

Workshop zur Dewey-Dezimalklassifikation DDC-Einsichten und -Aussichten 2007 Göttingen

Automatische Analyse von DDC-Notationen und DDC-Klassifizierung von GVK-PLUS- Titeldatensätzen

Ulrike Reiner

Verbundzentrale des GBV (VZG), Göttingen

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



DDC-Einsichten in ...

VZG-Projekt Colibri/DDC

{ Fragestellungen, Beispiel,
Colibri-Modell / DDC-Suchsystem,
Analyse von DDC-Notationen,
DDC-Klassifizierung von Titeldatensätzen,
Ergebnisse }

... DDC-Aussichten

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

VZG-Projekt Colibri/DDC (1)



CO ntext generation and
LI nguistic tools for
B ibliographic
R etrieval
I nterfaces

000 – 020 – 510 – 511.3

VZG-Projekt Colibri/DDC (2)



Vorphase (2002 – 2004)

- 25.01.02 VZG-Colibri-Projektstart, parallel zum GUM-Projekt (GUM: Guardian-Unix-Migration)
- 28.01.02 WebDewey (VZG-Abonnement bei OCLC PICA)
- 21.03.02 DDC Workshop – Web Access to DDC21 (Birmingham)



OCLC PICA Birmingham
7th Floor, Tricorn House
51-53 Hagley Road
Edgbaston
Birmingham, B16 8TP
United Kingdom
Tel +44 121 456 4656
Fax: +44 121 456 4680
OCLC PICA Birmingham
[<http://oclc-pica.org/?id=1022&ln=uk>]

[<http://www.ceg.co.uk/investment-strategy/property-examples-details.asp?ContentID=12&ContentSectionID=9>]

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

VZG-Projekt Colibri/DDC (3)



Vorphase (2002 – 2004)

- ...
- 29.04.03 Beginn Colibri-Modell
- 26.05.03 Beginn automat. DDC-Klassifizierung (*ul-ddc*)
- 26.01.04 Eintreffen der 22. DDC-Ausgabe als XML-Datei (*e22*)
- 18.03.04 Beginn automat. DDC-Notationsanalyse (*ul-ana*)

Hauptphase (2004ff)

- 15.10.04 Erneuerung und Erweiterung:
Algorithmus, Systemarchitektur, Datenstruktur,
Programmcode und -namen (*vc_...* : *vzg colibri_...*) , ...

... nun ins Detail ...

VZG-Projekt Colibri/DDC (4)



„ 1.1 From Problems to Programs

Half the battle is knowing what problem to solve. When initially approached, most problems have no simple, precise specification. In fact, certain problems, such as creating a „gourmet“ recipe or preserving world peace, may be impossible to formulate in terms that admit of a computer solution. Even if we suspect our problem can be solved on a computer, there is usually considerable latitude in several problem parameters. Often it is only by experimentation that reasonable values for these parameters can be found.“

„... 1.4 Running Time of a Programm ... 12 Memory Management ...“

[Alfred V. Aho; John E. Hopcroft; Jeffrey D. Ullman: Data Structures and Algorithms. Addison-Wesley Publ. Company, Reading, Massachusetts, p. 1]

=> Problemspezifizierung und -präzisierung

VZG-Projekt Colibri/DDC (5)



Colibri-Fragestellungen

Ziel: Verbesserung der Suche nach Objekten in heterogenen, digitalen Bibliotheks- und Wissensbeständen

F1: Ist es möglich, eine inhaltlich stimmige DDC-Titelklassifikation aller GVK-PLUS-Titel automatisch zu erzielen?

F2: Ist es möglich, molekulare DDC-Notationen automatisch (eindeutig) zu zerlegen?⁽¹⁾ **F3:** Ist es möglich, mit Hilfe atomarer DDC-Notationen die Klassifikation und Suche zu verbessern? ⁽¹⁾

F4: Seien K1 und K2 zwei beliebige Klassifikationssysteme. Sind K1 und K2 vergleichbar bzw. kompatibel?

[(1) Songqiao Liu: The automatic decomposition of DDC synthesized numbers. Los Angeles, Calif., Univ. of California, Los Angeles, Diss., 1993, p. 18]

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (1) - Fragen



gegeben: GVK-PLUS-Titeldatensatz T1, Fragen f1, f2, f3

gesucht: Antworten ai1, ai2, ai3 (intellektuell ermittelt)

Antworten aa1, aa2, aa3 (automatisch ermittelt)

f1: Wie kann T1 DDC-klassifiziert werden? (*vc_dcl*)

f2: Wie kann eine zu T1 ermittelte DDC-Notation *dno* analysiert werden? (*vc_day*)

f3: Gibt es eine zu *dno* vergleichbare Notation der Basisklassifikation *bno*? (*vc_DB_PLUS*, *vc_dcl*)

Beispiel (2) – ad Information Retrieval



“ Information retrieval is a field concerned with the structure, analysis, organization, storage, searching, and retrieval of information. In recent⁽¹⁾ years, this whole subject has received an increasing amount of attention ... That there are substantial problems in the information field, everyone is agreed upon: More and more information is generated and put into circulation; the existing tools, classification schedules, and storage arrangements are often inadequate, ...; and it generally becomes more difficult and more expensive to get to know what one needs to know. ... important theoretical problems are unsolved: **What exactly is the content or meaning of a document?** **What are the linguistic devices used to carry meaning? To what extent can individual words, or word groups, in a text be said to carry and maintain a well-defined, controlled meaning?** **How can one isolate the content-bearing units if they exist?** And so on.”

(1) Unterstreichungen und Einfärbungen von ul

[Gerard Salton: Automatic Information Organization and Retrieval.
McGraw-Hill, Inc., New York u.a., 1968, pp. v; 1; 3]

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (3) – GVK-Titeldatensatz T1



- 001@** **aA** (?)
001A **02016:13-06-06** (Kennung und Datum der Ersterfassung)
001B **01999:17-06-06** **t15:09:55.000** (Kennung und Datum der letzten Änderung)
001D **02016:13-06-06** (Kennung und Datum der Statusänderung)
001X **02999** (?)
002@ **0Aox** (Bibliographische Gattung und Status)
003@ **0781835577** (Identifikationsnummer (PPN))
006Y **003603423** (Identifikationsnummer Crosskonkordanz)
007D **0sw060613** (Verlags-, Produktions- und Bestellnummer)
011@ **a2006** (Erscheinungsjahr)
021A **a@Spatial relational learning in rufous hummingbirds (Selasphorus rufus)** (Hauptsachtitel, Zusätze, Parallelsachtitel, Verfasserangabe)
027D **aAnimal cognition** **pBerlin** **nSpringer** **014359448** **z14365558** (Titel in Bandsätzen (für Anzeigen usw.))
028C/01 **dJonathan** **aHenderson** (Personenangabe (Ansetzungsform der Library of Congress))
028C/02 **dT.Andrew** **aHurly**
028C/03 **dSusan D.** **aHealy**
031A **d9** **j2006** **e3** **h201-206** **g6** (Differenzierende Angaben zur Quelle)
039B **cin** **9246822589** **8Animal cognition. - Berlin : Springer** (Verknüpfung zur größeren Einheit)
 x200600000090003799
045W **apsy** (SSG-Selektionskennzeichen (OLC-Ausschnitte))

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (4) -

Intellektuell ermittelte Antworten ai1, ai2, ai3



- Spatial relational learning in rufous hummingbirds (Titel)
(Selasphorus rufus) (Zimtkolibri, Synonyme: Fuchskolibri, Rotrücken-Zimtelfe)
- Animal cognition (Titel in Bandsätzen)
- psy (OLC-SSG Naturwissenschaften und Medizin - Psychologie)

ai1: intellektuelle DDC-Klassifizierung von T1:

591.513 Intelligence

ai2: intellektuelle DDC-Analyse der *dno591.513* :

Science (500), Animals (Zoology) (590), Specific topics in natural history of animals (591), Behavior (591.5), General topics (591.51), Intelligence (591.513)

ai3: intellektuell ermittelte – zu *dno591.513* vergleichbare - *bno*
BK 42.66 Ethologie

Beispiel (5) -

Automatisch ermittelte Antworten aa1, aa2, aa3 (1)



aa1: automatische DDC-Klassifizierung von T1 (in_ppn_0781835577)

a) ohne Titelanreicherung

{302, 302.35, 304.2}

b) mit Titelanreicherung durch Lexikoneintrag

(lex(animal_cognition,cognition animals)):

{574, 591.5, 591.51, 591.513, 591.59, 596, 599.05, 599.8}

302	Social interaction
302.35	Social interaction in complex groups
304.2	Human ecology
574	Internal biological processes and structures (571-575, [574])
591.5	Behavior
591.51	General topics in behavior
591.513	Intelligence
591.59	Communication
596	Chordata
599.05	Mammals--serials
599.8	Primates

Beispiel (6) -

Automatisch ermittelte Antworten aa1, aa2, aa3 (2)



aa2a: automatische Analyse der für T1 ermittelten *dno591.513*

591.513 <T1_dno4_to_analyze; length: 7>

5--.---- Science <hatzen>

59-.---- Animals (Zoology) <hatzen>

591---- Specific topics in natural history of animals <hat>

591.5-- Behavior <hat>

591.51- General topics <hat>

591.513 Intelligence <hat>

vc_day-Systemantwort (als *vc_daygram*)

vc_day **vzg colibri_ddc** number *analyzer*

vc_daygram **vzg colibri_ddc** *analysis* **diagram**

<*hat*> hierarchy at class (*e22* xml tag)

<*hatzen*> gebildet aus <*hat*> und <*zen*>: zen built entry (*e22* xml tag)

Beispiel (7) -

Automatisch ermittelte Antworten aa1, aa2, aa3 (3)



aa3a: automatisch ermittelte zu *dno* vergleichbare *bno*s

```
mysql> select dno,descr_val as bno from dno_db where dno = "591.513" and  
descr like "%045Q%";
```

```
+-----+-----+  
| dno    | bno                                     |  
+-----+-----+  
| 591.513 | 18.42 klassische griechische literatur |  
| 591.513 | 42.66 ethologie <biologie>           |  
| 591.513 | 17.98 textsammlungen                 |  
| 591.513 | 42.44 pflanzengeographie             |  
| 591.513 | 77.31 kognition                      |  
| 591.513 | 17.93 literarische stoffe            |  
| 591.513 | 08.21 griechische philosophie       |  
| 591.513 | 43.31 naturschutz                   |  
| 591.513 | 44.21 ernÑhrung <medizin>          |  
| 591.513 | 42.66 ethologie <biologie>         |  
+-----+-----+
```

```
10 rows in set (0.00 sec)
```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (8) -



Automatisch ermittelte Antworten aa1, aa2, aa3 (4)

aa3b: automatisch ermittelte zu *dno* vergleichbare *bno* unter Berücksichtigung der am häufigsten vorkommenden *bno*

```
mysql> select dno,descr_val as bno, count(*) from dno_db where dno="591.513" and
      descr like "%045Q%" group by bno having count(*) > 1;
```

dno	bno	count(*)
591.513	42.66 ethologie <biologie>	2

1 row in set (0.03 sec)

formuliert in IQL(1) (Erläuterung später im Colibri-Modell)

$I((\lambda x) \text{ dno_bno } (591.513, x) = \{42.66 \text{ ethologie}\})$

(1) In Anlehnung an IQL aus [Reiner, Ulrike: „Anfragesprachen für Informationssysteme. Reihe Informationswissenschaft der Deutschen Gesellschaft für Dokumentation (hrsg. von W. Neubauer, DGD). Frankfurt am Main, 1991, S. 79ff]

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (9) -

Automatisch ermittelte Antworten aa1, aa2, aa3 (5)



aa3c: automatisch ermittelte *dnos* zu zuvor ermittelten vergleichbaren *bno*-Klassen (*vc_dcl*-Systemantwort)

number of ddc-classified title: 1
ppn (dno): 000000000 (XXX)
calculated cutoff value: 78
title:
considered descriptor values: |2| {<021A>-tierpsychologie[11], <021A>-verhaltensforschung[78]}
matched descriptor values: |2| {tierpsychologie, verhaltensforschung}
max. match value: |2|
calculated ddc classes (subdiv): |5| {573.86, 591.5, 591.51, 591.513, 595.7}
calculated ddc classes (sections): |3| {573, 591, 595}
calculated ddc classes (main): |1| {500}

vc_dcl *vzg colibri_ddc classification*

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (10) -

DDC-Klassifizierung eines (schon) DDC-klassifizierten GVK-Titeldatensatzes (in *vc_DB* enthalten) (*vc_dcl*-Systemantwort)



```
number of ddc-classified title: 2
ppn (dno): 0024246379 (594.58)
calculated cutoff value: 1
title: The @behavior and natural history of the Caribbean reef squid sepioteuthis
sepioidea dwith a consideration of social, signal, and defensive patterns, for
difficult and dangerous environments hMartin Moynihan; Arcadio F. Rodaniche
considered descriptor values: |30| {<028A_da>-martin#moynihan[4],
<028B>-arcadio f.#rodaniche[1], <021A>-caribbean[691], <021A>-reef[117],
<021A>-squid[24], <021A>-sepioteuthis[1], <021A>-sepioidea[1],
<021A>-behavior[1575], <021A>-natural[2377], <021A>-history[14891],
<021A>-defensive[85], <021A>-patterns[1664], <021A>-difficult[346],
<021A>-dangerous[548], <021A>-environments[687], <021A>-consideration[203],
<021A>-social[4233], <021A>-signal[292], <033A_n>-<033A_n>-parey[65],
<044A_a>-squids[13], <044A_x>-behavior[1575], <045A>-QL750[12],
<045Q/01>-42.73 mollusca[26], <145Z_a>-594.582.1[1],
<244Z_8>-korallentintenfisch: karibisches meer: verhalten[1],
<245Z_a>-182.278.91[1], <245Z_a>-vrs[9], <245Z_a>-vqw[36],
<245Z_8>-bid 770.800[3], <245Z_8>-bid[467]}
matched descriptor values |6| {korallentintenfisch: karibisches meer:
verhalten, 594.582.1, 182.278.91, sepioteuthis, sepioidea,
arcadio f.#rodaniche}
max. match value: |7|
calculated ddc classes |1| {594.58}
calculated ddc classes (sections): |1| {594}
calculated ddc classes (main): |1| {500}
http://www.gbv.de/
```

Beispiel (11) -



aa2b_en: Analyse der DDC-Notation 333.958764 (1)

```

333.958764 <trochili_1_to_analyze; length: 10>
3----- Social sciences <hatzen>
33----- Economics <hatzen>
333----- Economics of land and energy <hat>
333.9----- Other natural resources <hat>
333.95---- Biological resources <hat>
333.958--- Birds <hat>
333.9587-- Specific kinds of birds <hatspan:333.9583-333.9589>
333.958764 Hummingbirds--resource economics <hatien>
333.95---- Biological resources <nfa->333.7-333.9>
333.958--- Birds <nfa->333.7-333.9>
---.---7-- Miscellaneous orders of land birds <nalr1span:333.9583-333.9589:598.7>
---.---76- Apodiformes <nalr1span:333.9583-333.9589:598.76>
---.---764 Trochili (Hummingbirds) <nalr1span:333.9583-333.9589:598.764>

```

vc_daygram für **333.958764** in Englisch (lang=„en“)

<hatien> gebildet aus *<hat>* und *<ien>*: built schedule entry (main tag)
(*e22* xml tag)

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Beispiel (12) -

aa2b_de: Analyse der DDC-Notation 333.958764 (2)



```

333.958764 <trochili_1_to_analyze; length: 10>
3----- Sozialwissenschaften <hatzen>
33----- Wirtschaft <hatzen>
333----- Boden- und Energiewirtschaft <hat>
333.9----- Andere natürliche Ressourcen <hat>
333.95---- Biologische Ressourcen <hat>
333.958--- Vögel <hat>
333.9587-- Einzelne Vogelarten <hatspan:333.9583-333.9589>
333.958764 Kolibris Ressourcenökonomie <hatien>
333.95---- Biologische Ressourcen <nfa->333.7-333.9>
333.958--- Vögel <nfa->333.7-333.9>
---.---7-- Verschiedene Ordnungen von Landvögeln <nalrlspan:333.9583-333.9589:598.7>
---.---76- Apodiformes (Seglervögel) <nalrlspan:333.9583-333.9589:598.76>
---.---764 Trochiliformes (Kolibris) <nalrlspan:333.9583-333.9589:598.764>

```

vc_daygram für **dno333.958764** in Deutsch (lang=„de“)

Modell (1)



Atomare DDC-Notation (*dno_atom*)

Semantisch kleinste Einheit, die eine DDC-Klasse repräsentiert

dno_atom: 3 (Social sciences), ...

dno_atom: 333.958 (Birds), ...

dno_atom: 333.958764 (Hummingbirds--resource economics), ...

dno_atom: 598.764 (Trochili (Hummingbirds))

Molekulare DDC-Notation (*dno_mol*) (Synthetische DDC-Notation)

Nach Regeln (Synthese-Anweisung) zusammengesetzte DDC-Notation

dno_mol: 333.958764

333.95-<nfa>-rule

Add as instructed under 333.7-333.9 (1)

333.958-<nfa>-rule

Add as instructed under 333.7-333.9 (1)

333.9583-333.9589-<na1r1span>-rule

Add to base number 333.958 the numbers following 598 in 598.3-598.9, ... then add further as instructed under 333.7-333.9

(1) add table (Anhängetafel)

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Modell (2)



dno_mol 333.958764

DNB 5400 (PICA3) := 045F (PICA+)

GBV 5400 (PICA3) := 045H (PICA+)

Ausgezeichnete *dno_atoms*

- 1. base number *bano*** (Grundnotation) 045H/01
333.958, 598.07234, 82, 83, T1--080, T3C--382
- 2. prefix *dno predno*** (Ausgangsnotation) *vc_dayset_mab2* -
- 3. *dno_atom* beginnend mit 0, 1, ..., 9**
(Notation aus Haupttafel) 333.958 045H/02
- 4. *dno_atom* beginnend mit T1--, ..., T3A--, ..., T3C--, T4--, ..., T6--** (Notation aus Hilfstafel) T1--08 045H/03
- 5. wie in 3. oder 4. gefolgt von „+“ Ziffernfolge** (*vc_KB-Repräsentation*) 045H/04
598+09 (Historical, geographic, persons treatment)
823+914 (1945-1999), 833+914 (1945-1990), 833+92(1990-)
T1--08+01 (Forecasting and forecasts)
T1--08+074 (Museums, collections, exhibits)

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



Deskriptor (*descr*) (1)

PICA-Kategorie(*), deren Werte zur inhaltlichen Charakterisierung und/oder zur automatischen Klassifizierung von GVK-PLUS-Titeldatensätzen dienen:

PICA3	PICA+	Bedeutung
3000	028A	1. Verfasser
300x	028B	2. und weitere Verfasser
301x/2x	028C/D	Sonstige beteiligte Personen
303x	028E	Interpreten
304x	028F	Gefeierte Personen
305x	028G	Sonstige nichtbeteiligte Personen
4000	021A	Hauptsachtitel, Zusätze, Parallelsachtitel, Verfasserangabe
4222	046M	Angaben zum Inhalt
4241	039B	Verknüpfung zur größeren Einheit
5010	045F	DCC (Dewey Decimal Classification)
5020	045B/00	ASB (Allgemeine Systematik für Bibliotheken)
5021	045B/01	SSD (Systematik der Stadtbibliothek Duisburg)
5022	045B/02	SfB (Systematik für Bibliotheken)
5023	045B/03	KAB (Klassifikation für Allgemeinbibliotheken)
5024	045B/04	Systematik der ekz
5030	045A	LCC (Library of Congress Classification)
5040	045C	NLM (National Library of Medicine)

[(*) <http://www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/02Richtlinien/01KatRicht/update22.pdf>]

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



Deskriptor (*descr*) (2)

PICA3	PICA+	Bedeutung
5050	045E	Sachgruppen Der Deutschen Bibliothek
5051	045G	DDC-Sachgruppen Der Deutschen Bibliothek
5075	045T	NZN-Notation
5080	045U	ZDB-Notation
5100	041A/00	1. Kettenglied der 1. RSWK-Kette
5101	041A/01	2. Kettenglied der 1. RSWK-Kette
530x	045Q	Basisklassifikation
531x	045R	Zeitschlagwort (Jahreszahlen)
5500	044A	LoC subject headings
5510	044C	Medical subject headings
5520	044E	PRECIS
5530	044F	DDB-Schlagwörter bis 1986
5540	044G	British Library subject headings
555x	044K	Einzelschlagwort
5570	044S	Gattungsbegriffe bei alten Drucken
558x	044L	Einzelschlagwort (Projekte)
5950	020G	Angaben aus Inhaltsverzeichnissen
60xx	145Z	Lokale Notationen
65xx	144Z	Lokale Schlagwörter
67xx	245Z	Lokale Notationen
68xx	244Z	Lokale Schlagwörter

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Modell (5)



Deskriptorwert (*descr_val*)

Wert einer PICA-Kategorie

hummingbirds

Titeldatensatz *T*

Menge von Deskriptorwerten (von PICA-Kategorien)

in_ppn_0781835577 := {<028C>-t.andrew#hurly, ... , <021A>-learning, ... ,
<021A>- rufous, ... , <027D>-animal_cognition, ... , <039B>-cognition}

DDC-Klasse *D*

Menge von Deskriptorwerten (von PICA-Kategorien)

598.07 := {<028A_da>henry e.#childs, ... , <021A>birds, ...
<021A>california, ... , <045Q/01>42.83 aves, ... }

598.07234092 := {<021A>naturalists, ... , <044A_s>bird watching, ...
<044G_a>ornithology, ... , <044K>vogelbeobachtung, <045A>QL677.5, ... }

DDC-Basis *DB*

Menge von DDC-Klassen

{000, ... , 598.07, ... , 598.07234092, ... 999 }



DDC-Klassifizierung (eines GVK-PLUS-Titeldatensatzes)

Vektorkorrelationsmaß (Skalarprodukt) : $\alpha_{dt} = \sum_{i=1}^l d_i t_i$

Ein Titeldatensatz wird mit der (den) DDC-Klasse(n) klassifiziert, bei der (den) der größte Ähnlichkeitswert α_{\max} zwischen T und $D \in DB$ erzielt wird (werden)

Beispiel zur DDC-Klassifizierung

geg.: Modell-Datenbasis:

* Datenbasis DB bestehe aus den vier DDC-Klassen

508 := {A,C,D,E,F} 598 := {A,...,F} 736.4 := {A,F} 823.914 := {A,B,C,D,F}

* Die Deskriptorwerte $descr_vals$ {A,B, ... F} bedeuten:

A := hummingbirds B := feeding C := flight
D := breeding E := ecological F := system

* Drei Titeldatensätze T1, T2 und T3 seien wie folgt indexiert:

T1 := {A,B,E,F} T2 := {D,E} T3 := {A}

ges.: DDC-Klassifizierung der Titeldatensätze T1, T2 und T3



Klassen-Deskriptorwert-Ähnlichkeitsmatrix

Datenbasis DB	A	B	C	D	E	F
508 (Natural history)	1	0	1	1	1	1
598 (Aves (Birds))	1	1	1	1	1	1
736.4 (Wood)	1	0	0	0	0	1
823.914 (English fiction—1945-1999,...)	1	1	1	1	0	1
T1	1	1	0	0	1	1
T2	0	0	0	1	1	0
T3	1	0	0	0	0	0

$$\alpha(508, T1) = 3; \alpha_{\max}(598, T1) = 4; \alpha(736.4, T1) = 2; \alpha(823.914, T1) = 3;$$

=> nach Modell wird klassifiziert:

T1 mit {598}; T2 mit {508,598}; T3 mit {508,598,736.4,823.914}

Modell (8) / DDC-Suchsystem (1)

DDC-Suchsystem *vc_ds*

DDC-basierte
Anfragen →



→ DDC-basierte
Antworten

vc_ds: *vzg colibri_dewey search system*

Modell (9)



Beispiele für

DDC-basierte Anfragen

DDC-Notation (*dno_mol*)

DDC-Notation (*dno_atom*)

DDC-Klassenbenennung (*cap*)

DDC-Notation (*dno*)

RVK-Notation (*rno*)

LoC Subject Headings (*lcsh*)

DDC-Notation (*dno*)

unklassif. Titeldatensatz (*tit*)

allgemein:

DDC-Notation (*dno*)

O_i (i = 1, ..., n) : SuchObjekt, z. B. DDC-Notation, DDC-Klassenbenennung, BK-Notation, GVK-Titeldatensatz, Internetquelle, ...

DDC-basierte Antworten

-> {*dno_atom_1*, ..., *dno_atom_n*}

-> DDC-Klassenbenennung {*cap*}

-> DDC-Notation {*dno_atom*}

-> BK-Notation {*bno*}

-> DDC-Notation {*dno*}

-> DDC-Notation {*dno*}

-> British Library Subject Headings {*blsh*}

-> {*dno_1*, ..., *dno_n*}

-> {*O_1*, ..., *O_n*}



Anfrageklassen (4)

4a. Welche DDC-Notation(en) dno (dnos) treten gemeinsam mit

1. Library of Congress Subject Heading *lcs*h,
 2. British Library Subject Heading *bls*h,
 3. Relativ Index term *ri*,
 4. Einzelschlagwort(1) *sh*,
- ... auf?

(1) 555x (PICA3) bzw. 044K (PICA+)

IQL: $I((\lambda x) \text{ dno_lcs}h(x, \text{hummingbirds})) = \{598.764\}$

IQL-T₍₁₎: $I((\lambda x) \text{ dno_sh}(x, s^{\sim, \text{segler}})) = \{598.762\}$

mysql: `select * from dno_db where descr="<044K>" and descr_val like "%segler%";`

```

+-----+-----+-----+
| dno    | descr  | descr_val          |
+-----+-----+-----+
| 598.762 | <044K> | baumsegler <familie> |
| 598.762 | <044K> | segler             |
+-----+-----+-----+
2 rows in set (30.28 sec)

```

(1) <http://taipan.dyndns.org/~ul/freiall.pdf>

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Modell (16)



Anfrageklassen (5)

- 4b. Welches Schlagwort (welche Schlagwörter), Begriff (Begriffe)
1. *lcs*, 2. *bls*, 3. *ri*, 4. *sh*, ... tritt (treten) gemeinsam mit
der DDC-Klassenbenennung *cap* auf?

IQL: $I((\lambda x) \text{ cap_bls}(\text{hummingbirds}, x)) = \{\text{behavior, habitat, north america, physiology, pictorial works, hummingbirds}\}$

SQL: `select d1.dno as dno, d1.descr as dno_descr, d1.descr_val as dno_descr_val, d2.descr as bno_descr, d2.descr_val as bno_descr_val from dno_db d1, dno_db d2 where d1.descr_val like "%hummingbirds%" and d1.descr = "<hat>" and d1.dno = d2.dno and d2.descr like "<044G%";`

Modell (17)



Anfrageklassen (6)

5. Mit welcher (welchen) DDC-Notation(en) kann der Titeldatensatz *tit* DDC-klassifiziert werden? (vgl. Beispiel aus Modell(7))

$I((\lambda x) \text{ dno_tit}(x, T1)) =$

$I((\lambda x) \text{ dno_tit}(x, (\text{hummingbirds}, \text{feeding}, \text{ecological}, \text{system})))$
 $= \{598\}$

$I((\lambda x) \text{ dno_tit}(x, T2)) =$

$I((\lambda x) \text{ dno_tit}(x, (\text{breeding}, \text{ecological}))) = \{508, 598\}$

$I((\lambda x) \text{ dno_tit}(x, T3)) =$

$I((\lambda x) \text{ dno_tit}(x, (\text{hummingbirds})))$
 $= \{508, 598, 736.4, 823.914\}$

DDC-Suchsystem (2)



DDC-Suchsystem *vc_ds* mit den Hauptkomponenten

- * Analysekomponente *vc_day*
- * Klassifizierungskomponente *vc_dcl*
- * Wissensbasis *vc_KB*
- * Datenbasis *vc_DB*
- * invertierte Datenbasis *vc_IDB*

für die

- * Analyse von DDC-Notationen
- * DDC-Klassifizierung von Titeldatensätzen

Analyse von DDC-Notationen (1)



Stand der Technik (Zerlegung von DDC-/UDK-Notationen)

- [Liu1993]** Songqiao Liu: *The automatic decomposition of DDC synthesized numbers*. Los Angeles, Calif., Univ. of California, Los Angeles, Diss., 1993.
- [Riesthuis1997]** Gerhard J. A. Riesthuis: *Decomposition of complex UDC notations*. In: Knowledge Organization for Information Retrieval. Proceedings of the Sixth International Study Conference on Classification Research. University College London, 16-18 June 1997. International Federation for Information and Documentation (FID), The Hague, Netherlands, 1997, pp. 139-143.
- [Reiner2005]** Ulrike Reiner: *DDC-Notationsanalyse und –synthese*. Sept. 2004 - Feb. 2005. VZG-Colibri-Bericht 2/2004. Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG des GBV), Göttingen, 2005.

Analyse von DDC-Notationen (2)



Riesthuis1997 (Decomposition of complex UDC notations)

- * **complex notation: connectors (':', '::', '+')**
simple notation: splitted complex notation

- * **Beispiele von UDC-Notationen (p. 141):**

323.1/.2(47+57):81'01(47+57)

**National and ethnic minorities. Internal political activities.
Origins (of languages). Soviet Union**

327(44)(092)Talleyrand(044.4)

**International affairs. France. Biographical presentation. Talleyrand.
Official correspondence**

611.018.1:616-073.572-091.8:613.221(043)

**Cytology. Diagnosis by methods adopted from physics:
electrostatics. Cytopathology. Nutrition of infants. Dissertations**

- * **In natürlicher Sprache skizzierte Algorithmen (pp.141-143):**



Arbeitsschritt 1 (Syntheseanweisungen)

Liu1993

L1: study and analysis of the rules for synthesizing DDC numbers

L1: 17 rules for decomposition (class 700, tables and internal tables)

Reiner2005

R1: Studium und Analyse der Syntheseanweisungen, Auszug aus e22:

```

...
<hat><lnk refC="ddc::300" refR="ddc:333.958" span1="333.958"><txt>333.958</txt></lnk>
  <txt>Birds</txt></hat>
...
<nfa><txt>*Add as instructed under </txt><lnk refC="ddc::300" refR="ddc:333.7-.9"
  span1="333.7" span2="333.9"><txt>333.7-333.9</txt></lnk><txt>.</txt></nfa>
...
<hat><lnk refC="ddc::300" refR="ddc:333.7-.9" span1="333.7" span2="333.9"><txt>333.7-333.9
  </txt></lnk><txt>Natural resources and energy</txt></hat>
...
<ba5><txt>Except for modifications shown under specific entries, add to each subdivision identified
  by * as follows:</txt></ba5>
...
<ren><txt>11</txt></ren>
<reh><txt>Reserves</txt></reh>
...

```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (4)



Arbeitsschritt 1 (Syntheseanweisungen)

Reiner2005

R1: Klassifizierung der e22-DDC-Syntheseanweisungen, Beispiele:

- <na1> Add to base number ... the numbers following ... in ...
- <na5> Add to each subdivision identified by ... as instructed under ...
- <na6> Add notation ... from table under
- <nad> Add to ... the historical period numbers following ... that appear in subdivisions of ...
- <nfa> Add [to base number|historical periods] as instructed under ...
- <nfu> Use notation ... from Table 1 as modified at ...
- <ba4> Add to ... notation from the period table ... then add further as follows, ...

R1: Bestimmung der Anzahl der e22-DDC-Syntheseanweisungen (7.705)

```
colibri/ul-test> egrep -i "add to base number.*(notation|the numbers following.*[notation]*)|  
  add to each subdivision identified by.*(as follows|notation|the numbers following|as instructed  
  (at|under))|(add [to]*|add to base number).*as instructed (at|under)|(add.*notation.* from.*  
  table.*[under]*)|add [to]*.*(the [historical period|period division]*numbers following|notation)  
  " e22 | wc -l  
7705
```

R1: 25 DDC-Regeln für *vc_day* (vzg colibri_ddc number analyzer)

Analyse von DDC-Notationen (5)



Arbeitsschritt 1 (Syntheseanweisungen)

R1: 25 DDC-Regeln für *vc_day* (vzg colibri_ddc number analyzer)

r0 = "[Aa]dd to base number.*three-digit numbers"	# <na2>
r1 = "[Aa]dd to base number.*the numbers following.*in"	# <ba[13]>,<na[134]>
r2 = "[Aa]dd to.*notation"	# <ba4>,<na[246]>
r3 = "[Aa]dd to each subdivision identified by"	# <ba5>,<na5>
r4 = "told to add as instructed under"	# <ba5>
r5 = "[Aa]dd to.*the.*numbers following.*in"	# <na[13]>,<nad>
r6 = "[Aa]dd the numbers following.*in"	# <na1>
r7 = "[Aa]dd to base number.*notation"	# <na2>
r8 = "[Aa]dd to.*as instructed under"	# <na6>
r9 = "[Aa]dd as instructed under"	# <na6>,<nfa>
r10 = "[Aa]s modified under"	# <nst>
r11 = "before adding"	# <ba1>,<mt>
r12 = "[Aa]dd notation.*from table under"	# <na6>
r13 = "[Aa]dd to base number.*as instructed at beginning of"	# <na6>,<nfa>
r14 = "then.*add further as instructed"	# nasi16, nasi26, nasi34, nasi36, nasi46
r15 = "then.*add.*0.*and.*add notation"	# nasi24, nasi34, nasi44
r16 = "then.*add.*[01].*and to the result add.*the numbers following.*in"	# nasi31, nasi41, na1nad
r17 = "then to the result add the numbers following.*in"	# nasi31, nasi41
r18 = "then add.*0 and to the result add further as instructed"	# nasi46
r19 = "[Aa]dd to base number as instructed under"	# <nfa>
r20 = "Add historical periods as instructed under"	# <nfa>
r21 = "then.*add notation"	# <nad>
r22 = "then.*add standard sub"	# <nad>
r23 = "then.*add historical period numbers"	# <nad>
r24 = "[Aa]dd.*0.*to base number"	# <nad>
r25 = "[Uu]se notation.*under each subdivision identified"	# <nsi>

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (6)



Arbeitsschritt 2 (Stichprobe)

Liu1993

L2: Erstellung der Stichprobe (Umfang: 600)

Reiner2005

R2: Verwendung von Liu's Stichprobe.

Erstellung als Eingabedatei *in_liu_t* für Analysekomponente *vc_day* :

```
-----  
in_liu_1  
700.23  
in_liu_2  
700.90440747471  
...  
in_liu_600  
799.26092  
-----
```

**von den 600 DDC-Notationen von Liu's Stichprobe sind (nach 8 Jahren)
36 DDC-Notationen „veraltet“ (stillgelegte oder verlegte DDC-Notationen)**

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



Arbeitsschritt 2 (Stichprobe)

L2: Umfang der Stichprobe (sample size for desired margin of error) (1)

$$n = (2z^*/w)^2 p^* (1-p^*)$$

w : Fehlerbereich (margin of error), auch Genauigkeit

p* : geschätzter Anteil zerlegbarer DDC-Notationen

z* : tabellierte statistische Funktion

Für **p* = 0.5**

$$n = (z^*/2w)^2$$

Bei Konfidenzniveau 95% (z* = 1.96) und Fehlerbereich $-0.04 \leq w \leq 0.04$ (Genauigkeit 4%) => n = (1.96/.08)² = 600 (Stichprobenumfang)

[(1) David S. Moore; George P. McCabe: Introduction to the Practice of Statistics. 3rd ed., W. H. Freeman and Company, New York, 1998, p. 593]

Analyse von DDC-Notationen (8)



Arbeitsschritt 2 (Stichprobe)

L2: Recherche im OCLC Union Catalog (EPIC, telecommunication service)

gesucht mit: **700?, 701?, ..., 700.1?, 700.2?, ...**

gefunden: **11.662(*) bibliographische OCLC-Datensätze**

in Datenbank abgespeichert: **1.743 bibliographische OCLC-Datensätze**

(auf vollständiger, 20. DDC-Ausgabe beruhende, einmalig vorkommende, synthetische, von der LoC vergebenene DDC-Notationen)

L2: Grundgesamtheit (1.743-42) = 1.701 DDC-Notationen

(42 inkorrekt gebildete DDC-Notationen)

L2: Stichprobe: per Zufallsgenerator EPISTAT 600 DDC-Notation

bestimmt (EPISTAT: public-domain compute statistical package)

(*) zum Vergleich: 287.704 bibliographische GVK-Datensätze der Klasse 700 (Stand: Juli 2004)

Analyse von DDC-Notationen (9)



Arbeitsschritt 3 (DDC-Wissensbasis)

Liu1993

L3: Erstellung der DDC Datenbank/Wissensbank (OCLC Daten, 20th Ed. of DDC)

Reiner2005

R3: Erstellung der DDC-Wissensbasis *vc_KB*(1) mit Programm *vc_ckb*(2) aus *e22* (22nd Ed. of DDC). Wissensrepräsentation von Fakten und Regeln:

dno	ÿ(3)	descr	ÿ descr_val
333.958	ÿ	<hat>	ÿ Birds
333.958	ÿ	<nfa>	ÿ 333.7-333.9
333.7+11	ÿ	<ba5span:333.7-333.9+11>	ÿ Reserves

Bedeutung

Fakt: Die Klassenbenennung von „333.958“ ist „Birds“

nfa-Regel: Wenn es sich um die DDC-Notation „333.958“ handelt, dann ist es erlaubt, DDC-Notationsteile - wie unter 333.7-333.9 angegeben - hinzuzufügen

Fakt: Die Klassenbenennung von „333.7+11“ ist „Reserves“

(1) *vzg colibri_ddc Knowledge Base*; (2) *vzg colibri_create ddc knowledge base*

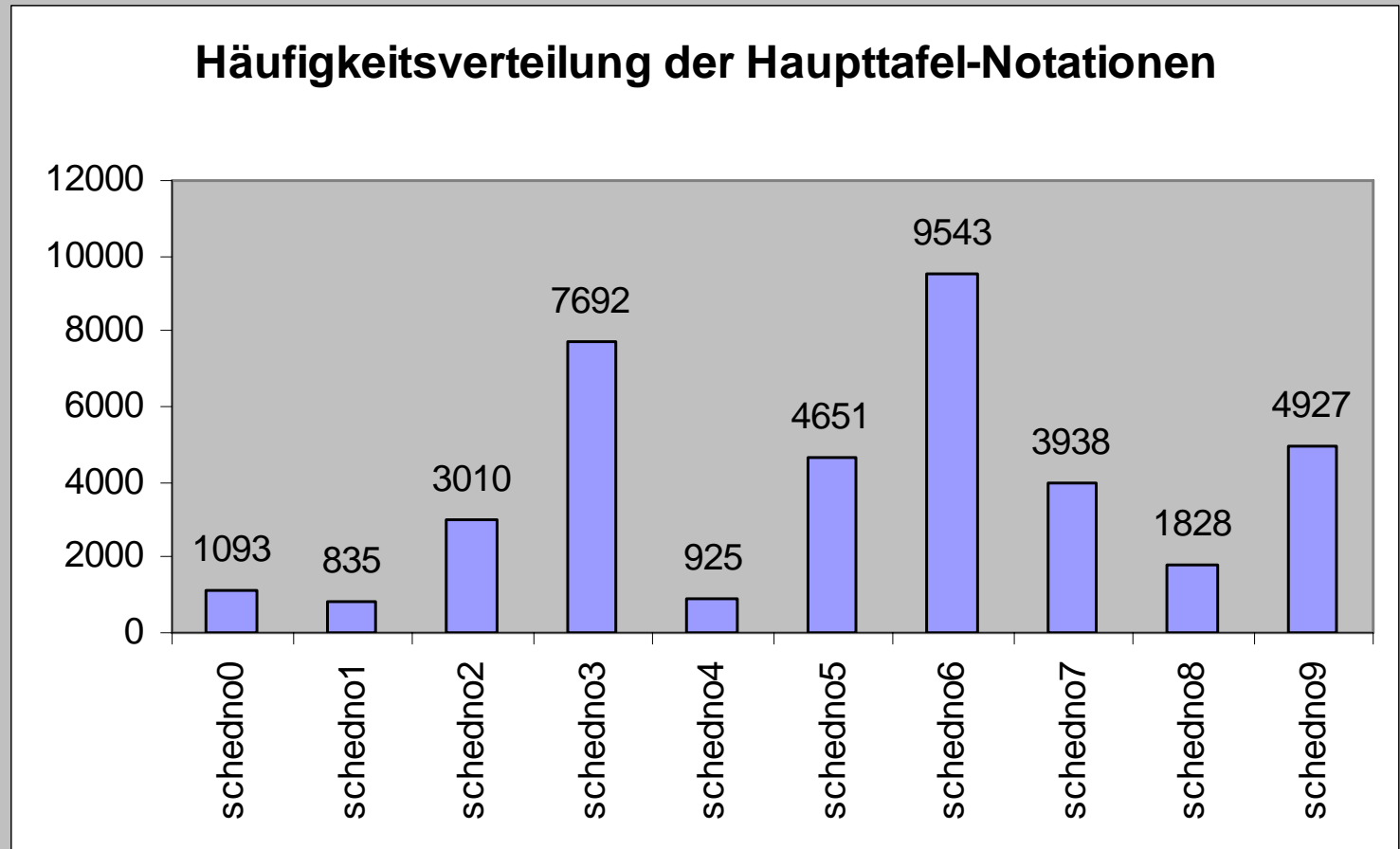
(3) ÿ: florin, zur Trennung von Datenfeldern

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



Arbeitsschritt 3 (DDC-Wissensbasis)

R3: DDC-Wissensbasis *vc_KB*



Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

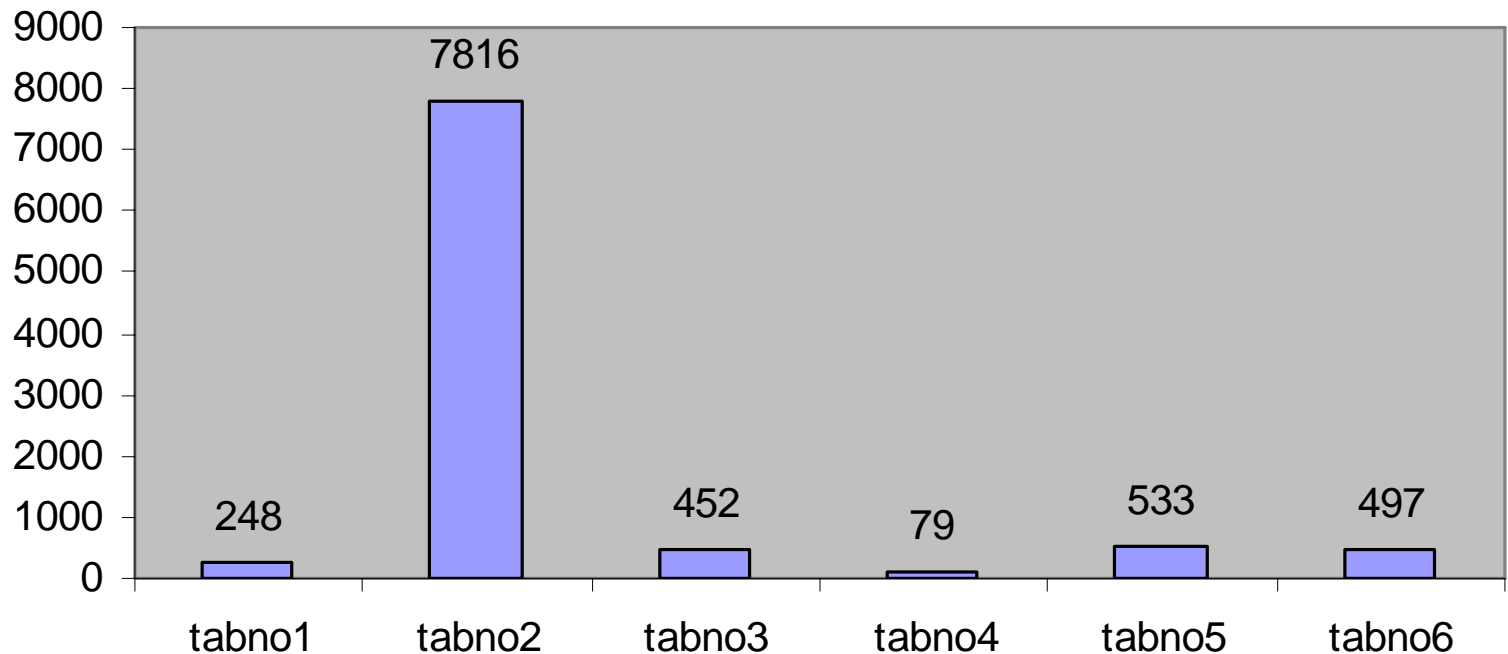
Analyse von DDC-Notationen (11)

Arbeitsschritt 3 (DDC-Wissensbasis)

R3: DDC-Wissensbasis *vc_KB*



Häufigkeitsverteilung der Hilfstafel-Notationen



tabno1: Standard Subdivisions; tabno2: Geographic Areas, Historical Periods, Persons; tabno3: Subdivisions for the Arts, for Individual Literatures, for Specific Literary Forms; tabno4: Subdivisions of Individual Languages and Language Families; tabno5: Ethnic and National Groups; tabno6: Languages

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (12)



Arbeitsschritt 3 (DDC-Wissensbasis) R3: Erstellung der DDC-Wissensbasis *vc_KB*

Internal DDC tables:

add table (Anhängetafel)	(27 table entries)	324.24-324.29
build table (Zeittafel)		830.1-838
generic table		342-349
option table		351.3-351.9
table of preference (Übersicht der Vorzugsreihenfolge)		T3B--3087
!! history table (in Marc21 tags 685, 761, 763)		
		„number discontinued“, „relocated to“, ...
manual build table, manual generic table, manual text table (Praxishilfe)		352-354
!! „Notation 001-999 replaces notation T7--09-T7--99 from Table 7 with the result that many numbers have been reused with new meanings“ (390.4-<nml>)		

Analyse von DDC-Notationen (13)



Arbeitsschritt 3 (DDC-Wissensbasis) R3: Erstellung der DDC-Wissensbasis *vc_KB*

vc_KB : 48.067 Fakten, 8.033 Regeln (2 MB)
vc_KB_spli : 121.035 Fakten, 80.324 Regeln (15 MB)

Ausschnitt aus *vc_kb*

...
 591.9<**hat**>Treatment of animals by specific continents, countries, localities
 591.9<**na4r7**>591.9T2--3-T2--9
 591.96<**hatien**>Animals--Africa
 591.978<**hatien**>Animals--Western States (U.S.)
 591.994<**hatien**>Animals--Australia
 592-599<**hat**>Specific taxonomic groups of animals
 592-599+01-08<**ba5**>Standard subdivisions590.1-590.8
 592-599+09<**ba5**>Historical, geographic, persons treatment
 592-599+0914-0919<**ba5**>[Treatment by areas, regions, places in general other than polar, temperate, tropical regions]
 592-599+1<**ba5**>General topics of natural history of animals<**na1r1**>591.3-591.7591
 592<**hat**>Invertebrates
 592<**nfa**>592-599
 592.165<**hatien**>Poisonous invertebrates
 592.16509162<**hatien**>Poisonous marine invertebrates
 592.176<**hatien**>Freshwater zooplankton,...
 592.177<**hatien**>Marine invertebrates
 592.1776<**hatien**>Zooplankton
 592.3<**hat**>Worms
 ...

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (14)



Arbeitsschritt 3 (DDC-Wissensbasis) R3: Erstellung der DDC-Wissensbasis *vc_KB*

vc_KB : 48.067 Fakten, 8.033 Regeln (2 MB)
vc_KB_spli : 121.035 Fakten, 80.324 Regeln (15 MB)

Ausschnitt aus *vc_KB_spli*

...
 591.9<hat>Treatment of animals by specific continents, countries, localities
 591.9<na4r7>591.9T2--3-T2--9
 591.96<hatien>Animals--Africa
 591.978<hatien>Animals--Western States (U.S.)
 591.994<hatien>Animals--Australia
 592-599<hat>Specific taxonomic groups of animals <--- 1. has to be splitted
 592<hatspan:592-599>Specific taxonomic groups of animals
 593<hatspan:592-599>Specific taxonomic groups of animals
 594<hatspan:592-599>Specific taxonomic groups of animals
 595<hatspan:592-599>Specific taxonomic groups of animals
 596<hatspan:592-599>Specific taxonomic groups of animals
 597<hatspan:592-599>Specific taxonomic groups of animals
 598<hatspan:592-599>Specific taxonomic groups of animals
 599<hatspan:592-599>Specific taxonomic groups of animals
 592-599+01-08<ba5>Standard subdivisions590.1-590.8 <--- 3. has to be splitted
 592-599+01-08<ba5>Standard subdivisions590.1-590.8 <--- 5. has to be splitted
 592+01<ba5span:592-599+01-08>Standard subdivisions590.1-590.8
 592+02<ba5span:592-599+01-08>Standard subdivisions590.1-590.8
 592+03<ba5span:592-599+01-08>Standard subdivisions590.1-590.8
 592+04<ba5span:592-599+01-08>Standard subdivisions590.1-590.8
 ...

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (15)



Arbeitsschritte 4-7

Liu1993

**L4: Erstellung Computersystem DND* (Dewey Number Decomposer)
realisiert in FoxPro** (* unseres Wissens nach nicht verfügbar)

L5: Durchführung der Zerlegung der 600 DDC-Notationen (Hauptklasse 700)

L6: Expertenbewertung der Ergebnisse der Zerlegung

L7: Schlussfolgerungen / Anwendungsbereiche

Reiner2005

R4: Erstellung der Analysekomponente *vc_day* realisiert in gawk

**R5: Durchführung der Analyse der 600 DDC-Notationen,
Durchführung der Analyse aller GVK-DDC-Notationen (alle DDC-Klassen)**

R6: Ergebnisse, Vergleich DND mit *vc_day*

R7: dto. Liu1993

Analyse von DDC-Notationen (16)



R4: Erstellung der Analysekomponente *vc_day* (1)

Syntaktische Betrachtungen (1)

Alphabet (endliche Menge von Symbolen)

- 1. Numerische Symbole (Ziffern): 0,1,2,3,4,5,6,7,8,9**
- 2. Alphabetische Symbole (Buchstaben): A, B, C, ... für Optionen**
- 3. Dewey-Punkt: "."**
- 4. Segmentierungssymbole: "' ' " bzw. "/" (ev. auch andere)**

Beispiele:

000, 576'.64, 551.7'92'0941, 598.7/64, 598.7/64/0981022, 05H (051-059 Option A: Hindi-language serial publications)

344.73/0125763347.30412576325763347.30412576 (ppn180304127)

720.2224945420903307449494542 (29-stellig)

Analyse von DDC-Notationen (17)



R4: Erstellung der Analysekomponente *vc_day* (2)

Syntaktische Betrachtungen (2)

Korrektheit

mögliche Fehler: zu wenig oder zu viele Ziffern, an falscher Stelle stehende Ziffern(folgen), Ziffernfolge in verdrehter Reihenfolge, fehlender "Deweypunkt", zuviele "Deweypunkte", an falscher Stelle stehende Buchstaben, ...

Entscheidbarkeit

Die Eigenschaft einer Zeichenkette, DDC-Notation zu sein, ist entscheidbar.

(Offene) Fragen:

- Welcher Sprachtyp wird durch die Menge aller DDC-Notationen definiert?
- Welcher Grammtiktyp wird durch die Syntheseanweisungen definiert?
- Ist die Notationssynthese eindeutig?
- Ist die Notationssynthese endlich?
- Komplexität des Algorithmus zur Analyse von DDC-Notationen?

Analyse von DDC-Notationen (18)



R4: Erstellung der Analysekomponente *vc_day* (3)

Syntaktische Betrachtungen (3)

Erzeugen synthetischer DDC-Notationen (1)

Anfügen von DDC-Notation(steil)en

- A1.) an eine DDC-Grundnotation aufgrund von Synthese-Anweisungen (instructions)**
- A2.) an eine DDC-Notation aus Hilfstafel 1 (Table 1: Standard Subdivisions) , falls nicht ausdrücklich ausgeschlossen (nicht beliebig oft erlaubt)**

19th ed. of the DDC

„ ... a formula for calculating ... potential classes in the system ...

$$21.504 \times [116 + (5 \times 6.933) + (2 \times 6.500) + 192 + 440] = 1.041.073.100^{(*)}$$

[(*) Francis L. Miksa: „The DDC, the Universe of Knowledge, and the Post-Modern Library“. Forest Press, Albany, New York, 1998, S. 24]

=> Notationssynthese endlich (Vermutung)

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (19)



R4: Erstellung der Analysekomponente *vc_day* (4)

Syntaktische Betrachtungen (4), Erzeugen synthetischer DDC-Notationen (2)

```

C          ::= ""
DI         ::= „0“ | „1“ | „2“ | „3“ | „4“ | „5“ | „6“ | „7“ | „8“ | „9“
DO         ::= „.“
bano       ::= schedno | tabno
clano      ::= schedno | tabno
dno        ::= dno_atom | dno_mol | dno_head dno_tail
dno_atom   ::= DI | DI C DI | schedno_atom | tabno_atom
dno_head   ::= sm | sd | ss | s
dno_mol    ::= DI C DI | bano dno | schedno_atom t1_atom | schedno_mol | t1_atom |
              schedno_atom t1_mol | schedno_atom t1_atom | dno_mol dno_atom
dno_span   ::= dno_span1-dno_span2
dno_span1  ::= dno
dno_span2  ::= dno
dno_tail   ::= [ DI C DO C ]* DI ([C DI ])*
s          ::= ss C DO C dno_tail
sd         ::= sm C DI
sm         ::= DI
ss         ::= sd C DI
schedno    ::= dno | dno_span | s
tabno      ::= dno | dno_span
...

```

=> reguläre Grammatik zur Erzeugung von DDC-Notationen (Vermutung)

Analyse von DDC-Notationen (20)



R4: Erstellung der Analysekomponente *vc_day* (5)

Semantische Betrachtungen

```
colibri/ul-test> egrep "47-" res/res_liu_t | egrep -v liu_ | sort -u
---.---4347----- Stuttgart and Tübingen districts
                        <ba4r2span:T1--093-T1--099:T2--4347>
---.---47-- Eastern Europe  Russia <na4r7:T2--47>
721.0447----- Metals <hat>
739.47-- [Historical, geographic, persons treatment] <hat>
747---- Interior decoration <hat>
---.-747- New York <ba4r2span:T1--093-T1--099:T2--747>
---.-----7947- Southern Coast Range counties <na4r7:T2--7947>
```

„47“ hat je nach Stelle und Kontext unterschiedliche Bedeutung

„74“ hat andere Bedeutung als „47“ (Reihenfolgerelevanz), z. B.

74 Drawing & decorative arts

=> semantische Analyse von DDC-Notationen erforderlich

Analyse von DDC-Notationen (21)



L4: Erstellung Computersystem DND

Arbeitsschritte zur Zerlegung synthetischer DDC-Notationen

- „1. Match the number to be decomposed against the Schedule by dropping digits on the right. If the complete number is found in the Schedule, the number is not a synthesized number and decomposition is complete. If no match is made after all digits are dropped, set the number aside as containing a potential error.**
- 2. When a match is made, search the entry in the Schedule for an Add Note. If one is found, determine which note type it is and apply the rule defined for that type. If none is found, apply appropriate rules for Standard Subdivisions.**
- 3. Repeat the above two steps for the remaining digits, but in repetitions, in the first step the number may be searched against the Tables rather than the Schedule.”**

Analyse von DDC-Notationen (22)



R4: Erstellung der Analysekomponente *vc_day* (6)

in 1	700.23	the subj as poj.			
in 2	700.0440747471	1940-1949	museum	new york	
in 3	700.942	historical	historical	england and wales	
in 4	700.972910904	historical	cuba	20th century	
in 5	700.97305	natural endowment	serial publ.		
in 6	700.97309034	19th century			
in 7	700.97309045	1950-1959			
in 8	700.974710904	historical	new york	20th century	
in 9	700.97949407479494	historical	los ang.	museums	los ang.
in 10	704.0396073	african americans			
in 11	704.03972	athabascan indians			
in 21	709.470744745	historical	russia	museums	tartar republic
in 22	709.4709048	historical	russia	1980-1989	
in 23	709.49209032	historical	netherlands	17th century	1600-1699
in 24	709.598	historical	indonesia and east timor		
in 25	709.730747445	american art	museums	essex	country
in 26	711.4092	city planners			
in 27	711.40977311	local	historical	chicago	
in 28	712.0942	historical	england and wales		
in 29	712.0945632	historical	rome		
in 30	712.0973092	historical	united states	persons	

„United States“ nicht entdeckt

... erste Schritte ...

nicht analysiert

grobe Zerlegung

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (23)



"Algorithms.

Once we have a suitable mathematical model for our problem, we can attempt to find a solution in terms of that model. Our initial goal is to find a solution in the form of an algorithm, which is a finite sequence of instructions, each of which has a clear meaning and can be performed with a finite amount of effort in a finite length of time.

"

[Alfred V. Aho; John E. Hopcroft; Jeffrey D. Ullman: Data Structures and Algorithms. Addison-Wesley Publishing Company, Reading, Massachusetts, 1982, p. 2]

Analyse von DDC-Notationen (24)



R4: Erstellung der Analysekomponente *vc_day* (7)

Algorithmus zur Analyse von DDC-Notationen (1)

BEGIN

```
1. FUNCTION fu_init_days_variables
2. FUNCTION fu_read_vc_kb_spli
3. WHILE (loop forever until break) DO BEGIN
4. | WHILE (port SRVPORT 7654 is connected) DO BEGIN
| | | a. count infile_lines
| | | b. determine dno, count number of (correct, incorrect) dnos
| | | c. if (vc_daygram_test_pri is 1) then print vc_daygram_test_pri_header
| | | d. phase1
| | |   1. FUNCTION fu_get_facts (from left to right)
| | | e. phase2
| | |   1. initialization
| | |   2. FUNCTION fu_get_facts_via_rules (from left to right)
| | | f. FUNCTION fu_chk_pri (determine output format, print)
| | | g. determine whether dno is totally complete or partially complete analyzed, count no. of dnos
| | | h. print -- FINI -- on SRVPORT
| | | i. FUNCTION fu_init_days_variables
| | END
| | print "client has closed the connection, classifying has finished"
| | close SRVPORT
| | PRINT statistics (number of [totally/partially complete analyzed dnos], ...)
| END
END
```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (25)



R4: Erstellung der Analysekomponente *vc_day* (8)

Algorithmus zur Analyse von DDC-Notationen (2)

```
FUNCTION fu_get_facts_via_rules
| 1. FUNCTION fu_init_phase2
| 2. WHILE (there is an unanalyzed dno_tail)
|   | a. determine dno_head, dno_tail and their length
|   |   (dno_head and dno_tail serves as data structure of dno,
|   |   analyzing begins with dno_head of length 1 and ends with
|   |   length of dno_head X (X: length of dno)
|   |   length of dno_tail is length of dno minus length of dno_head)
|   | b. FUNCTION fu_execute_phase2
|   | c. adjust dno_head, dno_tail
|   | d. if (try_tbl is 1)
|   |   1. if (bnst is not empty)
|   |     then FUNCTION fu_execute_bnst
|   |     else if (nst is not empty)
|   |       then FUNCTION fu_apply_nst
|   |     else
|   |       FUNCTION fu_execute_tbl
|   |   else (try_tbl is 1 and tried_tbl is 1)
|   |     1. FUNCTION fu_look_fwd
|   |     2. FUNCTION fu_apply_nfu
|   | e. adjust dno_head, dno_tail
|   END
| 3. FUNCTION fu_chk_ready
END
```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (26)



R4: Erstellung der Analysekomponente *vc_day* (9)

Algorithmus zur Analyse von DDC-Notationen (3)

```
FUNCTION fu_execute_phase2
| 1. determine fact_rule_type
| 2. if (dno_atom_pri_last is not empty)
|     then FUNCTION fu_apply_rule(dno_atom_pri_last)
| else if (dno_head_pri_last is not empty)
|     then FUNCTION fu_apply_rule(dno_head_pri_last)
| else if (dno_head_new is not empty)
|     then FUNCTION fu_apply_rule(dno_head_new)
| else if (try_tbl is 0 && is_hydra is 1)
|     then FUNCTION fu_apply_rule(npreno dno_hy_pri)
| else if (try_tbl is 0 && is_hydra is 0)
|     then FUNCTION fu_apply_rule(dno_head)
END
```

Analyse von DDC-Notationen (27)



R4: Erstellung der Analysekomponente *vc_day* (10)

Algorithmus zur Analyse von DDC-Notationen (4)

```
FUNCTION fu_apply_rule
| 1. determine fact_rule_type
| 2. if (nnotations matches "+:")           then FUNCTION fu_ref_to_add_tb
| 3. if (dno_nfu != "")                     then FUNCTION fu_look_fwd
| 4. determine dno_head, dno_tail
| 5. if (fact_rule_type matches "<hat" and matches not "<hatien>")
|                                           then FUNCTION fu_get_dno_atoms
|
| 6. FUNCTION fu_get_span
| 7. if (descr matches "<nfa")             then FUNCTION fu_apply_nfa
|     else if (descr matches "<ba")       then FUNCTION fu_apply_ba
|     else if (descr matches "<na")       then FUNCTION fu_apply_na
|     else if (descr matches "<nsi")      then FUNCTION fu_apply_nsi
|     else if (descr matches "<nfu")      then FUNCTION fu_apply_nfu
|     else                                 FUNCTION fu_apply_tbl
| 8. if (try_nad is 1)                       then FUNCTION fu_execute_nad
| 9. if (pri_line matches "T5--.*0.*>")     then FUNCTION fu_execute_nad
END
```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Analyse von DDC-Notationen (28)



Liu1993 (pp. 99-100)

„720.7073 has been decomposed as follows:

720: Architecture

0707: Geographical treatment

73: United States

The title of this book is:

#aVoices in architectural education : #bcultural politics and

The subject headings for this book are:

#aArchitecture #xStudy and teaching #zUnited States.

#aArchitecture and state #zUnited States.“

Reiner2006

720.7073 <ul-liu/liu_37_to_analyze; length: 8>

7----- Arts & recreation <hatzen>

72----- Architecture <hatzen>

720----- Architecture <hat>

--0.7--- Education, research, related topics <T1--07>

--0.707- Geographic treatment <T1--0707>

---.--7- North America <na4r7span:T1--0701-T1--0709:T2--7>

---.--73 United States <na4r7span:T1--0701-T1--0709:T2--73>

Analyse von DDC-Notationen (29)



```

780.710749 <liu_127_to_analyze; length: 10>
7----- Arts & recreation <hatzen>
78----- Music <hatzen>
780----- Music <hat>
780.7----- Education, research, performances, related
              topics <hat>
780.71---- Music--education <hatien>
--0.7----- Education, research, related topics <T1--07>
--0.71---- Education <T1--071>
--0.7107-- Geographic treatment <T1--07107>
---.---7-- North America <na4r7span:T1--07101-T1-07109:T2--7>(1)
---.---74- Northeastern United States (New England and
              Middle Atlantic states)
              <na4r7span:T1--07101-T1--07109:T2--74>
---.---749 New Jersey <na4r7span:T1--07101-T1--07109:T2--749>

```

(1) <na4r7span: Add to base number T1--0710 notation T2--1-T2--9>

zum Vergleich Liu1993: **780.7, (780), 07107, 749**



Ergebnisse (1)

1. Semantische Analyse von DDC-Notationen in einzelne Atome, nicht nur (grobe) Zerlegung (wie bei Liu1993, Riesthuis1997)

decompose

„to divide sth into smaller parts“ (1)

analyse

„to examine the nature or structure of sth, especially by separating it into its parts, in order to understand or explain it“ (1)

2. Zweiphasiger Analyseprozess von links nach rechts (Liu1993: von rechts nach links):

1. Phase: Ermittlung der Fakten (Haupttafeln)
2. Phase: Ermittlung der durch Regeln gewonnenen Fakten (aus Haupt- und Hilfstafeln und internen Tafeln)

[(1) Oxford Advanced Learner's Dictionary. Online: <http://www.cornelsen.de/oald/1.c.115282.de>]

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



Ergebnisse (2)

3. Erfolgreiche Reproduktion der Ergebnisse von Liu1993

- alle 600 DDC-Notationen der Stichprobe von Liu1993 analysiert (nach ev. notwendiger Korrektur von DDC-Notationen wegen fehlender oder überflüssiger Ziffern, Tippfehlern, ...)
- Unterschiede durch Verwendung unterschiedlicher DDC-Ausgaben (Liu1993: 20th Ed. of DDC, Reiner2005: 22nd Ed. of DDC)
- von 600 DDC-Notationen 27 DDC-Notationen zusätzlich analysiert, (wegen Beschränkung auf Hauptklasse 700 von DND nicht zerlegbar)

4a. Ergebnisse der Analyse aller GVK-DDC-Notationen (gesamt)

466.134 analysierte DDC-Notationen

297.782 total analysierte DDC-Notationen (63.9 %)

168.352 partiell analysierte DDC-Notationen (36.1 %)



Ergebnisse (3)

4b. Ergebnisse der Analyse aller GVK-DDC-Notationen (einzeln)

```

res_gvk_all: 466134
res_gvk0_all: 23946 analyzed dno(s) | 14887 (62.169%) totally complete analyzed dno(s)
res_gvk1_all: 4596 analyzed dno(s) | 4095 (89.0992%) totally complete analyzed dno(s)
res_gvk2_all: 20888 analyzed dno(s) | 16122 (77.1831%) totally complete analyzed dno(s)
res_gvk3_all:189246 analyzed dno(s) |119378 (63.0809%) totally complete analyzed dno(s)
res_gvk4_all: 6680 analyzed dno(s) | 3124 (46.7665%) totally complete analyzed dno(s)
res_gvk5_all: 33649 analyzed dno(s) | 27706 (82.3383%) totally complete analyzed dno(s)
res_gvk6_all: 51704 analyzed dno(s) | 44879 (86.7988%) totally complete analyzed dno(s)
res_gvk7_all: 52632 analyzed dno(s) | 43378 (82.4175%) totally complete analyzed dno(s)
res_gvk8_all: 20678 analyzed dno(s) | 4205 (20.3356%) totally complete analyzed dno(s)
res_gvk9_all 62115 analyzed dno(s) | 21036 (33.8662%) totally complete analyzed dno(s)

```

- Gründe für nicht vollständig analysierte DDC-Notation?
- wer sucht sie?

Analyse von DDC-Notationen (33)



Ergebnisse (4)

5. Analyse von DDC-Notationen der DDB-Datei "ddbt2"

285 analysierte DDC-Notationen

215 total analysierte DDC-Notationen (75.4 %)

70 partiell analysierte DDC-Notationen (24.6 %)

Ergebnisausgabe u.a. im MAB2-Format

6. Analyse von DDC-Notationen der DDB-Datei "0605all.ma2"

2.659 analysierte DDC-Notationen

2.210 total analysierte DDC-Notationen (83 %)

449 partiell analysierte DDC-Notationen (17 %)

Dauer: 105 sec.

7. Unterschiedliche Ausgabeformate

7a. Ausgabe als Analysediagramm: *vc_daygram*

7b. Ausgabe als Menge: *vc_dayset*

Formate (1)

Ergebnisse (5)



7a1. Ausgabeformat *vc_daygram*

- zu analysierende DDC-Notation (molekulare DDC-Notation)
- Name und Länge der zu analysierenden DDC-Notation in 1. Zeile
<liu_127_to_analyze ; length: 10>
- Dewey-Punkt an 4. Stelle der DDC-Notation
- Reihenfolge und Stellung der Ziffern (innerhalb) der DDC-Notation
- relevante DDC-Notationsteile pro Analyseschritt (atomare DDC-Notation),
- Kennzeichnung der im Analyseschritt irrelevanten Stellen mit „-“,
- Bedeutung der DDC-Notationsteile (Klassenbenennung) pro Analyseschritt,
- angewendete Regeln mit ermittelten DDC-Notationen

=> *vc_daygram* enthält Analyse- und Syntheseinformation

Formate (2)

Ergebnisse (6)



7a2. Ausgabeformat *vc_daygram*

```
-----
vc_daygram: 725.409430904 <H0605all_969_to_analyze>
-----
```

```
725.409430904 <H0605all_969_to_analyze; length: 13>
7----- Arts & recreation <hatzen>
72----- Architecture <hatzen>
725----- Public structures <hatspan:725-728>
725.4----- Industrial buildings <hat>
725.4----- Industrial buildings <nfa->721-729>
---.-09----- Historical, geographic, persons treatment <ba5span:721-729+09>
---.-09----- Historical, geographic, persons treatment <T1--09>
---.-094----- Treatment by specific continents, countries, localities;
extraterrestrial world <T1--094>
---.-4----- Europe Western Europe <ba4r2span:T1--093-T1--099:T2--4>
---.-43----- Central Europe Germany <ba4r2span:T1--093-T1--099:T2--43>
---.-09----- Historical and geographic treatment <ba4r2span:T1--093-T1--099+09>
---.-04----- 20thcentury, 1900-1999 <ba4r2span:T1--093-T1--099+09:na3r5:T1--0904>
```


Formate (3)

Ergebnisse (7)



7b1. Ausgabeformate *vc_dayset* (1)

vc_dayset_dno_cap: 725.409430904 <H0605all_969_to_analyze>

7;Arts & recreation

72;Architecture

725;Public structures

725.4;Industrial buildings

725+09;Historical, geographic, persons treatment

T1--09;Historical, geographic, persons treatment

T1--094;Treatment by specific continents, countries, localities; extraterrestrial worlds

T2--4;Europe Western Europe

T2--43;Central Europe Germany

T1--094+09;Historical and geographic treatment

T1--0904;20thcentury, 1900-1999

vc_dayset_fine: 725.409430904 <H0605all_969_to_analyze>

**H0605all_969:725.409430904;7;72;725;725.4;725+09;T1--09;T1--094;T2--4;T2--43;
 T1--094+09;T1--0904**

Formate (4)



Ergebnisse (8)

7b2. Ausgabeformate *vc_dayset* (2)

vc_dayset_fine: 725.409430904 <H0605all_969_to_analyze>

H0605all_969:725.409430904;7;72;725;725.4;725+09;T1--09;T1--094;T2--4;T2--43;T1--094+09;T1--0904

vc_dayset_mab2: 725.409430904 <H0605all_969_to_analyze>

705a□a725.409430904□p725.4□cT1--09□d725+09□f094;0904□g43□t725+09;T1--094+09

Analyse der DDC-Notation **725.409430904** im Vergleich:

DNB

705a□a**725.409430904**□c725.4□eDDC22ger□f0904□g43

vc_day

705a□a**725.409430904**□p725.4□cT1--09□d725+09□f094;0904□g43
□t725+09;T1--094+09

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Formate (5)



Auszug aus Katalogisierungsrichtlinie - Sacherschließung

PICA+	w	PICA3	Inhalt
--------------	----------	--------------	---------------

045H	5400		Synthetische DDC-Notation
045H/01	5401		Basisnotation
045H/02	* 5402		Notationen anderer Haupttafeln
045H/03	* 5403		Notationen aus Hilfstafeln
045H/04	* 5404		Notationen aus internal tables
045H/10	5410-14		Angaben zur zweiten DDC-Notation
045H/20	5420-24		Angaben zur dritten DDC-Notation
045H/30	5430-34		Angaben zur vierten DDC-Notation
045H/40	5440-44		Angaben zur fünften DDC-Notation

[<http://www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/02Richtlinien/01KatRicht/inhalt08.shtml>]



Auszug aus MAB2 Online-Kurzreferenz-Version

705 DDC (DEWEY DECIMAL CLASSIFICATION) ANALYTISCH

Unterfelder:

\$a = Vollstaendige Notation

\$b = Exemplarnummer (Item number)

\$c = Grundnotation

\$d = Notationen anderer Haupttafeln

\$e = Angabe der zugrunde liegenden DDC-Ausgabe

\$f = Notation aus Hilfstafel 1

\$g = Notation aus Hilfstafel 2

\$h = Notation aus Hilfstafel 3A

\$i = Notation aus Hilfstafel 3B

\$j = Notation aus Hilfstafel 3C

\$k = Notation aus Hilfstafel 4

\$l = Notation aus Hilfstafel 5

\$m = Notation aus Hilfstafel 6

\$t = Notation aus einer Anhaengetafel

\$A = Quelle der vergebenen Notation

0 = Notation von der Library of Congress vergeben

1 = Notation von Der Deutschen Bibliothek vergeben

4 = Notation von anderer Institution vergeben

[<http://www.ddb.de/standardisierung/txt/erw-mab.txt>]

Grundnotation (base number)



“Die Notationssynthese beginnt mit einer Grundnotation (die immer in der Anweisung angegeben ist), an die eine andere Notation angehängt wird. ... Es gibt vier Quellen für die Notationssynthese:

- (A) Standardschlüssel der Hilfstafel 1;
- (B) Hilfstafeln 2–6;
- (C) andere Teile der Haupttafeln;
- (D) spezielle Anhängetafeln in den Haupttafeln.„

[glossar_EG_OCLC_exp.pdf]

„Table 3A. ... The base number may be identified in an add note, e.g., at 820.1-828 (“add to base number **82**”) or another note, e.g., at 896 (...); otherwise, it is the number given for the literature, e.g., Dutch-language literature **839.31**. ...

...

896 (African literatures) ... Add to **896** ...” [22. WebDewey]

=> **es steht nicht immer** „ ... Add to base number ...” !

In e22 mind. 1580 DDC-Grundnotationen (automatisch ermittelt)

DDC-Suchsystem (2)



DDC-Suchsystem *vc_ds* mit den Hauptkomponenten

- * Analysekomponente *vc_day* ✓
- * Klassifizierungskomponente *vc_dcl*
- * Wissensbasis *vc_KB* ✓
- * Datenbasis *vc_DB*
- * invertierte Datenbasis *vc_IDB*

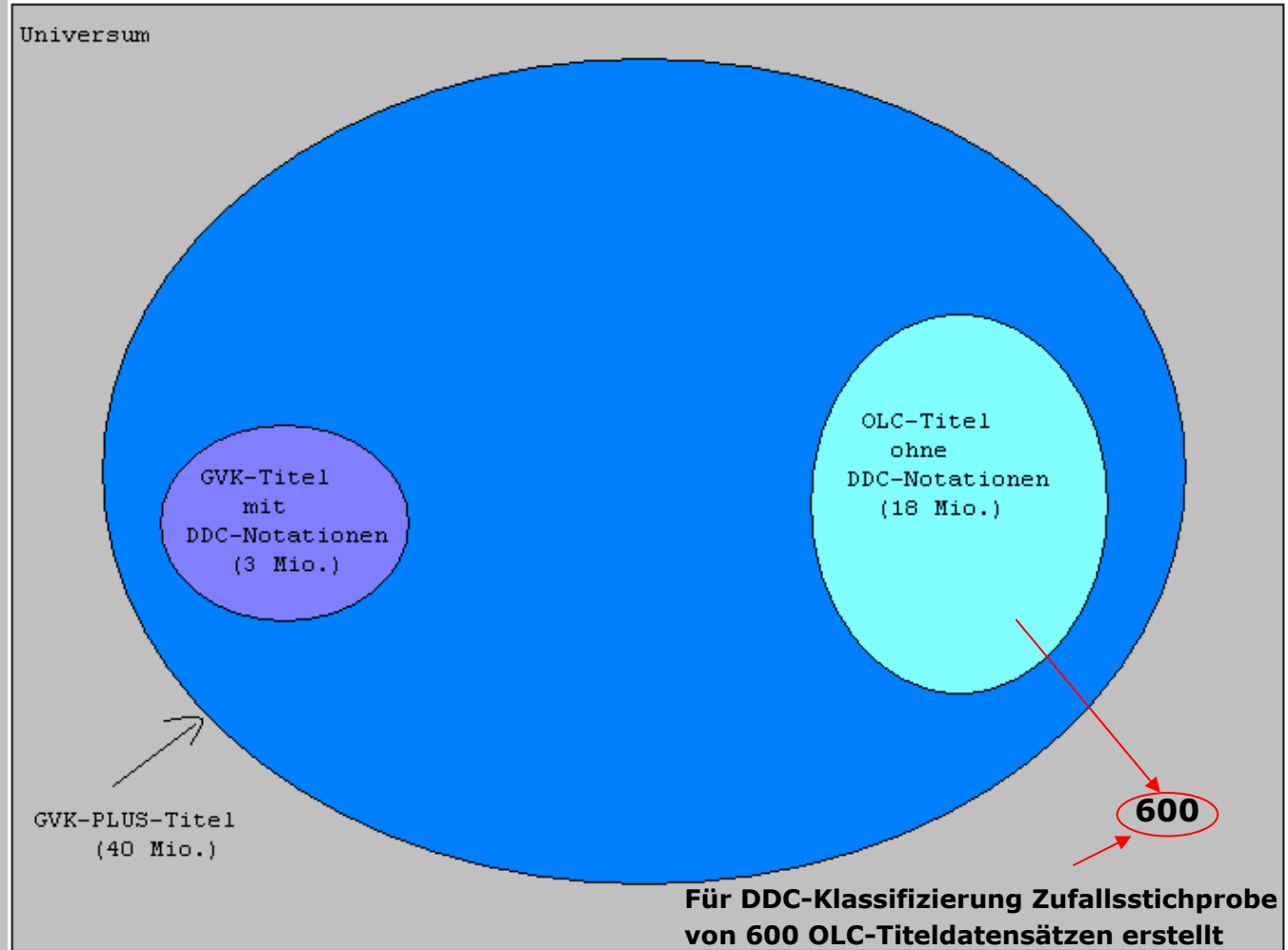
für die

- * Analyse von DDC-Notationen ✓
- * **DDC-Klassifizierung von Titeldatensätzen**



DDC-Klassifizierung (1)

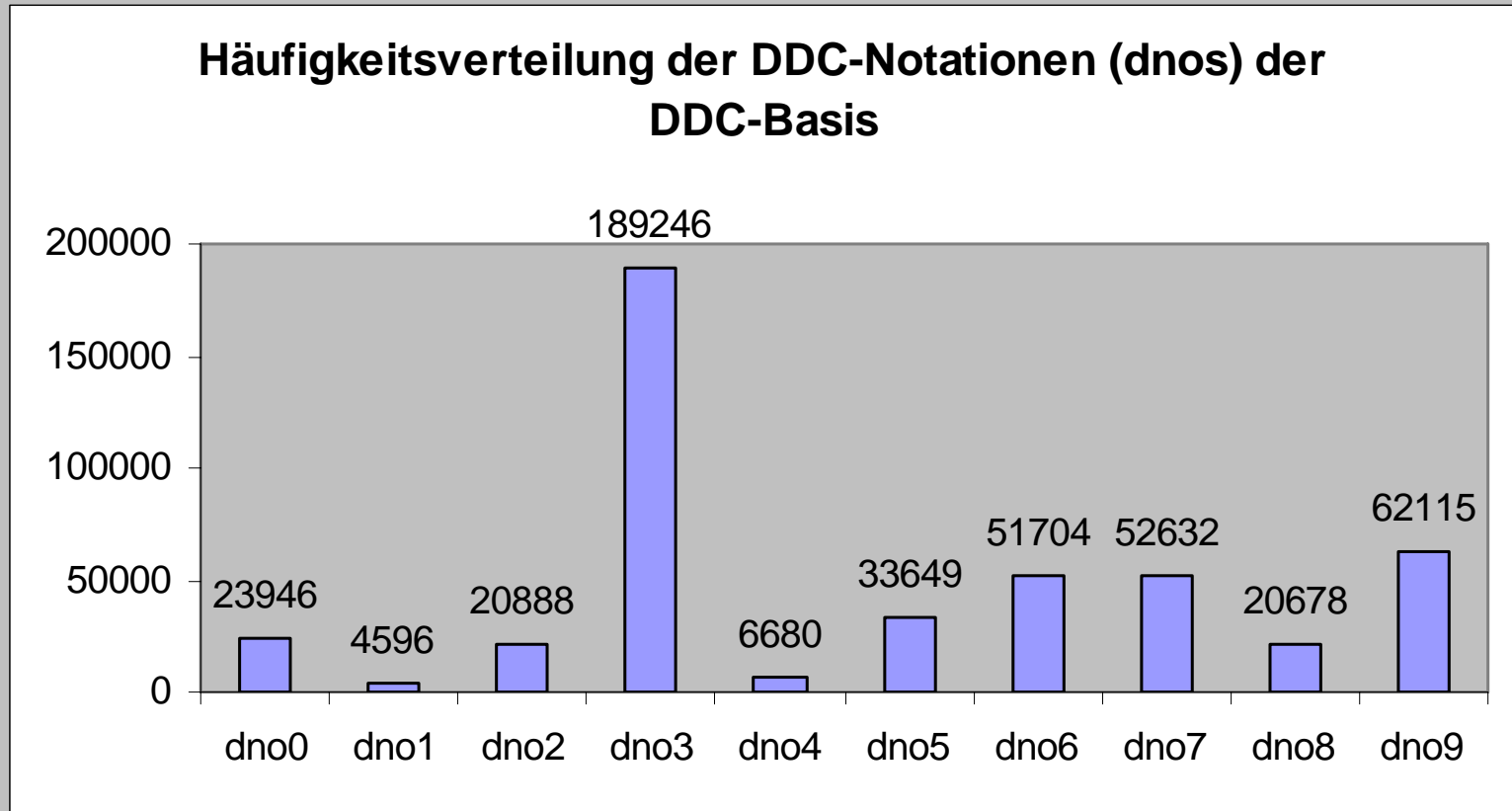
Bibliographische Titeldatensätze



Stand: Juli 2004

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (2)



0:Informatik&Informationswissenschaft & Allgemeines; **1** Philosophie & Psychologie; **2**: Religion;
3: Sozialwissenschaften; **4**:Sprache; **5**:Naturwissenschaften & Mathematik; **6**:Technik &
Technologie; **7**:Bildende Kunst & Kunsthandwerk, **8**:Literatur & Rhetorik, **9**: Geschichte & Geografie

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (3)



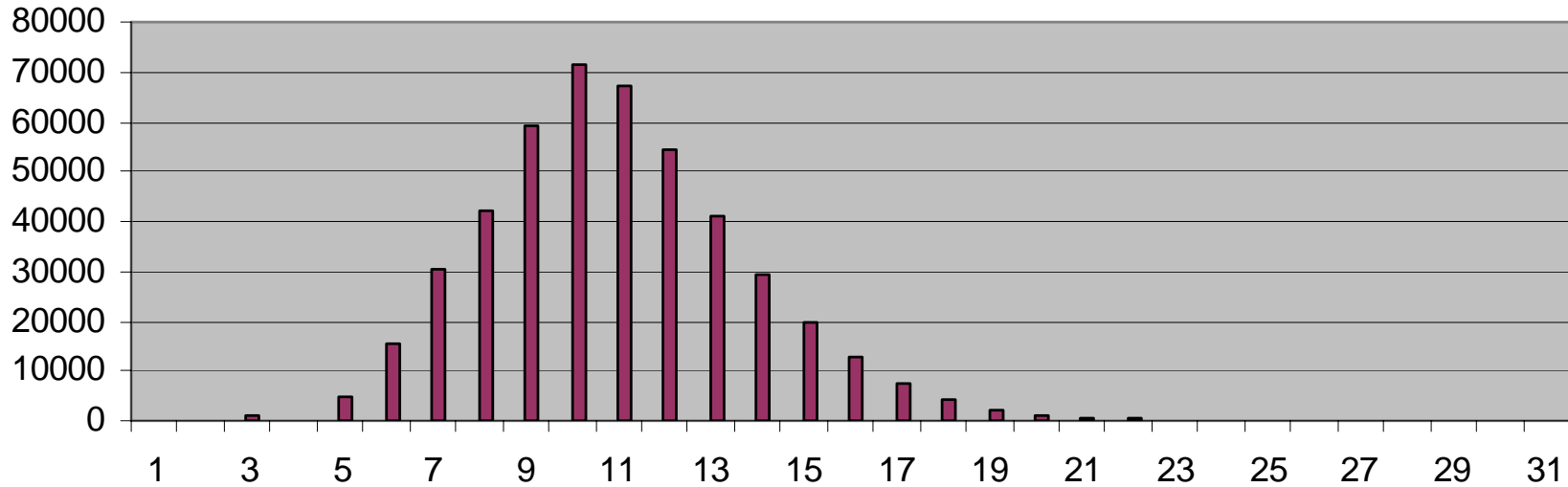
3.073.423 GVK-Titeldatensätze, die mind. eine DDC-Notation enthalten (Stand: Juli 2004), daraus werden ermittelt:

- * **3.339.717** (synthetische) DDC-Notationen
 - * **3.220.008** korrekte DDC-Notationen
 - * **119.709 inkorrekte DDC-Notationen** (3.8 % aller DDC-Notationen)
„3655474“, „322.10941 a) emancipation d) 1798-1822“, „B“, „081 s“
- => **466.134** unterschiedliche **DDC-Notationen** nach Eliminierung von Segmentierungszeichen, inkorrekt und mehrfach vorkommender DDC-Notationen

DDC-Klassifizierung (4)



Häufigkeitsverteilung der korrekten 466.134 DDC-Notationen
(nach Stelligkeit)



DDC-Klassifizierung (5)



Erstellung der Datenbasis *vc_DB* und *vc_DB_PLUS* (1)

1. Jeder DDC-klassifizierte GVK-Titeldatensatz wird konvertiert (vom Format PICA+ in die *vc_DB*-Repräsentationsform)

```

001A ^_02003:29-11-93
001B ^_02003:29-11-93^_t22:41:16.613
001D ^_09999:99-99-99
001X ^_00
002@ ^_0Aaxk
003@ ^_013137270X
...
021A ^_aThe @hummingbird book^_dthe complete guide to attracting, identifying,
    and enjoying hummingbirds^_hDonald and Lillian Stokes^_jillustrations and
    range maps by Donald Stokes
028A ^_dDonald W.^_aStokes
028C ^_dLillian Q.^_aStokes
032@ ^_alst ed
033A ^_pBoston^_nLittle, Brown and Co
...
037A ^_aOn cover: Plus a special section on attracting orioles
037A ^_aIncludes bibliographical references (p. 88-89)
044A ^_Sa^_aHummingbirds
044A ^_Sa^_aBird attracting
045A ^_aQL696.A558
045F ^_a598.8/99

```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (6)



Erstellung der Datenbasis *vc_DB* und *vc_DB_PLUS* (2)

2. Repräsentation der *vc_DB*-Deskriptorwerte analog *vc_KB*

```

598.899<9f><003@><9f>013137270X
598.899<9f><028A_da><9f>donald w.#stokes
598.899<9f><028C><9f>lillian q.#stokes
021A_line athe @hummingbird book^_dthe complete guide to attracting, identifying,
and enjoying hummingbirds^_hdonald and lillian stokes^_jillustrations and
range maps by donald stokes (1)
598.899<9f><021A><9f>hummingbird
598.899<9f><021A><9f>book
598.899<9f><021A><9f>identifying
598.899<9f><021A><9f>enjoying
598.899<9f><021A><9f>hummingbirds
598.899<9f><021A><9f>complete
598.899<9f><021A><9f>guide
598.899<9f><021A><9f>attracting
598.899<9f><033A_n><9f><033A_n>-little, brown
598.899<9f><044A_a><9f>hummingbirds
598.899<9f><044A_a><9f>bird attracting
598.899<9f><045A><9f>QL696.A558

```

3. Die im GVK vorkommenden DDC-Klassen der Haupttafeln mit klassifizierungsrelevanten Deskriptorwerten bilden die DDC-Basis *vc_DB*

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (7)



Erstellung der Datenbasis *vc_DB* und *vc_DB_PLUS* (3)

4. Irrelevante Deskriptorwerte werden mit der Funktion *fu_vc_elim* vor der Aufnahme in *vc_DB* eliminiert:

```

BEGIN {
  ELIM_1_SPEC_CHAR= " |\\(|\\)|\\[|\\]|\\^|'|Â´|`|?|!|\\\"|\\.|,|:|;|<|>|-|+|&"
  ELIM_1CHAR       = "[a-z0-9]"
  ELIM_2CHAR       = "[0-9][0-9]|An|ad|al|an|as|at|be|by|da|di|de|do|du|el|en|et|he
    |if|ii|il|im|in|is|it|la|le|li|lo|mo|na|nh|od|of|on|or|ou|to|up|vo|we|xi|zu"
  ELIM_3CHAR       = "[0-9][0-9][0-9]|all|als|and|ang|any|are|auf|aus|bis|can|das
    |del|dem|den|der|des|die|dun|ein|ems|fit|for|f\323r|f.r|hat|his|how|ias|iaa
    |iic|iid|iie|iii|iis|its|key|las|les|ltd|new|nur|our|que|see|sei|sin|[Tt]he
    |und|una|une|vii|vol|vom|von|you|zum|zur|var|was|xii"
  ELIM_4CHAR       = "[0-9][0-9][0-9][0-9]|acad|date|deck|dich|done|dual|from
    |gibt|have|hell|hrsg|iiav|iics|iigs|iiii| |iisg|iisi|jiia|less|lini|mkii
    |many|mein|more|nach|oder|ohne|open|over|phac|phar|publ|sich|with|than|that
    |upon|what|word|your|viii|xiii|xvii"
  ...
}

```

5. Die DDC-Basis *vc_DB* wird um die Fakten aus *vc_KB* erweitert => *vc_DB_PLUS*



Erstellung der invertierten Datenbasis *vc_IDB*

6. Für das Klassifizierungsverfahren wird die DDC-Basis *vc_DB_PLUS* invertiert => *vc_IDB*

```

...
hummingbirdÿ133
hummingbirdÿ299.7
hummingbirdÿ333.7 398.20899755 398.2452
hummingbirds--resource economicsÿ333.958764
hummingbirdsÿ333.958764 398.20899755 398.209 398.2452
hummingbirdsÿ428.6
hummingbirdÿ410 417.7
hummingbirdsÿ508 591.734 591.7 598 598.097 598.764 598.8
hummingbirdÿ598.764 598.8 598
hummingbirds petsÿ636.6
hummingbirdsÿ635.967 635.96 636.6 638.5789 639.978764 639.9788 639.9789
hummingbirdÿ635.967 636.6 639.978764 639.9788 639.9789
hummingbirds--conservation technologyÿ639.978764
hummingbirdsÿ736.4 782.254 785.25
hummingbirdÿ770.92
hummingbirdsÿ811.54 823.914
hummingbirdÿ811.54 813.52 813.54 813.6 814.52 823.914 823.92
hummingbirdÿ972.01 978.8
hummingbirdsÿ941.085
hummingbird rockshelter (colo.)ÿ978.8
...

```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)



7. Jeder nicht DDC-klassifizierte GVK-PLUS-Titeldatensatz wird genauso wie ein DDC-klassifizierter konvertiert, anstelle einer konkreten DDC-Notation steht „XXX“:

XXX <003@> 0781835577

XXX <028C> t.andrew#hurly

XXX <028C> susan d.#healy

XXX <028C> jonathan#henderson

021A_line a@spatial relational learning in rufous hummingbirds
(selasphorus rufus)

XXX <021A> rufous

XXX <021A> hummingbirds

XXX <021A> selasphorus

XXX <021A> rufus

XXX <021A> spatial

XXX <021A> relational

XXX <021A> learning

XXX <021A> cognition animals (über lex(animal_cognition,cognition animals))

XXX <027D_a> animal

XXX <027D_a> cognition

XXX <027D> animal_cognition

XXX <027D_n> <033A_n>-springer

XXX <039B> animal

XXX <039B> cognition

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (10)



Automatisches DDC-Klassifizierungsverfahren (1)

1. Grundlage: Nutzung der intellektuell geleisteten Arbeit

- am DDC-System (*e22*) und
- an den GVK-Titeldatensätzen (*vc_DB*)

=> Aufbau der DDC-Basis *vc_DB_PLUS*: *vc_DB* + *vc_KB_facts*
(+ aus *dno_atoms* gebildete Fakten, Fragestellung F3)

2. DDC-Klassenauswahl für die DDC-Basis

- In der DDC-Basis *vc_DB* sind die DDC-Klassen enthalten, die in der Haupttafel (*schedules*) verzeichnet sind
- Deskriptorwerte von synthetischen DDC-Notationen (Diskriminatoren !) werden bei ihren direkten Haupttafel-Oberklassen abgespeichert



Automatisches DDC-Klassifizierungsverfahren (2)

3. Statistisches Verfahren

Vektorraummodell nach Salton 1968

4. Exakte Übereinstimmung (*exact match*)

zwischen Deskriptorwerten der Datenbasis und denen des Titeldatensatzes

5. Lexikon (im Modell, noch nicht implementiert)

6. Heuristische Funktion (*cutoff_val*)

Der *cutoff_val* begrenzt die Höhe des Häufigkeitwertes von Deskriptorwerten. Er wird bei jedem zu klassifizierenden Titeldatensatz in Abhängigkeit der Häufigkeitsverhältnisse der in Betracht kommenden Deskriptorwerte dynamisch ermittelt. (*cutoff_val* auch statisch festlegbar)

DDC-Klassifizierung (12)



Automatisches DDC-Klassifizierungsverfahren (3)

7. Journal Title weighting (*JT_weight*)

Falls ein OLC (Online Contents)-Zeitschriftenartikel in einer DDC-klassifizierten Zeitschrift vorkommt, die bei der Library of Congress verzeichnet ist (LC Z39.50/SRW/SRU Server), wird mithilfe von *JT_weight* die dort angegebene DDC-Klasse durch einen hohen Gewichtungsfaktor (z.B. *JT_weight* =40) priorisiert

8. Qualität der Klassifizierung

Im Wesentlichen abhängig von Qualität der GVK-Ausgangsdaten (*vc_DB*) und heuristischer Funktion *cutoff_val*

DDC-Klassifizierung (13)



Automatisches DDC-Klassifizierungsverfahren (4)

9. Klassenaggregation für die Ergebnisausgabe (1)

- die DDC-Klassifizierungskandidaten werden auf DDC-Klassen der Ebene 3 (sections) und der Ebene 1 (main) reduziert (durch Längenbegrenzung der DDC-Notationen auf 3 Stellen bzw. 1 Stelle mit zwei angefügten Nullen) :

calculated ddc classes (subdiv): |5| {573.86, 591.5, 591.51, 591.513, 595.7}

calculated ddc classes (sections): |3| {573, 591, 595}

calculated ddc classes (main): |1| {500}

DDC-Klassifizierung (14)



Automatisches DDC-Klassifizierungsverfahren (5)

9. Klassenaggregation für die Ergebnisausgabe (2)

- die DDC-Klassifizierungskandidaten werden in Abhängigkeit der Anzahl ihres Vorkommens (Häufigkeiten) pro Ebene aggregiert. Die so ermittelten DDC-Klassen werden für die Ebenen 1, 2, 3 und ggf. weitere Ebenen mit Angabe der Häufigkeit (in eckigen Klammern) angezeigt:

calculated ddc classes (subdiv): |5| {573.86, 591.5, 591.51, 591.513, 595.7}
calculated ddc classes (subdiv): {591.5[3]}
calculated ddc classes (sections): {591[3]}
calculated ddc classes (div): {590[4]}
calculated ddc classes (main): {500[5]}

DDC-Klassifizierung (15)



Automatisches DDC-Klassifizierungsverfahren (6)

10. DDC-Klassifizierung von T1 mit *vc_ds*-Klassifizierungskomponente *vc_dcl* :

number of ddc-classified title: 3
 ppn (dno): 0781835577 (XXX)
 calculated cutoff value: 145
 title: a@spatial relational learning in rufous hummingbirds (selasphorus rufus)
 considered descriptor values: |17| {<028C>-t.andrew#hurly[0],
 <028C>-susan d.#healy[0], <028C>-jonathan#henderson[0],
 <021A>-rufous[1], <021A>-hummingbirds[25], <021A>-selasphorus[0],
 <021A>-rufus[52], <021A>-spatial[419], <021A>-relational[145],
 <021A>-learning[2250], <021A>-cognition animals[18],
 <027D_a>-animal[836], <027D_a>-cognition[411],
 <027D>-animal_cognition[0], <027D_n>-<033A_n>-springer[3321],
 <039B>-animal[836], <039B>-cognition[411]}
 matched descriptor values |5| {hummingbirds, cognition animals, rufus,
 relational, rufous}
 max. match value: |3|
 calculated ddc classes |8| {574, 591.5, 591.51, 591.513, 591.59,
 596, 599.05, 599.8}
 calculated ddc classes (sections): |4| {574, 591, 596, 599}
 calculated ddc classes (main): |1| {500}

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (16)



Algorithmus zur DDC-Klassifizierung von Titeldatensätzen (1)

BEGIN

1. FUNCTION fu_read_idb

| read inverted database idb

| END

2. initialization

3. WHILE (loop forever until break) DO BEGIN

4. | WHILE (port SRVPORT 7655 is connected) DO BEGIN

| | | a. read infile_line from SRVPORT

| | | b. count infile_lines

| | | c. ignore _line infile_lines for classification process

| | | d. split infile_lines and store elements in

| | | 1. in_dno 2. in_descr 3. in_descr_val

| | | e. FUNCTION fu_vc_elim

| | | eliminate irrelevant in_descr_val (comparison with negative dictionary)

| | | f. if in_descr_val is not empty

| | | 1. store all in_descr_val in in_descr_vals

| | | g. if in_descr is ppn_tag

| | | 1. in_descr_val is a ppn, store it in in_ppn

| | | 2. count the number of title records

• • • ...

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (17)



Algorithmus zur DDC-Klassifizierung von Titeldatensätzen (2)

```
. . . . .
| | | h. if in_descr_val is content relevant (and not empty)
| | |   1. count the number of in_descr_val
| | |   2. determine dno_set and dno_set_card (cardinality of dno_set)
| | |     of this in_descr_val
| | |   3. if in_descr is JT_tag (Titel in Bandsätzen)
| | |     a. set dno_set_card to JT_weight for this in_descr_val
| | |     b. label in_descr_val with descr "<027D>" and store it
| | |       in in_descr_val_JT
| | |     c. increase cardinality of JT's class by JT_weight
| | |   4. if dno_set of in_descr_val is not empty
| | |     a. for each in_descr_val increase the cardinality of its class
| | |     b. label each in_descr_val with its in_descr
| | |     c. if in_descr_val is not (already) in descr_cand
| | |       store in_descr_val in descr_cand
. . . . .
```

DDC-Klassifizierung (18)



Algorithmus zur DDC-Klassifizierung von Titeldatensätzen (3)

```

. . . . .
| | | 5. if in_descr is ppn_tag or client has sent "-- EOF --"
| | |   FUNCTION fu_res_match
| | |     determine the result of the matching process
| | |     a. FUNCTION fu_cutoff
| | |       implement the cutoff process
| | |       1. FUNCTION fu_get_cutoff_val
| | |         determine the dynamic cutoff_value
| | |     b. FUNCTION fu_get_cand_sets
| | |       determine the ddc candidat sets
| | |     c. FUNCTION fu_pri_res
| | |       print the result with calculated ddc classes
| | |       1. FUNCTION fu_pri_main
| | |         print the calculated (DDC) main classes
| | | 6. printf "-- FINI --" on SRVPORT;
| | END
| | print "client has closed the connection, classifying has finished"
| | close SRVPORT
| END
5. PRINT statistics (read time, main time, number of ddc classified titles)
END

```

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Suchsystem (2)



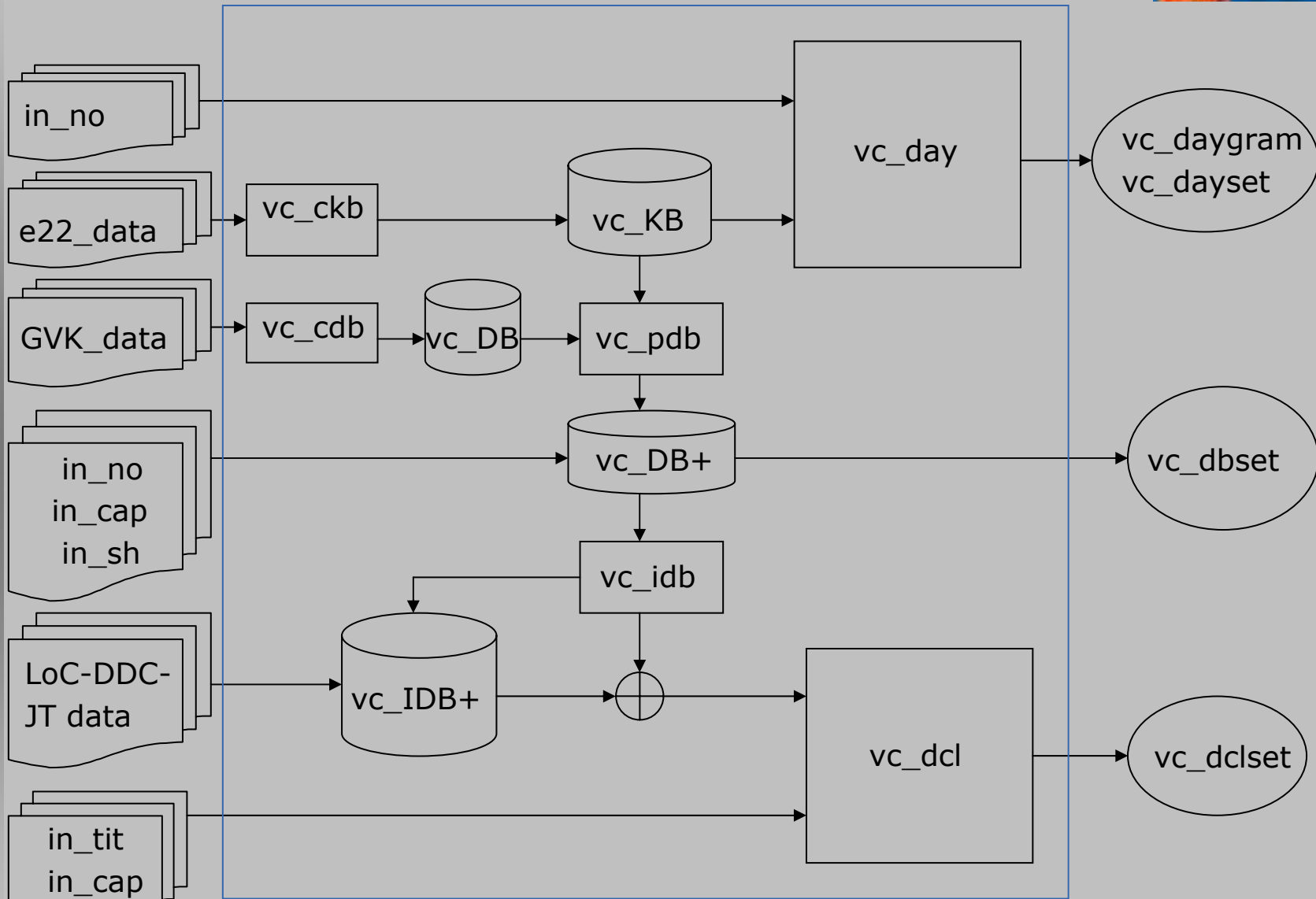
DDC-Suchsystem *vc_ds* mit den Hauptkomponenten

- * Analysekomponente *vc_day* ✓
- * Klassifizierungskomponente *vc_dcl* ✓
- * Wissensbasis *vc_KB* ✓
- * Datenbasis *vc_DB* ✓
- * invertierte Datenbasis *vc_IDB* ✓

für die

- * Analyse von DDC-Notationen ✓
- * DDC-Klassifizierung von Titeldatensätzen ✓

DDC-Suchsystem (3)



Systemarchitektur

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Entwicklung (1)



Stand: 05.01.07

loc(1)Name Colibri Entwicklungszeiten Tage(2)

pre-phase: predecessor program versions of vc_XXX programs

1112	ul-ana	analyse synthesized ddc notations	18.03.04 - 05.04.04	86
39	ul-base	build ddc_base	20.01.04 - 21.01.04	2
71	ul-ddc	automatic ddc classification	26.05.03 - 16.12.03	16
77	ul-ddc-inv	build inverted file for ddc class.	10.01.04	1
372	ul-ddc-pre	preparation of ddc_base, infile	21.04.03 - 14.01.04	29
219	ul-e22-pre	preparation of e22 ddc file	30.01.04 - 13.08.04	26
76	ul-elim	elimination of irrelevant terms	06.12.03 - 22.12.03	3
56	ul-freq	ddc_base frequencies	29.06.03 - 06.12.03	4
29	ul-get	get longest molecular dno	03.05.04	1
====				===
2051				168

2.051 Lines of code in 168 days = 5 month

- (1) loc: lines of code (pure code, without documentation and print statements)
(2) at least one line of code per day

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Entwicklung (2)



 main phase: development of vzg colibri programs vc_XXX

loc(1)	Name	Colibri	Entwicklungszeiten	Tage(2)
977	vc_cdb.awk	create ddc_base	12.10.04 - 28.11.06	33
3	vc_cdb_ksh	... for 10 ddc files		
9	vc_cin1.awk	create in_files for vc_days in: dno; out: {in,dno}	26.02.05 - 30.03.06	4
13	vc_cin2.awk	create in_files for vc_days in: {in,dno}; out: 1 {in,dno} pro 1 in_file	14.06.05	1
8	vc_ckb [ksh]	create knowledge base	17.10.04 - 01.01.05	2
216	vc_ckb.awk	create knowledge base	17.10.04 - 08.11.06	31
90	vc_dat1.awk	data analysis: card. of dno_all, dno_correct, dno_incorrect	25.12.06	1
38	vc_dat2.awk	data analysis: card. of x-arity-dnos	05.01.07	1
10	vc_dat3.awk	data analysis: count number of dnos	27.12.06	1
30	vc_dat4.awk	data analysis: count number of synthesized dnos	31.12.06	1
3	vc_dat4_ksh	... for 10 ddc files		
9	vc_day [ksh]	ddc notation analyzer	16.10.04 - 18.05.05	3

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Entwicklung (3)



loc(1)	Name	Colibri	Entwicklungszeiten Tage(2)	
8	vc_days [ksh]	ddc notation analyzer for vc_kb_spli(tted)	03.08.05	1
1542	vc_day.awk	ddc notation analyzer	15.10.04 - 24.07.05	115
3760	vc_days.awk	ddc notation analyzer for vc_kb_spli(tted)	02.08.05 - 27.09.06	246
18	vc_dcl [ksh]	ddc classification	07.09.06	1
334	vc_dcl.awk	ddc classification	07.09.06 - 06.12.06	38
	- vc_docu [text]	documentation file	15.10.04 - ongoing	
15	vc_idb [ksh]	create inverted ddc_base	07.09.06	1
40	vc_idb.awk	create inverted ddc_base	07.10.06 - 15.11.06	5
5	vc_idb_ksh	... for 10 ddc files		
30	vc_idba.awk	create inverted ddc_base idno-added (f.e. WWW)	13.12.06	1
9	vc_kbs [ksh]	knowledge base (range) splitter	01.06.05	1
301	vc_kbs.awk	knowledge base (range) splitter	01.06.05 - 06.12.05	12
692	vc_lib_KB.awk	program library for ddc knowledge base	15.10.04 - 02.08.06	39
95	vc_lib_DB.awk	program library for ddc_base	19.02.05 - ongoing	3
13	vc_low.awk	convert descr_val (letters) to lower case	16.11.06	1
33	vc_pdb.awk	prepare ddc_base	28.09.06	1
38	vc_pin.awk	prepare input file to ddc	28.09.06	1

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

Entwicklung (4)



loc(1) Name	Colibri	Entwicklungszeiten	Tage(2)
15 vc_pdl2in.awk	convert ppn_dno_list (pdl) file into file "in"	13.07.05	1
64 vc_polc.awk	prepare olc data	30.11.06 - 06.12.06	2
36 vc_psam.awk	prepare sample	09.12.06	1
18 vc_psru.awk	prepare sru requests	07.12.06	1
11 vc_ran.awk	random number generator	27.11.06	1
====			===
8483			550

8.483 Lines of code

550 days = 18 month = 1 1/2 years.

DDC-Klassifizierung (19)



Ergebnisse (9)

- 1. Ca. 2 Jahre Programmentwicklungszeit für DDC-Suchsystem *vc_ds* mit Analysekomponente *vc_day* und Klassifizierungskomponente *vc_dcl*. Der verwendete Zeitanteil beträgt ca. 3:1 (*vc_day*:*vc_dcl*).**
- 2. Erfolgreiche DDC-Testklassifizierung mit ca. 21.000 DDC-vor-klassifizierten GVK-Titeldatensätzen aus allen DDC-Klassen. In 0.3% aller 21.000 DDC-Klassifizierungen unterscheiden sich die Ergebnisse in der intellektuellen und automatischen Klassifizierung. Es kann untersucht werden, welches (im Allgemeinen) die bessere DDC-Klassifizierung ist.**
- 3. Für die Bestimmung der Qualität der automatischen DDC-Klassifizierung wurde eine repräsentative OLC-Stichprobe (Umfang: 600) gezogen und es wurde automatisch mit und ohne *JT_weight* DDC-klassifiziert. Das Ergebnis sollte vor Fortführung weiterer Arbeiten intellektuell überprüft werden.**

DDC-Klassifizierung (20)



Ergebnisse (10)

BK-DDC-Konkordanz (1)

4. Mit Hilfe der Klassifizierungskomponente *vc_dcl* wurde folgende BK(Hauptklassen)-DDC-Konkordanz (semi-)automatisch erstellt:

ohne #: semi-automatisch ermittelt

mit #: automatisch ermittelt (mit *vc_dcl* und *vc_DB_PLUS*)

dno000 Computer science, information & general works

dno100 Philosophy & psychology

dno200 Religion

dno300 Social sciences

dno400 Language

dno500 Science

dno600 Technology

dno700 Arts & recreation

dno800 Literature

dno900 History & geography

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (21)



Ergebnisse (11) BK-DDC-Konkordanz (2)

bno01 Allgemeines	-> {dno000}
bno02 Wissenschaft und Kultur allgemein	-> {dno300#, dno500, dno900#}
bno05 Kommunikationswissenschaft	-> {dno300}
bno06 Information und Dokumentation	-> {dno000}
bno08 Philosophie	-> {dno100}
bno10 Geisteswissenschaften allgemein	-> {dno000#, dno300}
bno11 Theologie, Religionswissenschaft	-> {dno200}
bno15 Geschichte	-> {dno300#, dno900}
bno17 Sprach- und Literaturwissenschaft	-> {dno400, dno800#}
bno18 Einzelne Sprachen und Literaturen	-> {dno400#, dno800}
bno20 Kunstwissenschaften	-> {dno300#, dno700, dno800#, dno900#}
Kunstwissenschaft	-> {dno700#}
bno21 Einzelne Kunstformen	-> {dno700}
Einzelne Kunstform	-> {dno700#}
bno24 Theater, Film, Musik	-> {dno700}
bno30 Naturwissenschaften allgemein	-> {dno500}
bno31 Mathematik	-> {dno500}
bno33 Physik	-> {dno500}
bno35 Chemie	-> {dno500}
bno38 Geowissenschaften	-> {dno900}
bno39 Astronomie	-> {dno500}

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (22)



Ergebnisse (12)

BK-DDC-Konkordanz (3)

bno42	Biologie	-> {dno500}
bno43	Umweltforschung, Umweltschutz	-> {dno300#, dno500, dno600}
bno44	Medizin	-> {dno600}
bno46	Tiermedizin	-> {dno600}
bno48	Land- und Forstwirtschaft	-> {dno300#, dno600}
bno49	Hauswirtschaft	-> {dno300}
bno50	Technik allgemein	-> {dno600}
bno51	Werkstoffkunde	-> {dno600}
bno52	Maschinenbau, Energietechnik, Fertigungstechnik	-> {dno600}
bno53	Elektrotechnik	-> {dno600}
bno54	Informatik	-> {dno000}
bno55	Verkehrstechnik, Verkehrswesen	-> {dno300}
	Verkehrstechnik	-> {dno300#, dno600#}
bno56	Bauwesen	-> {dno600}
bno57	Bergbau	-> {dno300, dno600}
bno58	Chemische Technik, Umwelttechnik, verschiedene Technologien	
	Chemische Technik	-> {dno500, dno600#}
	Umwelttechnik	-> {dno300, dno600#}
	verschiedene Technologien	-> {dno300#, dno600#}
bno70	Sozialwissenschaften allgemein	-> {dno300}

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Klassifizierung (23)



Ergebnisse (13) BK-DDC-Konkordanz (4)

bno71 Soziologie	-> {dno300}
bno73 Ethnologie, Volkskunde	-> {dno300}
bno74 Geographie, Raumordnung, StÄädtebau	-> {dno300#}
Geographie	-> {dno900#}
Raumordnung	-> {dno300#}
StÄädtebau	-> {dno300#}
bno76 Sport, Freizeit, Erholung	-> {dno300#}
Sport	-> {dno700#}
Freizeit	-> {dno300#}
Erholung	-> {dno300#}
bno77 Psychologie	-> {dno300#}
bno79 SozialpÄädagogik, Sozialarbeit	-> {dno300}
bno80 PÄädagogik	-> {dno300}
bno81 Bildungswesen	-> {dno300}
bno83 Volkswirtschaft	-> {dno300}
bno85 Betriebswirtschaft	-> {dno600}
bno86 Recht	-> {dno300}
bno88 Verwaltungslehre	-> {dno300}
bno89 Politologie	-> {dno300}
bno42.83 Aves	-> {dno568, dno574#, dno581#, dno591#, dno598, dno599#}
Syn.: VÄägel	
Hier: Ornithologie	

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

DDC-Aussichten

[<http://pie.midco.net/dougback/miscphotos/AZ2005/Broad-billed%20hummingbird.jpg>]



Es ist blauer Himmel, ein paar Schönwetterwolken behindern die Aussicht nicht. Ein gutes Stück des Weges ist bewältigt, aber es liegen noch steile Berge auf dem Weg nach ganz oben ...



Vielen Dank
für Ihr
Interesse
am
VZG-Projekt
Colibri

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

?Analyse folgender DDC-Notationen?



in_liu_20

709.17671

in_liu_62

728.6089970728

in_liu_63

728.80103

in_liu_66

729.092

in_liu_175

796.323092

in_liu_180

796.35202

in_liu_263

728.042

in_liu_323

769.497963570973

in_liu_346

781.550943613

in_liu_355

782.421621300268

in_liu_359

783.714207

in_liu_525

781.24111

in_liu_532

782.10268

in_liu_554

791.43028092273

Liu's Schlussfolgerungen



„Since the success rate of the decomposition is more than 95% (actually 100%), it is concluded that synthesized DDC numbers can be *accurately* decomposed by computer. Although the study covers only DDC synthesized class numbers in the main class Arts (700), I strongly believe that synthesized numbers in all the DDC main classes can be accurately decomposed automatically, on the ground that the synthesizing rules for the Arts (700) are representative of the rules throughout the schedule. Synthesizing rules for the main classes Literature (800) and Language (400) may differ somewhat from those for other classes. ... thus, with some modification or augmentation, the decomposing rules that were defined by this study could be extended to all synthesized numbers in all DDC main classes. Nevertheless, it seems a logical next step to extend and apply the methodology to other DDC schedules. Future research on automatic decomposition of synthesized numbers in the other nine main DDC classes would validate and supplement the findings and conclusions of the present study.“ [Liu1993], p. 65

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)

GVK-Titeldatensatz (in_liu_37)



001A 02000:12-05-92
 001B 00009:11-04-03 t14:56:49.000
 001D 00009:11-04-03
 001X 00
 002@ 0Aau
 003@ 0019617852
 004A 00897892534
 006A 0919593
 006B 00897892534
 006B 0b9222726
 006R 0FF31/07K
 010@ aen
 011@ a1991
 019@ aus
 021A aVoices in architectural education dcultural politics and pedagogy
 hThomas A. Dutton [Hrsg.]
 028C dThomas A. aDutton BHrsg.
 033A pNew York,NY [u.a.]: Bergin & Garvey
 034D axxix, 295 p
 034M aill
 036E a@Critical studies in education and culture series
 037A aIncludes bibliographical references (p. [279]-288) and index
 044A Sa aArchitecture Sx aStudy and teaching Sz aUnited States
 044A Sa aArchitecture and state Sz aUnited States
 044G Sa aArchitecture Sb aProfessional education
 045A aNA2105
 045F a720'.7'073
 045F a720/.7/073
 045F a720.7073

Workshop DDC-Einsichten und -Aussichten (ul, 1. März 2007)