.6450510202 943.645051 943.6450522087 943.6450532086914 943.645
```

Abb. 72: Auszug aus der invertierten *DDC-Datenbasis vc_IDB* („vorarlberg“)

DNB-K096 {DNB0981|23910} SW gematcht

Je zwei Deskriptorwerte (darunter auch Deskriptorwert „frank#huss“ des Deskriptors Personennamen) von vier Deskriptorwerten kommen gemeinsam in den DDC-Notationskandidaten vor (Abb. 73), mit *vc_dcl / vc_IDB* (Dezember 2008) ist *CN=0.75*:



```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB0981123910 (940.2,940.2)
DNB DDC notation (MAB2 field 700):  {940}
DDC notation (MAB2 field 705):      {940.2}
calculated cutoff value:            174
title:                              Grosse Maenner, grosse Schwaechen!
title (remainder):                  die Wahrheit ueber unvergessene Personen der
Weltgeschichte
considered descriptor values:       |13| {<100>-frank#huss[1], <331>-schwaechen[7],
<331>-grosse[2021], <331>-maenner[9159], <335>-unvergessene[10], <335>-
personen[174], <335>-weltgeschichte[742], <335>-wahrheit[661], <335>-ueber[9176],
<412@410>-<033A>-koenig@greiz[0], <540a>-3-934673-69-4[0], <902g>-europa[8171],
<902sl>-herrscher[142]}
matched descriptor values:          |4| {frank#huss, personen, herrscher,
schwaechen}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):   |9| {150, 610, 833.914, 900, 929.73, 940,
940.1, 943, 943.02}
calculated1 ddc classes (sections): |4| {900, 929, 940, 943}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):   |3| {929.73[1], 940.1[1], 943.02[1]}
calculated2 ddc classes (sections): {940[2], 943[2]}
calculated2 ddc classes (divisions):{940[4]}
calculated2 ddc classes (main):     {900[6]}
correlation (A0745_DNB0981123910,940.2): 111.0xxx xxx xxx (0.75)

```

Abb. 73: Klassifizierungsergebnis von *in_dnb_DNB0981123910* mit *vc_dcl / vc_IDB* (Dezember 2008)

DNB-K097 {DNB098379846X} Zufallstreffer??

Nein. Der Titeldatensatz DNB098379846X mit dem Titel „... weitertragen – Anneliese Knoop-Graf ueber Willi Graf und die Weisse Rose ; Live-Mitschnitt eines Vortrages am 2. Dezember 2006 in Muenchen.“ enthält *dno_like_i=943* („Geschichte Deutschlands“), *dno_j=943.0864092* und *schedno=943.0864* („Zeit des Zweiten Weltkriegs, 1939–1945“). *vc_dcl / vc_IDB* (Dezember 2008) bestimmt mit den Deskriptorwerten „{anneliese#knoop-graf, weitertragen, willi}“ in *calc2 (subdiv)* die DDC-Notationskandidaten „943.086“ („Zeit des Dritten Reichs, 1933–1945“) und „953.086“ (953: “Arabische Halbinsel und benachbarte Gebiete“)⁶⁴, daher ist $CN=6/7=0.85714$. In *vc_IDB* sind zu DNB98379846X zwei inhaltlich sehr ähnliche Titeldatensätze⁶⁵ enthalten, deren Deskriptorwerte zu dem Klassifizierungsergebnis beitragen (Abb. 74).

```

number of ddc-classified title:      1
identifier (dno,schedno):           DNB098379846X (943.0864092,943.0864)
DNB DDC notation (MAB2 field 700):  {943}

```

⁶⁴ Höchstwahrscheinlich Tippfehler in *ppn_0339011122* (Pica+-Kategorie <045F>): müsste „943.086“ statt „953.086“ heissen.

⁶⁵ 1. *ppn_0339011122*: ‚Weitertragen – Studien zur "Weissen Rose" ; Festschrift fuer Anneliese Knoop-Graf zum 80. Geburtstag‘ mit *dno_j=953.086*; 2. *ppn_017949029X*: ‚Gewalt und Gewissen – Willi Graf und die "Weisse Rose" ; eine Dokumentation von Klaus Vielhaber. In Zsarb. mit Hubert Hanisch u. Anneliese Knoop-Graf mit *dno_j=943.086*



```

DDC notation (MAB2 field 705):      {943.0864092}
calculated cutoff value:          57
title:                            ... weitertragen
title (remainder):                Anneliese Knoop-Graf ueber Willi Graf und die
Weisse Rose ; Live-Mitschnitt eines Vortrages am 2. Dezember 2006 in Muenchen
title (series):                    Yeotone-Dokumente
considered descriptor values:      |13| {<100b>-anneliese#knoop-graf[3], <331>-
weitertragen[1], <335>-willi[57], <335>-muenchen[852], <335>-weisse[474], <335>-
rose[2315], <335>-live-mitschnitt[2], <335>-vortrages[9], <335>-dezember[566],
<335>-ueber[9176], <412@410>-<033A>-yeotone@walzbachtal[0], <451b>-yeotone-
dokumente[0], <540a>-3-9811526-0-3[0]}
matched descriptor values:         |3| {anneliese#knoop-graf, weitertragen, willi}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):  |2| {943.086, 953.086}
calculated1 ddc classes (sections): |2| {943, 953}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):  |2| {943.086[1], 953.086[1]}
calculated2 ddc classes (sections): {943[1], 953[1]}
calculated2 ddc classes (divisions): {940[1], 950[1]}
calculated2 ddc classes (main):     {900[2]}
correlation (A0747_DNB098379846X,943.0864): 111.111 0xx xxx xxx (0.857143)

```

Abb. 74: Klassifizierungsergebnis von *in_dnb_DNB098379846X* mit *vc_dcl* / *vc_IDB* (Dezember 2008)

DNB-K098 {DNB0986077488} hier fehlt die linguistische Komponente (Rückführung von Weltkrieges auf Weltkrieg hätte sicher bessere Ergebnisse gebracht)

Nein. DNB986077488 ist mit *dno_like_i=943* („Geschichte Deutschlands“) und *dno_i=943.430864* und *schedno=943* („Mitteleuropa Deutschland“) klassifiziert. Mit und ohne Deskriptorwert des Deskriptors „weltkrieg“ bestimmt *vc_dcl* / *vc_IDB* (Dezember 2008) den *cutoff_val_dyn=23*. Der Deskriptorwert „weltkrieges“ kommt jedoch (in mehr als 23, nämlich) in 151 DDC-Klassen und der Deskriptorwert „weltkrieg“ in 991 DDC-Klassen vor (vgl. Abb. 75). *CN=1*, da *schedno=943* und *dno_a=943* in *calc2* (*sections*) enthalten ist. In *vc_IDB* kommen keine Titeldatensätze mit *dno_i=943.430864* vor, sondern nur welche mit *dno_i=943.43* und DDC-Oberklassen hierzu.

```

number of ddc-classified title:    1
identifier (dno,schedno):          DNB0986077488 (943.430864,943)
DNB DDC notation (MAB2 field 700): {943}
DDC notation (MAB2 field 705):     {943.430864}
calculated cutoff value:          23
title:                            Archaeologie des II. Weltkrieges am Mittelrhein
title (remainder):                Bodendenkmaeler, Bauten und Ruinen erinnern
considered descriptor values:      |12| {<100b>-wolfgang#gueckelhorn[1], <331>-
weltkrieges[151], <331>-mittelrhein[23], <331>-archaeologie[1357], <331>-ii.[2],
<335>-ruinen[29], <335>-erinnern[125], <335>-bodendenkmaeler[4], <335>-bauten[237],
<412@410>-<033A>-helios@aachen[12], <902g>-mittelrhein-gebiet[10], <902s>-
weltkrieg[991]}
matched descriptor values:         |3| {<033A>-helios@aachen, mittelrhein-gebiet,
mittelrhein}
max. match value of matched descriptor values: |3|
calculated1 ddc classes (subdiv):  |2| {940, 943}
calculated1 ddc classes (sections): |2| {940, 943}
calculated1 ddc classes (main):     |1| {900}
calculated2 ddc classes (subdiv):  |0| {}
calculated2 ddc classes (sections): {940[1], 943[1]}
calculated2 ddc classes (divisions): {940[2]}

```



```
calculated2 ddc classes (main):      {900[2]}
correlation (A0746_DNB0986077488,943): 111.xxx xxx xxx (1)
```

Abb. 75: Klassifizierungsergebnis von *in_dnb_DNB0986077488* mit *vc_dcl / vc_IDB* (Dezember 2008)

DNB-K099 {DNB0982811934} hier fehlt linguistische Komponente, Erkennen von römische(n) Republik hätte besseres Ergebnis gebracht

Welches Ergebnis wäre besser (**F52**)? Vgl. Argumentation analog DNB-K098. Der Titeldatensatz DNB982811934 enthält *dno_like_i=930* („Alte Geschichte, Archaeologie“). *vc_dcl* schlägt *dno_a=830* vor, daher *CN=0*. Mit der zusätzlichen heuristischen Regel H4. (vgl. DNB-K084) und dem Deskriptorwert „roemische geschichte“ (*QTIT*-Verbesserung) des Deskriptors Sachschlagwort kann *CN=1* erreicht werden (Abb. 76). Durch H4. (*QCUT*-Verbesserung) wird der *cutoff_val_dyn=118* (*vc_dcl*, April 2008), der (nur) zu den (unzureichenden) Deskriptorwerten „{inszenierung, anton-heinz#schmidt}“ führt, auf *cutoff_val_dyn=903* erhöht (*vc_dcl / vc_IDB*, Dezember 2008). Dadurch tragen je zwei der vier Deskriptorwerte „{inszenierung, anton-heinz#schmidt, roemische geschichte, roemischen}“ zu dem automatischen Klassifizierungsergebnis bei.

```
number of ddc-classified title:      1
identifier (dno,schedno):            DNB0982811934 (930,930)
DNB DDC notation (MAB2 field 700):   {930}
calculated cutoff value:             903
title:                               Politik und Inszenierung in der Roemischen
Republik
considered descriptor values:        |7| {<100>-anton-heinz#schmidt[2], <331>-
roemischen[903], <331>-republik[1753], <331>-politik[7087], <331>-
inszenierung[166], <412@410>-<033A>-a.-h.a schmidt@aigen-voglhub[0], <902s>-
roemische geschichte[16]}
matched descriptor values:           |4| {inszenierung, anton-heinz#schmidt,
roemische geschichte, roemischen}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):     |19| {225.61, 270.1, 300, 302.23, 321.6, 430,
720, 759, 759.4, 780, 780.92, 800, 820, 830, 937.01, 937.06, 937.07, 940, 943}
calculated1 ddc classes (sections):   |6| {720, 759, 780, 937, 940, 943}
calculated1 ddc classes (main):       |2| {700, 900}
calculated2 ddc classes (subdiv):     |5| {759.4[1], 780.92[1], 937.01[1], 937.06[1],
937.07[1]}
calculated2 ddc classes (sections):   {937[3]}
calculated2 ddc classes (divisions):  {930[3]}
calculated2 ddc classes (main):       {700[5], 900[5]}
correlation (B0745_DNB0982811934,930): 11x.xxx xxx xxx (1)
```

Abb. 76: Klassifizierungsergebnis von *in_dnb_DNB0982811934* mit *vc_dcl / vc_IDB* (Dezember 2008), Verbesserung von *QCUT* und Deskriptorwert „roemische geschichte“

DNB-K100 {ZDB2098423-6} Ausstellungskatalog sollte Stopwort sein

Ob Deskriptorwerte mit der Zeichenkette „ausstellungskatalog“ aus *vc_IDB* eliminiert werden sollten (vgl. Abb. 77), muss an mehreren Titeldatensätzen überprüft werden. Im Fall des Titeldatensatzes



ZDB2098423-6 mit *dno_like_i=940* („Geschichte Europas“) ist das Vorhandensein der Deskriptorwerte „ausstellungskatalog“ und „ausstellungskataloge“ nicht ausschlaggebend, sondern entscheidend ist hier, dass *QTIT* verbessert werden sollte, z. B. mit dem Deskriptorwert „<033A>-vorarlberger landesarchiv@bregenz“ (Abb. 78).

```

aausstellungskatalog ausstellung coburg band einzelnÿ769.4994303
aausstellungskatalog ausstellung wolfenbuettel ausstellung hamburgÿ220.531
aausstellungskatalogÿ745.670949310740365
ausstellungskatalog koelnÿ306
ausstellungskatalog koelnÿ730.0993 746.92095980740355
ausstellungskatalog kunstsammlung schoenborn <familie>ÿ709.407443324
ausstellungskatalogeÿ398.210922
ausstellungskatalogeÿ759.2 760.0446
ausstellungskatalogÿ011.3109434645 011.44 015.4771 016.83080928209041 016.83099282
016.8316 016.83 016.87101 016.943004924 027.543 027.670904 060 070.509431550904
070.92 074.09033 083.1 090.7401471 098.120943
ausstellungskatalogÿ193 199.492
ausstellungskatalogÿ200.9363 220.531 238.20744336 270 271.1204336 271.7914007443324
271.900434351 274.347 282.0924 282.092 282.43612 291.0951074431554 291.14 291.213
296.46074464164 296.461 296.8341092 299.9
ausstellungskatalogÿ301.208 303.482094 303.48243059807459822 303.482489704359709033
303.48251040903 305.48892400747471 305.8007479467 305.892404333 305.8924043431
305.89240533074 305.892405667 305.89435043074435538 306.07401444 306.089974
306.089980811 306.098346 306.09953 306.74094307443515 325.21089924051132
331.625610431554 344.30552 345.43052 363.731 365.4509470904 385.094359
387.223074031554 390 391.0089971 391.0089972 391.008998 391.0094380740384
391.2094409034 391.434089971 391.434097281 391.43 394.1094 394.2509433 399.0899915
ausstellungskatalogÿ590 595.7648
ausstellungskatalogÿ607.434646 610.89924 615.5320943 629.1 630 641.013 641.5123
652.3 676.2207443122 686.4 688.6
ausstellungskatalogÿ700.411 700.89969883074013 700.924 700.92 700.94307404361
700.965 700 704.03924044361 704.03924 704.03970074016686 704.039707477311 704.03972
704.0397 704.03980810740336 704.039915094295 704.303970711074 704.9423094307403515
704.9482094343 704.94855 704.94894095496 704.94896074569442 704.9489967
704.9499794053 707.40355 708.161 708.332 708.336 708.341 708.3 709.01108997
709.01109795074019777 709.011099307494 709.01109931074013 709.01109942907409944
709.01109950740956944 709.0110996 709.011 709.01507401471 709.01 709.040074031554
709.17498 709.24 709.2 709.366 709.410740214 709.42074 709.42 709.43074031554
709.4307403 709.4332407443324 709.43 709.44362 709.549607477311 709.56074443155
709.56709045 709.701 709.70902 709.72074013 709.720747471 709.7252 709.7281 709.94
709.95307494 709.9 709 711.409443607444361 711.40944361 711.4094933 712.092 720.924
720.92 720 721.822 728.810288 728 729 730.0899635 730.08997 730.09364
730.096707403515 730.0972074017866 730.097253 730.098 730.09953074094933 730.09
730.8996333 730.924 730.95074449457 730.9670744494 730.9701 730.973 730.996 730 732.2
732 736.98 737 738.089970789 738.0951074434164 738.3096 738.37 738.382 739.20936
739.2278074443825 739.2278507483315 739.227861074493222 739.2308998 739.2708997079
739.3074346 739.752 740 741.943 741.94409031074 745.08931071 745.08997073074016686
745.08997073 745.0899942 745.0940740346 745.0940740355 745.096707477173 745.09701
745.09721 745.0973074753 745.098507401444 745.09958507443471 745.2074435954
745.449411 745.52 745.5922108997007401471 745.5 745.67094507404361 745 746.074449454
746.089972 746.0899742 746.094074094945 746.098207409821 746.308997 746.397194
746.39791 746.41 746.662082 746.6709176710740336 746.9208998 746.92 746
758.3095407477311 759.011099429 759.1 759.36 759.3909034074435911 759.3 759.5632
759.7074435917 759.94965 759.94972 759.9561 759.9940899915 759.994140899915
759.994295 759 760.0443609435 760.0924 760.092 769.0740347 779.0924 779.092 779.443
779.89453 779.9301 779.9943155085 779.994765 779.9978 779 780.924 780.92 780
781.743612 791.43624940954
ausstellungskatalogÿ813.52 830.8 830 831.208 831.209354 831.21 831.7 831.912

```



```

833.912 834.709321812 838.609
ausstellungskatalogÿ900 909.04924 909.4 909 910.4 910.916472 910 912.074 912.434
912 914.563204928 929.2089924043 929.607444361 930 932.0074431554 932.007443563 932
934 936.302 936.3 936.4007444361 936.4 937 938.3 939.2800740436 939.8 940.1
940.24074 940.4150924 940.5314409433022 940.531503924 940.5318074753 940.5318092
940.53180943155 940.531844 940.5318 940.53432142 940.5405074 940.542143610222
940.5421707443155 940.547243 940 942.055092 942.1004924 943.004924007443155
943.004924008621 943.02074435511 943.022092 943.0260924 943.08074031554
943.1087707443613 943.15508707443155 943.17 943.182200222 943.18 943.2103
943.347004924 943.36408507443364 943.3 943.416074074434715 943.41640049240074434164
943.41 943.4605 943.4607092 943.4626087 943.471 943.500740359 943.5500792400744355
943.55380049240074435538 943.553 943.55 943.560049240074 943.565 943.5914004924
943.591705074435917 943.6004924007443613 943.600492422 943.6004924 943.60307443615
943.613004924007443613 943.613004924 943.6130407404361 943.645 943.7004924007443712
943.71004924007443712 943.86 943 944.013092 944.01 944.033092 944.04082 944.06
944.10074434124 944.36104074 944.361 944.362014074 944.5270207444527
945.00492400747471 945.083074435514 946.004924 947.004924 947.010743364
947.52004924 949.6180150222 949.742007474423 949.84 950 951.5060744336 954
960.0710434164 963.0074 970.00497 970.3 970.65717 972.0180744212 972.81016074435958
972.8721 972.976007472976 973.04924 973.049700922 975.26049244043 978.00497
978.020924 979.00497 979.007479493 979.502 980.01074435954 980.65717 985.010740359
985 988.3 995.60074
berlin baudenkmale ausstellungskatalogÿ720.28809431554074431554
china schatz ausstellungskatalogÿ709.51074432142
ggr 920.5 alte karten ausstellungskatalogeÿ911.0744947222 912.07403 912.07443416
912.09
kreml peter grosse ausstellungskatalog uebersee-museum bremenÿ947.05092
malerei graphik:ausstellungskataloge:kataloge 20. jhd.ÿ709.04807443155
709.4307403512 709.4360744365 760.094307443155
malerei graphik:ausstellungskataloge:kataloge ueber mehrere
jahrhunderteÿ759.9407443912
malerei/graphik ausstellungskataloge 20. jh.ÿ303.48241052
malerei/graphik ausstellungskataloge 20. jh.ÿ709.04007443
mexic/kunst/keramik/maya/mexiko/ausstellungskatalog/art ceramicsÿ738
mexic/kunst/keramik/maya/mexiko/ausstellungskatalog/art ceramicsÿ972.01
mikroskop <ausstellungskatalog>ÿ681.41307402574
muenchen ausstellung ausstellungskatalogÿ688.74107443324
plastik:ausstellungskataloge 20. jhd.ÿ730.92
plastik:ausstellungskataloge 20. jhd.ÿ730.949407449479
plastik:ausstellungskataloge:kataloge 20. jhd.ÿ709.04807443155
rom nazarener rom" ausstellungskatalogÿ759.307405632

```

Abb. 77: Auszug aus der invertierten DDC-Datenbasis vc_IDB (mit „ausstellungskatalog“)

```

number of ddc-classified title: 1
identifier (dno,schedno): ZDB2098423-6 (940,940)
DNB DDC notation (MAB2 field 700): {940,z-140}
calculated cutoff value: 108
title:
Landesarchivs
Ausstellungskatalog des Vorarlberger
title (other):
Landesarchivs
Ausstellungskataloge des Vorarlberger
considered descriptor values: |5| {<331>-landesarchivs[108], <331>-
ausstellungskatalog[411], <331>-vorarlberger[26], <370a>-ausstellungskataloge[3],
<412@410>-<033A>-vorarlberger landesarchiv@bregenz[2]}
matched descriptor values: |4| {vorarlberger, landesarchivs, <033A>-
vorarlberger landesarchiv@bregenz, ausstellungskataloge}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |3| {350, 759.2, 940}
calculated1 ddc classes (sections): |3| {350, 759, 940}
calculated1 ddc classes (main): |3| {300, 700, 900}
calculated2 ddc classes (subdiv): |1| {759.2[1]}

```



```

calculated2 ddc classes (sections): {350[1], 759[1], 940[1]}
calculated2 ddc classes (divisions): {350[1], 750[1], 940[1]}
calculated2 ddc classes (main):      {300[1], 700[1], 900[1]}
correlation (B0748_ZDB2098423-6,940): 1|x.xxx xxx xxx (1)

```

Abb. 78: Klassifizierungsergebnis von *in_dnb_ZDB2098423-6* mit *vc_dcl / vc_IDB* (Dezember 2008), mit Deskriptorwert „<033A>-vorarlberger landesarchiv@bregenz“

DNB-KI01 {DNB0983540004} Ziffern nicht matchen! (hier 17, 18) - Stoppwörter!, sondern sollte auch Stoppwort sein

Ja, s. Fußnote 10. Durch Eliminierung der Deskriptorwerte „17.“ und „18.“ (*QELIM*-Verbesserung) wird der mit *dno_like_i=943* („Geschichte Deutschlands“), *dno_i=943.00491497* und *schedno=943* intellektuell klassifizierte Titeldatensatz *DNB983540004* von *vc_dcl / vc_IDB* (Dezember 2008) mit *dno_a=940.5318*, d. h. $CN=2/3=0.66$ automatisch klassifiziert.

DNB-KI02 {DNB0971652821} Zahlen sollten Stoppwörter sein (hier achten, neunten)

Ja, s. Fußnote 10. Der Titeldatensatz *DNB971652821* mit dem Titel „Karolingische Rekognitionszeichen – die Kanzlerzeile und ihre graphische Ausgestaltung auf den Herrscherurkunden des achten und neunten Jahrhunderts. Elementa diplomatica ; 10.“ enthält *dno_like_i=943* („Geschichte Deutschlands“) und *dno_i=943.014* („751–843“). *DNB971652821* wird (aufgrund *cutoff_val_dyn=30*) von *vc_dcl / vc_IDB* (Dezember 2008) (nur) mit den Deskriptorwerten „{achten, neunten}“ mit *dno_a=831.91* („Deutsche Versdichtung--1900–1990, . . .“) klassifiziert, somit ist $CN=0$. Trotz der Eliminierung der Deskriptorwerte „achten“ und „neunten“ bleibt $CN=0$, es werden mit „3“ beginnende DDC-Notationskandidaten vorgeschlagen. Mit dem (zusätzlichen) Deskriptorwert „mittelalter geschichte“ des Deskriptors „Sachschlagwort“ kann $CN=0.66$ erreicht werden (Abb. 79).

```

number of ddc-classified title:      1
identifier (dno,schedno):            DNB0971652821 (943.014,943.014)
DNB DDC notation (MAB2 field 700):  {943}
DDC notation (MAB2 field 705):      {943.014}
calculated cutoff value:            30
title:                              Karolingische Rekognitionszeichen
title (remainder):                  die Kanzlerzeile und ihre graphische
Ausgestaltung auf den Herrscherurkunden des achten und neunten Jahrhunderts
title (series):                      Elementa diplomatica ; 10
considered descriptor values:        |17| {<100>-peter#worm[0], <331>-
karolingische[30], <331>-rekognitionszeichen[0], <335>-ihre[3643], <335>-
graphische[113], <335>-ausgestaltung[53], <335>-herrscherurkunden[0], <335>-
jahrhunderts[3106], <335>-kanzlerzeile[0], <412@410>-<033A>-univ.-
bibliothek@marburg[1], <451>-elementa[30], <451>-diplomatica[93], <540a>-3-8185-
0390-7[0], <902>-4794213-7 kanzlei[0], <902s1>-koenigsurkunde[11], <902s2>-
gegenzeichnung[1], <902s2>-mittelalter geschichte[20]}
matched descriptor values:           |2| {koenigsurkunde, mittelalter geschichte}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |3| {940.1, 943.02, 943.024}
calculated1 ddc classes (sections):  |2| {940, 943}
calculated1 ddc classes (main):      |1| {900}
calculated2 ddc classes (subdiv):    |3| {940.1[1], 943.024[1], 943.02[1]}

```



```

calculated2 ddc classes (sections): {943[2]}
calculated2 ddc classes (divisions):{940[3]}
calculated2 ddc classes (main):      {900[3]}
correlation (H0711_DNB0971652821,943.014): 111.100 xxx xxx xxx (0.666667)

```

Abb. 79: Klassifizierungsergebnis von *in_dnb_DNB0971652821* mit *vc_dcl* / *vc_IDB* (Dezember 2008), ohne Deskriptorwerte „achten“ und „neunten“ und mit Deskriptorwert „mittelalter geschichte“

DNB-K103 {DNB0982833849} Sachgruppen 914.3, 940, 943 werden natürlich von der Maschine nicht korrekt erkannt, da wir hier in DNB von den hundreds abweichen

Was ist damit genau gemeint (**F53**)? Zu „914.3“ vgl. auch DNB-K088, ZDB2391538-9. *vc_dcl* / *vc_IDB* kann tatsächlich DDC-Notationen (*dno_i*'s) und DDC-ähnliche Notationen (*dno_like_i*'s) nicht voneinander unterscheiden, wenn in den GVK-DDC-Titeldatensätzen in der Pica+-Kategorie <045F> („DDC“) nicht (nur) *dno_i*'s, sondern auch *dno_like_i*'s (, die in der Pica+-Kategorie <045G>“ stehen sollten) abgespeichert sind. In Abb. 80 sind die Klassenbenennungen der DDC-(ähnlichen)-Notationen „914.3“, „940“ und „943“ (die alle in *vc_IDB* enthalten sind) wiedergegeben. Der Titeldatensatz DNB0982833849 ist mit *dno_like_i=940* („Geschichte Europas“) klassifiziert. *vc_dcl* ermittelt mit den Deskriptorwerten „{buendnis, bern, netzwerk}“ *dno_a=940* („Geschichte Europas Geschichte Westeuropas“) in *calc2 (divisions)* und *dno_a=943* („Mitteleuropa Deutschland“) in *calc2 (sections)*, daher CN=1.

	<i>dno_i</i>	<i>dno_like_i</i>
914.3	Geografie Mitteleuropas und Reisen in Mitteleuropa Geografie Deutschlands und Reisen in Deutschland	Landeskunde Deutschlands
940	Geschichte Europas Geschichte Westeuropas	Geschichte Europas
943	Mitteleuropa Deutschland	Geschichte Deutschlands

Abb. 80 Klassenbenennungen der DDC-(ähnlichen)Notationen „914.3“, „940“ und „943“

DNB-K104 {DNB098400971X, DNB0979445949} Die erste DNB-Notation wurde nicht herangezogen! Matching läuft wohl immer nur auf die jeweils letzte DNB-Notation?! - Fehler im Programm!; wiederum nur 2. DNB-Notation verglichen

Ja, s. Fußnote 26.

Der Titeldatensatz DNB98400971X mit dem Titel „Augustus und die Juden – Rechtsstellung und Interessenpolitik der kleinasiatischen Diaspora. Studien zur Alten Geschichte ; Bd. 6.“ enthält *dno_like_i=290* („Andere Religionen“), *dno_like_i=930* („Alte Geschichte, Archaeologie“), *dno_i=296.09014* („Zweite Tempelsperiode, 586 v. Chr.–70 n. Chr.“) und *dno_i= 937.06* („Zeit des Römischen Kaiserreichs, 31 v.Chr.–476 n.Chr.“). Mit je zwei der sechs Deskriptorwerte



„{kleinasiatischen, monika#schul, provinzialverwaltung, interessenpolitik, roemisches recht, anatolien} bestimmt *vc_dcl* mit „3“ beginnende DDC-Notationskandidaten: *dno_a*=320.9 („Politische Situation und politische Bedingungen“) und *dno_a*=338.04 („Unternehmertum“) in *calc2 (subdiv)*, daher ist *CN*=0 (Abb. 81). Auch die Version *vc_dcl / vc_IDB* (Dezember 2008) liefert *CN*=0 mit 939.2 („Westliches Kleinasien bis 640“), 955.03 („Zeit der persischen Dynastien, 1499–1794“) und 956.1015 („1300–1918“). Von den vergebenen Deskriptorwerten des Deskriptors „Schlagwort“ liegen die Deskriptorwerte „judentum“, „roemisches reich“, „rechtsprechung“ und „roemisches recht“ über dem *cutoff_val_dyn*=95. Ob die automatische Klassifizierung bei diesem Titeldatensatz verbessert werden kann, muß an anderer Stelle untersucht werden, es ist keine einfache / schnelle Lösung sichtbar.

```

number of ddc-classified title: 1
identifier (dno,schedno): DNB098400971X (296.09014,296.09014)
DNB DDC notation (MAB2 field 700): {290,930}
DDC notation (MAB2 field 705): {296.09014,937.06}
calculated cutoff value: 95
title: Augustus und die Juden
title (remainder): Rechtsstellung und Interessenpolitik der
kleinasiatischen Diaspora
title (series): Studien zur Alten Geschichte ; Bd. 6
considered descriptor values: |18| {<100>-monika#schul[2], <331>-
juden[2945], <331>-augustus[384], <335>-kleinasiatischen[4], <335>-diaspora[904],
<335>-rechtsstellung[286], <335>-interessenpolitik[80], <412@410>-<033A>-verl.
antike@frankfurt[0], <451>-geschichte[20632], <451>-alten[2067], <540a>-3-938032-
16-2[0], <902g4>-anatolien[95], <912s1>-judentum[1176], <912s2>-
religionspolitik[73], <902g>-roemisches reich[988], <907s2>-rechtsprechung[757],
<907s1>-provinzialverwaltung[19], <907s1>-roemisches recht[116]}
matched descriptor values: |6| {kleinasiatischen, monika#schul,
provinzialverwaltung, interessenpolitik, religionspolitik, anatolien}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv): |5| {320, 935, 939.2, 955.03, 956.1015}
calculated1 ddc classes (sections): |4| {935, 939, 955, 956}
calculated1 ddc classes (main): |1| {900}
calculated2 ddc classes (subdiv): |3| {939.2[1], 955.03[1], 956.1015[1]}
calculated2 ddc classes (sections): {935[1], 939[1], 955[1], 956[1]}
calculated2 ddc classes (divisions): {930[2], 950[2]}
calculated2 ddc classes (main): {900[4]}
correlation (H0709_DNB098400971X,296.09014): 000.000 00x xxx xxx (0)

```

Abb. 81: Klassifizierungsergebnis von *in_dnb_DNB098400971X* mit *vc_dcl / vc_IDB* (Dezember 2008),

Der Titeldatensatz [DNB979445949](#) mit dem Titel „Kuenstlerischer Opportunismus in der Malerei und Plastik des Dritten Reiches“ wurde mit *dno_like_i*=750 („Malerei“), *dno_like_i*=730 („Plastik, Numismatik, Keramik, Metallkunst“), *dno_like_i*=943 („Geschichte Deutschlands“), *dno_i*=759.3 („Deutsche Malerei, ...“) und *dno_i*=943.086 („Zeit des Dritten Reichs, 1933–1945“) intellektuell klassifiziert. *vc_dcl* (April 2008) ermittelt *dno_a*=943, *CN*=0.5, wg. fehlerhaftem Vergleich mit 2. *dno_i*=943.086 (s. Fußnote 26). *vc_dcl / vc_IDB* (Dezember 2008) ermittelt mit „7“ beginnende DDC-Notationskandidaten, weshalb *CN*=1/4=0.25.



4. Weiterführende Betrachtungen und Fortführung

In diesem Kapitel werden einige der vorstehend beschriebenen Sachverhalte vertieft. Weiterhin werden die im Verlauf des Berichtes gewonnenen Erkenntnisse⁶⁶ zusammengefasst, die für die Fortführung der automatischen Klassifizierung relevant sind.

4.1 Klassifizierungsergebnisse mit *vc_dcl*-Programmänderungen

Wie oben an mehreren Stellen erwähnt, sind seit April 2008 (bis Dezember 2008) einzelne Verbesserungen und Erweiterungen an der Klassifizierungskomponente *vc_dcl*, der DDC-Datenbasis *vc_DB* und den Komponenten *vc_cdb* und *vc_cdb2*⁶⁷ vorgenommen worden. Die wichtigsten *vc_ds*⁶⁸-Programmänderungen VC1., ..., VC5. sind:

- VC1. Wenn ein DNB-Titeldatensatz mehrere intellektuelle DDC-Notationen (*dno_i*'s) enthält, findet die automatische Bewertung (Maß *CN*) zwischen der als erste vorkommenden DDC-Notation (*dno_i*)⁶⁹ und einem automatisch ermittelten DDC-Notationskandidaten (*dno_a*) statt.
- VC2. Aufnahme von Deskriptorwerten des Deskriptors „Weitere Sachtitel“ (MAB2-Feld 370) in die Repräsentation der automatisch zu klassifizierenden Titeldatensätze (Dezember 2008).
- VC3. Aufhebung der *schedno*-Begrenzung für *vc_IDB* (s. Fußnote 17), Berücksichtigung aller DDC-Notationen (*dno*'s). Dadurch vergrößert sich *vc_IDB* von 417MB (Oktober 2008) auf 702MB (November 2008).
- VC4. Um in *vc_IDB* schon vorhandene Titeldatensätze zu identifizieren, wurden über die DNB-*id*-Prüfungen (MAB2-Feld 026 („regionale Identifikationsnummer“)) hinaus weitergehende ISBN-Prüfungen (MAB2-Felder 004A, 004B, 004D, 005A) eingeführt. Falls Identifizierung erfolgreich, wird dies in der „*correlation*“-Zeile mit „!“ markiert (s. z. B. Abb. 82).
- VC5. Erweiterungen in der Ausgabe der Klassifizierungsergebnisse (ab Abbildungen Abb. 10ff), in Abb. 82 fettgedruckt wiedergegeben:

⁶⁶ U. a. aufgrund der (von der Autorin durchgeführten) intellektuellen Analyse einer erheblichen Anzahl von automatischen Klassifizierungsergebnissen; hauptsächlich in Kapitel 3 dokumentiert.

⁶⁷ *vzg colibri_create data base2*: Konvertierung vom MAB2-Format in die *vc_DB*-Repräsentation (vgl. auch [Reiner 2008c], Fußnote 20)

⁶⁸ *vc_ds*: *vzg colibri_dewey search system*, s. z. B. [Reiner 2007], S. 27, 37, 98 und [Reiner 2008b], S. 17.

⁶⁹ von oben nach unten



```

number of ddc-classified title:      330
id (title record in vc_DB!,schedno):DNB097533946X (343.49404,343)
DNB DDC notation (MAB2 field 700):  {340,330}
DDC notation (MAB2 field 705):    {343.49404,343.49407}
calculated cutoff value:           3972
title:                             Investitions- und Steuerstandort Schweiz
. . .
calculated2 ddc classes (sections): {330[1]}
. . .
correlation (dnb_ABH_DNB097533946X,343): 100.xxx xxx xxx xxx (0.333333) !

```

Abb. 82: Erweiterte Ausgabe eines Klassifizierungsergebnisses

Wie am Ende des Kapitels 2.1.4 (Fußnote 26) erwähnt, werden nun die Klassifizierungsergebnisse aufgrund der durchgeführten *vc_dcl*-Programmänderung VCI. vorgestellt. Die veränderten CN-Werte (*vc_dcl*, April-Oktober 2008 im Vergleich), bei der die 1. DDC-Notation aus MAB2-Feld berücksichtigt wird, sind in Abb. 83 wiedergegeben:

Grundgesamtheit <i>res_in_dnb_ABH</i> (24.642-(2.021 bzw. 2.019) (X))			
Maß CN (Kap. 2.1.1)	CN = 0	CN > 0	CN = 1
April 2008 (08-04-24)	36,06%	40,45%	23,49%
Oktober 2008 (08-10-18)	35,98%	40,78%	23,24%
Stichprobe <i>Sample I</i> (1.000)			
Maß CN (Kap. 2.1.1)	CN = 0	CN > 0	CN = 1
April 2008 (08-04-24)	45,4%	31,2%	23,4%
Oktober 2008 (08-10-18)	45,3%	31,2%	23,5%

Abb. 83: CN-Werte-Vergleich mit *vc_dcl*-Programmversionen vom April 2008 (letzte DDC-Notation berücksichtigt) und Oktober 2008 (erste DDC-Notation berücksichtigt)

Wie aus Abb. 83 zu erkennen ist, unterscheiden sich die Klassifizierungsergebnisse vom April 2008 und Oktober 2008 bzgl. der CN-Werte minimal voneinander. Der geringe CN-Unterschied liegt daran, dass lediglich 4,4% (1.089 von insgesamt 24.642) Klassifizierungsergebnisse betroffen sind, bei denen mehr als 2 (bis maximal 4) DDC-Notationen intellektuell vergeben wurden.

In Abb. 84 ist die mit dem Maß C (Kapitel 2.2.1) gemessene prozentuale Verteilung der Stellenübereinstimmungen zwischen intellektueller und automatischer Klassifizierung für die mit *vc_dcl* / *vc_IDB* (Dezember 2008) erzeugten Klassifizierungsergebnisse der Grundgesamtheit und *Sample I* wiedergegeben. Abb. 84 zeigt, daß – mit Maß C gemessen – in ca. 35% (Grundgesamtheit) bzw. ca. 44% (*Sample I*) keine Stelle und entsprechend in ca. 65% (Grundgesamtheit) bzw. ca. 56% (*Sample I*) wenigstens eine Stelle zwischen intellektueller und automatischer Klassifizierung übereinstimmen. Dieses Ergebnis wird im nächsten Kapitel 4.2 (Klassifizierungsergebnisse – erneut betrachtet) noch weitergehend untersucht.



C = 0	C = 1	C = 2	C = 3	C = 4	C = 5	C = 6	C = 7	C = 8	C = 9
Grundgesamtheit <i>res_in_dnb_ABH</i> (24.642)									
34,50%	14,31%	33,01%	8,24%	4,74%	2,33%	1,85%	0,67%	0,24%	0,11%
Stichprobe <i>Sample I</i> (1.000)									
44,4%	15,7%	24,6%	7,4%	3,3%	2,5%	1,4%	0,4%	0,2%	0,1%

Abb. 84: Mit C gemessene *dno_(like_)i-dno-a*-Übereinstimmungen (*vc_dcl* / *vc_IDB*, Dezember 2008)

4.2 Klassifizierungsergebnisse – erneut betrachtet

Die Klassifizierungsergebnisse aus dem Jahr 2008 werden erneut betrachtet. Die Bewertungen bilden die Referenz für die weiteren Klassifizierungstests. **1.** Die Ergebnisse basieren auf den Programmänderungen VCI.-VC5. **2.** Die Klassifizierungsergebnisse werden bereinigt. D. h., Ergebnisse, die sich auf Titeldatensätze bzw. deren Deskriptorwerte in *vc_IDB* beziehen, werden eliminiert. Dies betrifft 1.760 (Grundgesamtheit) bzw. 96 (*Sample I*) Klassifizierungsergebnisse von Titeldatensätzen, demnach 7,14% der Grundgesamtheit bzw. 9,6% der Stichprobe *Sample I*. Außerdem werden doppelte Klassifizierungsergebnisse (0,19% der Grundgesamtheit)⁷⁰ aus den Klassifizierungsergebnissen entfernt. **3.** Die Übereinstimmungen zwischen intellektueller und automatischer Klassifizierung werden getrennt für DDC-Notationen (*dno_i's*) und DDC-ähnliche-Notationen (*dno_like_i's*) für die Grundgesamtheit (*res_in_dnb_ABH-09-02-03*) und die Stichprobe (*res_sample I-09-02-03*) mit Maß *CN* (Abb. 86) und *C* (Abb. 87) berechnet.

<i>L_i = 1</i>	<i>L_i = 2</i>	<i>L_i = 3</i>	<i>L_i = 4</i>	<i>L_i = 5</i>	<i>L_i = 6</i>	<i>L_i = 7</i>	<i>L_i = 8</i>	<i>L_i = 9</i>	<i>L_i ≥ 10</i>
<i>dno_i's</i> <i>res_in_dnb_ABH</i> (15.374)									
0,07%	1,08%	10,02%	26,69%	24,31%	23,61%	8,76%	3,27%	1,33%	0,86%
<i>dno_i's</i> <i>res_sample I</i> (581)									
-	3,10%	14,11%	24,61%	24,96%	21,17%	7,06%	2,93%	1,72%	0,34%
<i>dno_like_i's</i> <i>res_in_dnb_ABH</i> (5.475)									
8,35	85,48	4,66	1,51						
<i>dno_like_i's</i> <i>res_sample I</i> (323)									
12,69%	74,61%	12,08%	0,62%	-	-	-	-	-	-

Abb. 85: Längenverteilungen der DDC-Notationen und DDC-ähnlichen Notationen (ohne in *vc_IDB* vorkommende Titeldatensätze und nur einmalig vorkommende DNB-Titeldatensätze)

⁷⁰ In *dnb_ABH* kommen 46 Titeldatensätze doppelt vor, z. B. DNB980568153 (beide in Datei A0745ti2.dis); DNB986051284 (in Dateien A0746ti2.dis und A0748ti2.dis); ZDB2060460-9 (in Dateien A0745ti2.dis und A0746ti2.dis); DNB986156078 (in Dateien A0748ti2.dis und B0748ti2.dis).



Aus Abb. 85 ist abzulesen, daß ca. 85% (Grundgesamtheit und *Sample1*) der von der DNB intellektuell vergebenen DDC-Notationen 3-stellig bis 6-stellig sind und ca. 85% (Grundgesamtheit) bzw. ca. 75% (*Sample1*) der von der DNB intellektuell vergebenen DDC-ähnlichen-Notationen 2-stellig sind. Demnach wäre für die automatische DDC-Klassifizierung ein anzustrebendes Ziel, 3- bis 6-stellige DDC-Notationskandidaten zu ermitteln.

Wie in Beispiel 24 (DNB2-4) im letzten Satz angesprochen, ist es einfach möglich, CN-Bewertungen für DDC-Notationen (*dno_i*'s) und DDC-ähnliche-Notationen (*dno_like_i*'s) separat vorzunehmen. Die Prozentzahlen des Maßes CN für die Klassifizierungsergebnisse der Grundgesamtheit (*res_in_dnb_AB*) und die der Stichprobe *Sample1* (*res_sample1*) sind in Abb. 86 wiedergegeben. In Klammern dahinter ist die Anzahl der betrachteten Klassifizierungsergebnisse (nach Bereinigung, s. Anfang des Kapitels unter 2.) enthalten:

Maß CN	CN = 0	CN > 0	CN = 1
<i>dno_i</i>'s			
<i>res_in_dnb_ABH</i> (15.374)	37,13%	50,42%	12,45%
<i>res_Sample1</i> (581)	47,16%	41,82%	11,02%
<i>dno_like_i</i>'s			
<i>res_in_dnb_ABH</i> (5.475)	33,95%	14,25%	51,80%
<i>res_Sample1</i> (323)	47,06%	10,22%	42,72%

Abb. 86: CN-Werte für *res_dnb_ABH* und *res_Sample1* nach *dno_i*'s und *dno_like_i*'s getrennt mit *vc_dcl* / *vc_IDB*, Dezember 2008 bestimmt (bereinigt nach Kap. 4.2, 2., d. h. ohne Klassifizierungsergebnisse a. der in *vc_IDB* vorkommenden Titeldatensätze, b. die schon ermittelt wurden (Entfernung doppelt vorkommender Ergebnisse wegen doppelt vorkommender DNB-Titeldatensätze), c. die in der *correlation*-Zeile mit „(X)“ markiert wurden (= keine korrekte intellektuelle DDC-Notation))

Aus Abb. 86 folgt: bzgl. des Maßes CN stimmen intellektuell vergebene DDC-Notationen mit automatisch ermittelten in ca. 63% (Grundgesamtheit) bzw. in ca. 53% (*Sample1*) in mindestens einer Stelle überein. Entsprechend stimmen intellektuell vergebene DDC-ähnliche-Notationen mit automatisch ermittelten DDC-Notationen in ca. 66% (Grundgesamtheit) bzw. ca. 53% (*Sample1*) in mindestens einer Stelle überein. Daraus folgt: der Unterschied von DDC-Notationen und DDC-ähnlichen Notationen mit automatisch ermittelten DDC-Notationen im Vergleich macht lediglich 3% aus. Weiter ist zu erkennen, daß *Sample1* bzgl. des Maßes CN einen Fehlerbereich von ca. 13% hat. Der Grund hierfür ist z. Zt. unklar. Weitere Testkollektionen (Grundgesamtheiten) und Stichproben können ev. Aufschluß darüber geben.

Als letzte Betrachtung zur automatischen Bewertung werden die C-Werte aus Abb. 84 nochmals neu berechnet (Bereinigung nach Kapitel 4.2., 2.) und in Abb. 87 dargestellt:



C = 0	C = 1	C = 2	C = 3	C = 4	C = 5	C = 6	C = 7	C = 8	C = 9
dno_i's res_in_dnb_ABH (15.374)									
37,13%	13,20%	25,89%	10,01%	6,53%	3,25%	2,61%	0,90%	0,33%	0,15%
dno_i's res_sample1 (581)									
47,16%	14,98%	15,66%	9,98%	4,99%	3,96%	2,24%	0,52%	0,34%	0,17%
dno_like_i's res_in_dnb_ABH (5.475)									
33,95%	17,60%	46,32%	1,54%	0,59%	-	-	-	-	-
dno_like_i's res_sample1 (323)									
47,06%	16,72%	33,74%	2,48%	-	-	-	-	-	-

Abb. 87: Mit Maß C gemessene Klassifizierungsergebnisse (bereinigt nach Kap. 4.2, 2.)

Aus Abb. 87 geht hervor, daß auch hier der Fehlerbereich bei ca. 10% liegt. Außerdem folgt: Der momentane Stand der automatischen DDC-Klassifizierung ergibt folgende Übereinstimmungen zwischen intellektueller und automatischer Klassifizierung: ca. 63% in mindestens der ersten Stelle, ca. 50% in mindestens den ersten beiden Stellen und ca. 24% in mindestens den ersten drei Stellen (Prozentzahlen jeweils aufgerundet).

Am Ende dieses Kapitels wird nun auf (aus Kapitel 3) offen gebliebene Feststellungen eingegangen.

DNB5-3 „... Anhaltspunkte für Ursachen des schlechten Ergebnisses lassen sich kaum erkennen“.

Was genau ist mit „schlechtem Ergebnis“ gemeint? Welche (Wunsch-)Vorstellungen herrschen bzgl. der Möglichkeiten eines automatischen Klassifizierungsverfahrens vor (**F54**)? Dazu gehört u. a. die Beantwortung der Frage, welche DDC-Notation ein Titeldatensatz erhalten soll, der bspw. nur den Deskriptorwert „Apfel“ oder „Hamster“ enthält.

DNB5-1 „...Eine exakte Übereinstimmung, d.h. alle Stellen der Colibri-Notationen stimmten mit allen Stellen der DNB-Notationen überein, hatten nur 5% der Titel.“ **DNB-K035**: „die unglaublich hohe Trofferquote von 20%, die so also keineswegs stimmt, ich schätze maximal 5-10%, mehr auf keinen Fall.“

Aus Abb. 86 folgt, daß 12,45% (191400/15374) Klassifizierungsergebnisse mit DDC-Notationen ermittelt werden, bei denen CN=1 ist.

DNB5-4 „... So ergab sich für die gesamte Datenbasis eine Hauptsachgruppentrefferquote von 39% durch Colibri.“



Aus Abb. 86 folgt, daß 51,80% (283600/5475) Klassifizierungsergebnisse mit DDC-ähnlichen-Notationen ermittelt werden, bei denen CN=1 ist. Entspricht dies der „Hauptsachgruppentrefferquote“ (**F55**)?

DNB5-3: „... Oftmals war sogar schwer nachvollziehbar, wie das Programm innerhalb der 5%-Marge zu den exakten Notationen kam.“

Gründe für eine exakte Übereinstimmung zwischen intellektueller und automatischer Klassifizierung können sein:

G1. Einschlägige Deskriptorwerte im Titeldatensatz enthalten.

G2. Mehrere Deskriptorwerte des Titeldatensatzes bestimmen gemeinsam (eindeutig) DDC-Notationskandidaten.

G3. Titeldatensätze mit ähnlichen Deskriptorwerten bzw. (sehr) ähnliche Titeldatensätze (= Dubletten?) sind in *vc_IDB* enthalten (vgl. DNB-K097 und z. B. bei DNB983972206; H_DNB984676171; DNB984691839; DNB974093424; DNB984489347; DNB985660910; DNB98471927X; DNB98528983X; DNB984647775; DNB984691154; DNB984487514; DNB98621552X)

G4. Titeldatensätze vom selben Autor sind mit anderem / ähnlichen Titel in *vc_IDB* enthalten

Es werden einige Beispiele aus *res_in_dnb_ABH-09-02-03* (Beispiel 30, ..., Beispiel 43) herausgegriffen und in Kurzform folgendermassen dargestellt:

Beispiel (Gründe):

Regionale Identifikationsnummer:

|...| Anzahl unterschiedlicher Deskriptorwerte, die zu DDC-Notationskandidaten führen können
 {Deskriptorwerte} Schlagwörter unterstrichen

|...| max. Anzahl der Deskriptorwerte => führt zu |...| Anzahl {DDC-Notationskandidaten}
dno_i unterstrichen, in Klammern Klassenbenennung

Beispiel 30 (G1., G2.):

DNB983312109

|9| {standard, publication, creation, internet, using, management, documents, components, digital}

|9| => |3| {005.1[1], 006.6[1], 025.04[1]}

(„Informationsspeicherungs- und Information-Retrieval-Systeme“)



Beispiel 31 (G1., G2.):

DNB979876680

|2| {psychologija, sigmund#freud}

|2| => |7| {131.34[1], 150.1952[1], 150.195[1], 150.19[1], 153.8[1], 155.232[1], 155.633[1]}
(„Freudsches System“)

Beispiel 32 (G1., G2.):

DNB975183621:

|3| {kindergottesdienste, kindergottesdienst, kirchenjahr}

|3| => {264.0083} („Kindergottesdienst“)

Beispiel 33 (G1., G2.):

DNB985303646

|2| {verhaeltniswahl, wahlssystem}

|2| => |7| {324.622[1], 324.63[1], 324.6[1], 324.9417[1], 324.993038[1], 328.334722[1], 328.3347[1]}
(„Verhältnswahl“)

Beispiel 34 (G1., G2.):

DNB983835772

|3| {lamaholot-sprache, <033A>-lincom europa@muenchen, lamaholot}

|3| => |1| {499.22[1]} („Balinesisch, . . .“)

Beispiel 35 (G1., G2.):

ZDB207562-3:

|3| {vorzeit, <033A>-schweizerbart@stuttgart, palaeobotanik}

|3| => |1| {560[1]} (dno_i=560.5) („Paläontologie Paläozoologie-- Fortlaufende Sammelwerke“)

Bemerkung: T1--05 („Fortlaufende Sammelwerke“) wird automatisch nicht ermittelbar sein.

Beispiel 36 (G1., G2.):

ZDB1449764-5:

|4| {bluttransfusion, immunoematology, immunhaematologie, transfusionsmedizin}

|4| => |1| {615.39[1]} („Produkte aus Humanblut und Blutersatzmittel“)

Beispiel 37 (G3. ppn_0341237493, id_OBV0AC00303928):

DNB550799222

|4| {maschinenelement, hermann#roloff, maschinenelemente, <033A>-vieweg@braunschweig}

|4| => |2| {621.815[1], 621.82[1]} („Maschinenelemente“)

Beispiel 38 (G1.):

DNB984162542

|1| {friedrich#schoenfelder}

|1| => |1| {792.028092[1]} (“Schauspieler, . . .”)Beispiel 39 (G1., G2.):

DNB984207996

|4| {opernregisseur, nabucco, d'hoffmann, opernproduktion}

|4| => |3| {782.1092[1], 782.12[1], 792.5023[1]} („Dramatische Vokalmusik Opern“)Beispiel 40 (G4. ppn_0118419951,id_DNB0920397816):

DNB98528983X

|3| {herbert#schnierle-lutz, hesses, calw}

|3| => |1| {838.91209[1]} („Autoren (Literatur)--deutsche Literatur--mehrere Gattungen--1900–1945, . . .“)Beispiel 41 (G1.):

DNB98406060X

|3| {marjorie, decolonization, kinnan}

|2| => |4| {811.52[1], 813.52[1], 818.5209[1], 823.914[1]} („Amerikanische Erzählprosa--1900–1945, . . .“)Beispiel 42 (G1.):

DNB981448763

|2| {eletrajza, skorzeny}

|1| => |5| {940.540092[1], 940.54009[1], 940.5421[1], 940.548143[1], 943.086092[1]} („Deutschland--1933–1945--Biografien, . . .“)Beispiel 43 (G1., G2.):

DNB985848421

|2| {israels, palaestinafrage}

|2| => |3| {956.9404[1], 956.94054[1], 956.9405[1]} („Geschichte Asiens des Orients des Fernen Ostens--Naher Osten (Mittlerer Osten)--Palästina Israel --1948--“)



4.3 Zusammenfassung der Ergebnisse und Fortführung

Die ausführliche Auseinandersetzung mit den Klassifizierungsergebnissen (in den Kapiteln 3 und 4) ist das Ergebnis umfangreicher zeitaufwendiger Daten- und Programm-Analysen der Autorin, um Erklärungen für die automatischen Klassifizierungsergebnisse zu liefern und mögliche Verbesserungsansätze zu finden.

Die Verbesserungsmöglichkeiten liegen – wie oben erläutert – in der Erhöhung der Qualität von *QIDB*, *QTIT*, *QELIM*, *QCUT*, *QCALC* und *QSIM*.

QIDB und *QTIT* muß vor allem für die Titeldatensätze verbessert (, d. h. die DDC-Datenbasis *vc_IDB* muß erweitert) werden, bei denen kein einziger Deskriptorwert des Titeldatensatzes in *vc_IDB* vorkommt (vgl. z. B. DNB-K071), erkennbar an folgender Ausgabe (Abb. 90):

```

number of ddc-classified title:      11982
identifier (dno,schedno):           DNB0983744475 (620.5043,620.5)
DNB DDC notation (MAB2 field 700):  {620}
DC notation (MAB2 field 705):       {620.5043}
calculated cutoff value:            0
title:                              Nano-Initiative - Action-Plan 2010
title (series):                    High-Tech-Strategy
considered descriptor values:        |4| {<331>-nano-initiative[0], <331>-action-
plan[0], <412@410>-<033A>-federal ministry education research public relation
div.@berlin[0], <451b>-high-tech-strategy[0]}
matched descriptor values:          |0| {}
max. match value of matched descriptor values: |0|
calculated1 ddc classes (subdiv):    |0| {}
calculated1 ddc classes (sections):  |0| {}
calculated1 ddc classes (main):      |0| {}
calculated2 ddc classes (subdiv):    |0| {}
calculated2 ddc classes (sections):  {}
calculated2 ddc classes (divisions): {}
calculated2 ddc classes (main):      {}
correlation (dnb_ABH_DNB0983744475,620.5): XXX.Xxx xxx xxx xxx (0)

```

Abb. 90 Klassifikationsergebnis eines Titeldatensatzes (*in_dnb_ DNB0983744475*), von dem kein einziger Deskriptorwert in *vc_IDB* vorkommt

QIDB kann zusätzlich u. a. durch weitergehende Transliteration von Deskriptorwerten (Fußnote 11), weitergehende Eliminierung von Sonderzeichen in Deskriptorwerten (Fußnote 12) und die Integration deutscher Deskriptorwerte (Fußnote 13) verbessert werden. Was die in *vc_IDB* vorhandenen Titeldatensätze (bzw. deren Deskriptorwerte) betrifft, so ist z. Zt. für die Autorin noch unklar, welche Titeldatensätze bzw. welche nicht in *vc_IDB* (prinzipiell) enthalten sind. Ev. müssen noch weitere (über Identifikationsnummern und ISBN hinausgehende) Prüfungen vorgenommen werden, um vorhandene Titeldatensätze in *vc_IDB* zu identifizieren. Die von der Deutschen Nationalbibliothek (DNB) im November 2007 zur Verfügung gestellten 25.653 Titeldatensätze (12 Wochen-/Monatslieferungen aus dem Jahr 2007) der Deutschen Nationalbibliografie der Reihen A, B und H



müssten (laut VZG-Protokoll) alle (!) in *vc_IDB* enthalten sein, da alle (!) bis spätestens Dezember 2007 in GVK „eingespielt“ wurden und für die Autorin ein „Abzug“ der GVK-DDC-Titeldatensätze im Januar 2008 erstellt wurde. Zu klärende Frage in der VZG (vgl. „liegen in der VZG vor“, DNB6-3, 1.): Wo verstecken sich die 92,86% (100% - 7,14%) DNB-Titeldatensätze (**F56**)? Die Beantwortung dieser Frage ist für weitere Tests wegen (der Korrektheit) der Testmethodologie wesentlich.

QTIT sollte auch für Titel verbessert werden, wenn die Anzahl der Deskriptorwerte zu gering (z. B. nur einer) und / oder die Deskriptorwerte nicht aussagekräftig sind, wie z. B. bei ZDB2167911-3: {salzburger}; DNB983529167: {geist}; ZDB2287942-0: {grupos}; DNB972057552: je einer von beiden Deskriptorwerten kommt in vielen DDC-Klassen vor: {phantasie, meditationen}; ZDB2389239-0: {epistemology} und ZDB1500201-9: {schriftenreihe}. Gute Deskriptorkombinationen sind z. B. DNB971901775: {„waermedaemmschichtsystemen“, „haftvermittler“}; DNB985281928: {feature, mapping, product, oriented, architecture, software, development}; DNB981454445: {„bpg“, „configuration“}; DNB0975230638: {roboterarm, taktile sensor, greifmanipulator}; DNB981978711: {aktienportefeuille, anlagepolitik}; DNB0984487514: {horvaths, oedoen} und DNB984837167: {nierenarterienstenose, nierentransplantation}.

Eine *QELIM*-Verbesserung (Fußnote 10) ist relativ einfach, muß allerdings mit Bedacht realisiert werden. Oftmals ist auch die Eliminierung nicht unbedingt notwendig, da irrelevante Deskriptorwerte auch über *cutoff_val_dyn* wegfallen können (vgl. Fußnote 21). Zu eliminieren sind z. B.: „10.“, „11.“, „28.“, „16.“, „11.“, „19.“, „20.“, ..., „chr.“, „editi“, „usum“, „gemaess“, „schriftenreihe“, „während“, „unvollendete“ (allerdings: Schubert's Symphonie Nr. 8!?), „habilitationsschrift“. Vielleicht sollte „unvollendete habilitationsschrift“ (vgl. DNB-K018) als Phrase entfernt werden. Vielleicht sollten nicht eliminiert werden: „deutsche“ bzw. „german“ in „deutsche malerei“ bzw. „german painting“ (759.3) (DNB979445949).

QCUT: Die heuristischen Regeln sind für die Qualität der Klassifizierungsergebnisse entscheidend, die Erweiterung / Verfeinerung der heuristischen Regeln muß kontrolliert erfolgen / getestet werden, z. B. ob H4. im Durchschnitt eine sinnvolle Erweiterung der heuristischen Funktion ist (vgl. DNB-K084).

QCALC: Es muß untersucht werden, welche Ergebnisausgaben im Allgemeinen die besten Ergebnisse liefern, d. h. *calc1*, *calc2*, oder ggf. eine andere (DNB-K006). Ist die automatische Bewertung auf der Basis von *shednos* ok (vgl. Fußnote 17 bzw. Maß S7) (**F57**)?

QSIM: Als zusätzliches Ähnlichkeitsmaß zum Vektorprodukt (vgl. Fußnote 22) wurde begonnen, die Hamming-Ähnlichkeit [<http://de.wikipedia.org/wiki/Hamming-%C3%84hnlichkeit>] zu implementieren (vgl. auch [Pfeffer 2007], S. 23; dort „Hamming-Abstand“ genannt).



Der vorliegende Bericht dient der weiteren gegenseitigen Klärung der Sachverhalte, um ein gemeinsames Verständnis zu entwickeln und eine Übereinkunft über Vorstellungen (vor allem über die Bewertung) zu erzielen. Es bietet sich der Aufbau von festen Testbeständen / Stichproben für unterschiedliche Fragestellungen an, anhand derer die automatischen Klassifizierungsverfahren getestet werden können. Was in diesem Bericht keinen Platz und keine Zeit mehr findet, aber u. a. noch zu untersuchen ist (Todo's):

T1: Ev. Schlagwörter mit <902>-, <912>, ... spezifizieren und Schlagwörter nur mit Schlagwörtern vergleichen.

T2: Spielt die Sprache der Deskriptorwerte (deutsche, englische oder anderssprachige), die in der *vc_IDB* enthalten sind, eine Rolle?

T3: Erstellung der prozentualen *CN*-Verteilungen bzgl.

- a) der Längen, bei denen *CN*=1 ist (Kap. 2.2.1),
- b) der DDC-Klassen
- c) der Reihen A, B, H
- d) mit / ohne Schlagwörter

T4: Vollständige Testklassifikation aller *GVK-DDC*-Titeldatensätze (DNB2-2, 1.)

Das Suchsystem *vc_ds* mit der Klassifizierungskomponente *vc_dcl* wird (wenn Zeit verfügbar) weiter verbessert und ergänzt, hier z. B. geplante (kleinere) Erweiterungen:

- VC6. In *correlation*-Zeile *dno* bzw. *dno_like* als Markierung einführen
- VC7. Angabe der Anzahl vor *calc2 (sections)*, *calc2 (divisions)* und *calc2 (main)*
- VC8. Korrektur: z. B. „779“ ist keine *calc1 (subdiv)* (Fußnote 56)
- VC9. Wieso fehlt *dno_a=940.54050947* in *calc2 (subdiv.)* (DNB-K087)?
- VC10. Ausschluss von stillgelegten DDC-Notationen (DNB-K040)
- VC11. Ggf. Anpassung / Verbesserung / Ergänzung / Implementierung der Bewertungsmaße (*C*, *CN*, *CN'*, *CS*, ...). Thematische Nähe wird durch *CP*, *CN* nicht erfaßt (vgl. Beispiel 7) bzw. Maß S5 („Correlated“), Kapitel 2.2.2

Wer – so wie die Autorin – von Seite 1 bis hierhin gekommen ist, ohne ein einziges (!) Zeichen zu überspringen, kann sich mit Kolibris messen:

„Kolibris haben auch eine beachtliche Ausdauer bei ihren Wanderbewegungen über Gebirge, Seen und das Meer“

[<http://freenet-homepage.de/jfeldhusen/Kolibris.htm>]