

DDC-based Search in the Data of the German National Bibliography

Abstract. In 2004, the German National Library began to classify title records of the German National Bibliography according to subject groups based on the divisions of the Dewey Decimal Classification (DDC). Since 2006, all titles of the main series of the German National Bibliography are classified in strict compliance with the DDC. On this basis, an enhanced DDC-based search can be realized – e.g., searching the data of the German National Bibliography for title records using number components of synthesized classification numbers or searching for DDC numbers using unclassified title records. This paper gives an account of the current research and development of the DDC-based search. The work is conducted in the VZG project Colibri that focuses on the automatic analysis of DDC-synthesized numbers and the automatic classification of bibliographic title records.

I. Introduction

The Dewey Decimal Classification (DDC) system is the world's most widely used library classification system. Since 2006, the German National Library DNB (Deutsche Nationalbibliothek) is among the libraries of 60 countries that use the DDC to organize their national bibliographies. The DDC has been translated into more than 30 languages. While the first German version of the DDC's 22nd print edition was published in October 2005, other translations are still in progress or are proposed for translation. E.g., the Bibliotheca Alexandrina is currently producing an Arabic translation of Edition 22, the National Library of China has proposed a Chinese translation of Edition 21, and an exploratory study for a Swedish translation is underway. The German translation of the DDC is the first besides the English standard edition that is published simultaneously in a Web version and in a printed version.¹ The German work on the DDC was preceded by two workshops at the DNB (in October 1998 and 1999): Here, workgroup members could convince themselves of the DDC's development potential.² Further DDC workshops and an inspiring traveling exhibition provide evidence of the successful implementation of the DDC in the German-speaking countries.³ A great deal of this success is due to Magda Heiner-Freiling (DNB), who always promoted the work on the DDC with concentrated passion. Her thoughts about the future of the DDC can be best expressed in her own words: "Die DDC-Einsichten gefallen mir gut, vielleicht könnten wir sie sogar noch mit den Aussichten anreichern?"⁴ Some of these prospects that were discussed at the workshop "DDC-Einsichten und -Aussichten 2007" have – as Magda Heiner-Freiling would have wanted it – meanwhile turned into new

¹ Dewey Services : Overview : DDC Translations [<http://www.oclc.org/dewey/about/translations/default.htm>].

² Foreword by Magda Heiner-Freiling in "Introduction and Use of the DDC in the German-speaking countries", conducted by the Subject Indexing working group on behalf of the Conference for Cataloging Rules, Frankfurt a. M., 2000 [<http://www.ddc-deutsch.de/publikationen/pdf/machbarkeit.pdf>].

³ - Presentations held during the final workshop "Presentation of the German DDC 22 and Perspectives for the Application of the DDC in German-speaking Countries and International" of the DDC Deutsch project (DNB, Frankfurt a. M., April 20, 2005) [<http://www.ddc-deutsch.de/projekt/workshop2005.htm>],
- Workshop "DDC-Einsichten und -Aussichten 2007" (SUB Göttingen, Göttingen, March 1, 2007) [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2836.pdf,
http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2837.pdf],
- DDC-DACHS (D= Deutschland, A= Österreich, CH= Schweiz, S= Südtirol) [<http://www.oegdi.at/DDC-DACHS/>].

⁴ E-mail communication between Magda Heiner-Freiling and Ulrike Reiner (January 23, 2007) concerning the planning of the workshop "DDC-Einsichten und -Aussichten 2007" in Göttingen.

DDC results. Presenting an integrated view of the DDC-based search in the data of the German National Bibliography, this paper summarizes the progress of the VZG project Colibri (*CO*ntext generation and *L*inguistic tools for *B*ibliographic *R*etrieval *I*nterfaces). The project's work is conducted at the Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG), Göttingen. After an outline of the use of the DDC in the German National Bibliography in section 2, we will introduce the DDC-based search in section 3. Section 4 contains a description of the DNB test data. The paper closes with the current results and prospects in section 5.

2. The Use of the DDC in the German National Bibliography

The data for the German National Bibliography is created by the DNB and covers German books, serials, maps, government documents, dissertations, etc. It is published in Germany (in machine-readable form since 1966)⁵ and is updated quarterly. The German National Bibliography includes, among others, Series A (monographs and periodicals from the publishers' book trade), B (monographs and periodicals from outside the publishers' book trade), C (maps), and H (university publications, dissertations and postdoctoral theses from German universities, and dissertations and postdoctoral theses in German language from foreign universities, regardless of their form of publication).⁶ The printed versions can be obtained via the book trade.⁷ Electronic versions of the German National Bibliography are also available – e.g., as html or pdf files. Moreover, the German National Bibliography can be searched online using public networks.⁸ On November 22, 2006, the DNB started using the Dewey Decimal Classification to classify all titles for the main series of the German National Bibliography. On the following day, Magda Heiner-Freiling communicated to the OCLC: “It is a huge step forward, and the project DDC German has now become a regular service. Of course it means a real challenge for us as we had classified mainly grey literature, reports and theses in 2006, and now we are confronted with the whole diversity of regular book trade publications. However, we can make use of many Dewey classes now which we never needed before. This shows the possibilities Dewey offers for everyday (non-academic) publications much better.”⁹

3. DDC-based Search

As the DDC has become a regular DNB service, a DDC-based search, e.g., in the data of the German National Bibliography will enhance nowadays' search possibilities. Several steps have to be made to offer a DDC-based search for DDC experts as well as for DDC laypersons. Today's possibilities of library systems have to be improved so that both DDC-based queries

⁵ Deutsche Nationalbibliographie [auf CD-ROM]
[http://www.bsz-bw.de/depot/media/3400000/3421000/3421308/94_0352.html].

⁶ Deutsche Nationalbibliografie [<http://www.d-nb.de/eng/service/zd/dnb.htm>].

⁷ MVB Marketing- und Verlagsservice des Buchhandels GmbH
[<http://mvb-online.de/sixcms/detail.php?id=67394>].

⁸ Deutsche Nationalbibliografie online (Katalogdatenbank ILTIS)
[http://z3950gw.d-nb.de/z3950/zfo_get_file.cgi?fileName=DDB/searchForm.html].

⁹ 025.431: The Dewey blog, November 24, 2006, German National Bibliography
[http://ddc.typepad.com/025431/2006/11/german_national.html].

can be posed and DDC-based answers¹⁰ can be given by the system. For further considerations, we introduce:

bno: BK (Basisklassifikation) number
cap: caption of a DDC number
dno: DDC number
dno_atom: semantically indecomposable string (of symbols) that represents a DDC class
dno_mol: string that is syntactically decomposable into atomic DDC numbers
blsh: British Library Subject Headings
lcsh: Library of Congress Subject Headings
rno: RVK (Regensburger Verbundklassifikation) number
schedno: DDC number of the DDC schedules

Presented in a simplified way, a DDC-based search system looks as follows:

DDC-based queries → <i>vc_colibri_dewey</i> search system (<i>vc_ds</i>) → DDC-based answers
--

Fig. 1 Simple model of a DDC-based search system

The architecture of the system *vc_ds* with its components can be found elsewhere.¹¹ The main components of *vc_ds* are *vc_day* (*vzg_colibri_ddc* number analyzer) and *vc_dcl* (*vzg_colibri_ddc* classification). Here are some examples of DDC-based queries that could be posed to *vc_ds*:

- Q1. Is “616” a *dno_atom* of *dno_mol* “362.196994220092”?
- Q2. What is the *cap* of *dno* “616”?
- Q3. What is the *dno* of *cap* “Diseases”?
- Q4. Which *cap*(s) are contained in *dno_mol* “362.196994220092”?
- Q5. Which *bno*(s) and which *rno*(s) are most comparable with *dno* “T6--94387”?
- Q6. What is one of the *lcsh* of *dno* “616.994”?
- Q7. What is the *dno* of *blsh* “Islam”?
- Q8. Which *dno*(s) has (have) or fits (fit) the title record with the local identifier “DNB985469137”?

While the answers to questions Q1.-Q4. will be produced by the *vc_ds* component *vc_day*¹², for Q5.-Q7. this will be done by the *vc_ds* component *vc_dqa*¹³ and for Q8. by the *vc_ds* component *vc_dcl*¹⁴. We will discuss the answers to questions Q1.-Q8. in section 5 (“VZG-Colibri Results and Prospects”).

¹⁰ DDC-based queries (DDC-based answers) means that elements of the DDC system are part of the queries (answers).

¹¹ [Reiner2007] Ulrike Reiner: Automatische Analyse von DDC-Notationen und DDC-Klassifizierung von GVK-PLUS Titeldatensätzen [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2836.pdf], p. 98. GVK (Gemeinsamer Verbundkatalog des GBV)-PLUS – Common Union Catalogue with Online Contents [<http://gso.gbv.de/DB=2.2/LNG=EN/>].

¹² Stored in *vc_daygram*(s) and/or *vc_dayset*(s).

¹³ The *vc_ds* component *vc_dqa* (*vzg_colibri_ddc* question answering) is not yet contained in the system architecture [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2836.pdf], p. 98. For *vc_dqa*, the *vzg_colibri_ddc* database *vc_DB* is loaded into the database system MySQL. With *vc_dqa*, questions can be answered via the relational database language SQL. Answers are contained in *vc_dbset*.

¹⁴ Stored in *vc_dclset*.

4. Test Data

After the workshop “DDC-Einsichten und -Aussichten 2007”, test data from the German National Library were placed at the disposal of the VZG project Colibri for experimenting purposes. For these data, both *vc_ds* components *vc_day* and *vc_dcl* were updated and tested. Here are the DDC numbers contained in the handout “Instructions for the representation of number components of synthesized DDC numbers”¹⁵:

```
dnbt1_1: 253.092
dnbt1_2: 540016 [DDC22ger]331.11423015118 / 5401 331.11423 / 5402 511.8 /
5403 -T1--015
dnbt1_3: 5400 [DDC22ger]305.89435 / 5401 305.8 / 5403 -T5--943 /
5403 -T6--9435
dnbt1_4: 5400 [DDC22ger]362.196994220092 / 5401 362.19 / 5402 616.994 /
5402 611.22 / 5403 -T1--092
dnbt1_5: 5400 [DDC22ger]914.3595400222 / 5401 914 / 5403 -T1--0222
(eigentlich 00222 aus der internen Anhängetafel bei 913-919) /
5403 -T2--435954
dnbt1_6: 5400 [DDC22ger]920.0432224 / 5401 920 (nicht: 920.0) /
5403 -T2--432224
dnbt1_7: 5400 [DDC22ger]550.604341 / 5401 550 / 5403 -T1--06
(nicht: -T1--060) / -T2--4341
dnbt1_8: 5400 [DDC22ger]631.64 (nicht: 631.640943858, da Stehplatz) /
5401 631.64 / 5403 -T2--43858 (nicht angehängt, aber einzeln
abgelegt)
dnbt2_1: 305.894387
dnbt2_2: 331.11423015115
```

Fig. 2 DDC numbers by [AlexHeiner-FreilingJahnsScheven2007] (cf. footnote 15)

```
in_dnbt1
in_dnbt1_1 # DDC numbers from handout sent via e-mail (H. Alex, M. Heiner-Freiling,
253.092 # Y.Jahns, E. Scheven; May 31, 2007)
in_dnbt1_2
331.11423015118
in_dnbt1_3
305.89435
in_dnbt1_4
362.196994220092
in_dnbt1_5
914.3595400222
in_dnbt1_6
920.0432224
in_dnbt1_7
550.604341
in_dnbt1_8
631.640943858
in_dnbt2
in_dnbt2_1 # DDC numbers sent via e-mail (H. Alex; June 14, 2007)
305.894387
in_dnbt2_2
331.11423015115
```

Fig. 3 Synthesized DDC numbers as input files *in_dnbt1* and *in_dntb2* for *vc_day*

¹⁵ [AlexHeiner-FreilingJahnsScheven2007] “DDC: Richtlinien für die Einzelablage von Notationsbestandteilen”. DNB handout, May 31, 2007. In Fig. 2, the representation of the given synthesized DDC numbers and their components differs slightly from the original; the DDC numbers are numbered as well.

¹⁶ 5400 (synthesized DDC-Notation), 5401 (base number), 5402 (schedule number), and 5403 (add table number) are fields of the Pica format PICA3:

[<http://www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/02Richtlinien/01KatRicht/update23.pdf>].

The DDC numbers given in Fig. 2 were stored in the files *in_dnb1* and *in_dnb2* as input files for *vc_day* as shown in Fig. 3. The results of the automatic analysis of these DDC numbers will be presented in section 5 (“VZG-Colibri Results and Prospects”).

In terms of the classification data, it was initially expected that the DNB could intellectually evaluate the automatic classification results of *vc_dcl* of the representative sample (600 bibliographic title records of the “Online Contents”)¹⁷ that were presented at the workshop “DDC-Einsichten und -Aussichten 2007”¹⁸. However, it turned out that the DNB didn’t have most of the titles contained in the sample in its collections. Moreover, the effort to intellectually classify this sample proved to be too high for them. Instead, the DNB put bibliographic title data on their ftp server for automatic classification by *vc_dcl* and subsequent evaluation by the DNB. These data are 12 weekly/monthly lists of Series A, B, and H of the German National Bibliography (created in November, 2007) in MAB2 format.¹⁹ The test data for *vc_dcl* consist of 25,653 title records (stored in the file *dnb_ABH*). One title record of *dnb_ABH* (its relevant parts) is shown in Fig. 4. Similar to the input files *in_dnb1* and *in_dnb2* for *vc_day*, the title records of *dnb_ABH* are transformed and stored in the file *in_dnb_ABH*. This is done by *vc_cdb2*²⁰ (*vzg colibri_create data base*; “2” stands for DNB data in MAB2 format); the result of this process for the title record given in Fig. 4 is shown in Fig. 5. After the presentation of the DNB input data *in_dnb1*, *in_dnb2*, and *in_dnb_ABH* for *vc_ds*, we will turn to the results obtained by *vc_ds* in the following section.

5. VZG-Colibri Results and Prospects

To this day, questions Q1.-Q8. of section 3 cannot be answered by one single public library, information or Dewey search system. As far as we know, Q2., Q3., Q6., and Q7. can be answered by WebDewey²¹ (Q2., Q3., and Q7. by MelvilClass²²). Q5. can possibly be answered by an experimental (concordance) system, but Q1. and Q4. can only be answered by the *vc_ds* component *vc_day*²³. The only question that can be answered by the DNB information system ILTIS²⁴ is Q8. It is formulated by using the search menu, inserting the

¹⁷ Online Contents [<http://gso.gbv.de/LNG=EN/DB=2.3/>].

¹⁸ [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2836.pdf], p. 79, p. 103: “3. Für die Bestimmung der Qualität der automatischen DDC-Klassifizierung wurde eine repräsentative OLC-Stichprobe ... gezogen und es wurde automatisch ... DDC-klassifiziert. Das Ergebnis sollte vor Fortführung weiterer Arbeiten intellektuell überprüft werden.”

¹⁹ DNB MAB, MAB2 [<http://www.d-nb.de/standardisierung/formate/mab.htm>]. At present, the program component *vc_dcl* takes the following MAB2 fields into consideration: 001 (ID number of the title record), 026 (regional ID number), 100 (personal names), 310 (title proper in heading form), 331 (title proper in descriptive or hybrid form), 335 (remainder of title), 341 (1st parallel title in descriptive or hybrid form), 410 (place(s) of 1st publication, printing, etc.), 412 (name of 1st publisher, printer, etc.), 451 (1st series title), 540 (ISBN), 542 (ISSN), 700 (notation of a classification system), 705 (DDC analytical), 902 (chain link of 1st subject heading chain), and 907 with subfields s (topical heading) and g (geographical/ethnographical heading).

²⁰ Each (GVK, OLC, DNB) title record is transformed in the same way: title records in Pica format (GVK, OLC) with the *vc_ds* program component *vc_cdb* and title records in MAB2 format (DNB) with the *vc_ds* program component *vc_cdb2*. All GVK title records containing DDC number(s) contribute to the *vc_DB* (*vzg colibri_ddc DataBase*). For matching processes, the representation of each title record is the same as in *vc_KB* (*vzg colibri_ddc Knowledge Base*): *dno SEP descr SEP descr_val* (*dno*: DDC number; *descr*: descriptor; *descr_val*: descriptor value; *SEP*: field separator). If a title record has no *dno*, “XXX” is stored instead.

²¹ WebDewey [<http://connexion.oclc.org/>].

²² MelvilClass [<http://melvil.d-nb.de/melvilclass/>].

²³ To our knowledge, Liu’s DND (Dewey Number Decomposer) is not available. Moreover, DND could only decompose DDC numbers of main class 700 [cf. <http://taipan.dyndns.org/~ul/colibri03-05-04-11.pdf>], p. 9-42.

²⁴ “ILTIS ist das Integrierte Literatur-, Tonträger- und Musikalien-Informationssystem der Deutschen National-

```

001 985469137
...
026 DNB985469137
...
037beng
...
100 Waardenburg, Jean Jacques
102a119476568
331 Muslims as actors
335 Islamic meanings and Muslim interpretations in the perspective of the
study of religions
359 Jacques Waardenburg
410 Berlin ; New York
412 adea Gruyter
...
451 Religion and reason ; Vol. 46
...
501 Literaturverz. S. 398 - 459
540aISBN 978-3-11-019142-4 Gewebe : EUR 98.00 (freier Pr.), ca. sfr 157.00
(freier Pr.)
...
700 |290ÎDNB
705aa297.2c297.2eDDC22ger
902s 4040921-1 Muslim
902s1 4049426-3 Religionswissenschaft
902s2 4162468-3 Islamwissenschaft
903 231Î321

```

Fig. 4 Title record (with ID 985469137) of series A of the German National Bibliography

```

297.2ÿ<026>ÿDNB985469137
297.2ÿ<100>ÿjean#wardenburg
331_line Muslims as actors
335_line Islamic meanings and Muslim interpretations in the perspective of
the study of religions
297.2ÿ<331>ÿmuslims
297.2ÿ<331>ÿactors
297.2ÿ<335>ÿmuslim
297.2ÿ<335>ÿinterpretations
297.2ÿ<335>ÿperspective
297.2ÿ<335>ÿstudy
297.2ÿ<335>ÿreligions
297.2ÿ<335>ÿislamic
297.2ÿ<335>ÿmeanings
297.2ÿ<412@410>ÿ<033A>-adea gruyter@berlin
451_line Religion and reason ; Vol. 46
297.2ÿ<451>ÿvol.
297.2ÿ<451>ÿreligion
297.2ÿ<451>ÿreason
297.2ÿ<540a>ÿ3-11-019142-3
700_line: {290}
297.2ÿ<902s>ÿmuslim
297.2ÿ<902s1>ÿreligionswissenschaft
297.2ÿ<902s2>ÿislamwissenschaft

```

Fig. 5 A sample input file (in_dnb_ABH_985469137) for vc_dcl

bibliothek. In seinem Zentralkatalog findet man neben den Normdaten ca. 5 Millionen Titel der Deutschen Nationalbibliographie seit 1945."

value “985469137”, and choosing the search field “Lokale Identifikationsnummer”. From the ILTIS system response, the DDC number “297.2” can be obtained as answer to Q8.²⁵:

1.	<p>Waardenburg, Jean Jacques: Muslims as actors : Islamic meanings and Muslim interpretations in the perspective of the study of religions / Jacques Waardenburg. - Berlin ; New York : de Gruyter, 2007. - XVII, 471 S. ; 24 cm (Religion and reason ; Vol. 46) Inhaltstext Literaturverz. S. 398 – 459 ISBN 978-3-11-019142-4 Gewebe : EUR 98.00 (freier Pr.), ca. sfr 157.00 (freier Pr.) ISBN 3-11-019142-3 Gewebe : EUR 98.00 (freier Pr.), ca. sfr 157.00 (freier Pr.)</p> <p>SW: Muslim ; Religionswissenschaft ; Islamwissenschaft SG: 290 DDC: 297.2 Signatur: F 2007 A 72116 IDN: 985469137</p>
----	---

We will return to Q8. later. The answers to Q1.-Q4. can be derived from the results of *vc_day* from the various output formats as *vc_dayset_fine*, *vc_dayset_dno_cap*, and *vc_daygram*. Here, the interested reader will find details.²⁶ In the following figures, the analysis results of both input files *in_dnbt1* and *in_dnbt2* as *vc_dayset_mab2* (Fig. 6) and *vc_dayset_fine* (Fig. 7) are given:

dnbt1_1:	705a	\$a	253.092	\$p	253	\$f	092
dnbt1_2:	705a	\$a	331.11423015118	\$p	331.11423	\$c	T1--015 \$d 511.8 \$f 015118
dnbt1_3:	705a	\$a	305.89435	\$p	305.8	\$c	305.8;T5--943 \$l 9435 \$m 9435
dnbt1_4:	705a	\$a	362.196994220092	\$p	362.19	\$c	362.19;616.994 \$d 616.99422;611.22 \$t 618.2+0092
dnbt1_5:	705a	\$a	914.3595400222	\$p	91	\$c	91 \$g 435954 \$t 913-919+00222
dnbt1_6:	705a	\$a	920.04322224	\$p	920	\$c	920.0 \$g 432224
dnbt1_7:	705a	\$a	550.604341	\$p	55	\$c	T1--060 \$f 0604 \$g 4341
dnbt1_8:	705a	\$a	631.640943858	\$p	631.64	\$c	T1--09 \$f 094 \$g 43858
dnbt2_1:	705a	\$a	305.894387	\$p	305.8	\$c	305.8;T5--943 \$l 94387 \$m 94387
dnbt2_2:	705a	\$a	331.11423015115	\$p	331.11423	\$c	T1--015 \$d 511.5 \$f 0151

Fig. 6 Analysis result of *in_dnbt1* and *in_dnbt2* as *vc_dayset_mab2* with its subfields²⁷

²⁵ Deutsche Nationalbibliografie online (Katalogdatenbank ILTIS) [http://z3950gw.d-nb.de/z3950/zfo_get_file.cgi?fileName=DDB/searchForm.html]. Slight changes in the presentation of the results were made (e.g., deleted output lines).

²⁶ Ulrike Reiner: “Automatic Analysis of Dewey Decimal Classification Notations”. Full paper based on a presentation [<http://archiv.tu-chemnitz.de/pub/2007/0139/>] that will appear in “Data Analysis, Machine Learning and Applications. Proceedings of the 31st Annual Conference of the Gesellschaft für Klassifikation e.V., Albert-Ludwigs-Universität Freiburg, March 7-9, 2007”. [<http://www.springer.com/computer/artificial/book/978-3-540-78239-1>]

²⁷ MAB2 Online-Kurzreferenz-Version. Juni 2006 [<http://www.d-nb.de/standardisierung/txt/erw-mab.txt>]. \$a: synthesized DDC number; \$c: base number; \$d: schedule number; \$f: table1 number; \$g: table2 number; \$l: table5 number; \$m: table6 number; \$t: add table number. \$p (prefix dno) (cf. [Reiner2007], p. 21).

```

dnbt1_1:253.092;200;250;253;253.09;253.092;T1--09;T1-092
dnbt1_2:331.11423015118;300;330;331;331.1;331.11;331.114;331.1142;
      331.11423;T1--01;T1--015;T1--0151;T1--015118;510;511;511.8
dnbt1_3:305.89435;300;305;305.8;305.89;T5--9;T5--94;T5--943;T5--9435;
      T6--9435
dnbt1_4:362.196994220092;300;360;362;362.1;362.19;362.196;362.1969;
      362.196994;362.1969942;616;616.9;616.99;616.994;616.9942;
      616.99422;611.2;611.22;618.2+009;618.2+0092
dnbt1_5:914.3595400222;900;910;914;T2--4;T2--43;T2--435;T2--4359;T2--43595;
      T2--435954;913-919+002;913-919+0022;913-919+00222
dnbt1_6:920.0432224;900;920;920.04;T2--4;T2--43;T2--432;T2--4322;T2--43222;
      T2--432224
dnbt1_7:550.604341;500;550;T1--06;T1--0604;T2--4;T2--43;T2--434;T2--4341
dnbt1_8:631.640943858;600;630;631;631.6;631.64;T1--09;T1--094;T2--4;T2--43;
      T2--438;T2--4385;T2--43858
dnbt2_1:305.894387;300;305;305.8;305.89;T5--9;T5--94;T5--943;T5--94387;
      T6--9438;T6--94387
dnbt2_2:331.11423015115;300;330;331;331.1;331.11;331.114;331.1142;
      331.11423;T1--01;T1--015;T1--0151;510;511;511.5

```

Fig. 7 Analysis result of *in_dnbt1* and *in_dnbt2* as *vc_dayset_fine* (printed in bold: synthesized DDC numbers given in Fig. 2 with their components)

As an example, three other output formats are also presented here for *in_dnbt1_4*. Fig. 8 shows the *vc_daygram*, Fig. 9 the *vc_dayset_dno_cap*, and Fig. 10 the *vc_dayset_marc21*.²⁸ It should be noted that the format *vc_dayset_fine* contains all *dno_atom(s)* of a *dno_mol* and that the format *vc_dayset_dno_cap* contains all *dno_atom(s)* with their corresponding captions.

```

362.196994220092 <dnbt1_4_to_analyze; length: 16>
3----- Social sciences <hatzen>
36----- Social problems & social services <hatzen>
362----- Social welfare problems and services <hat>
362.1----- Physical illness <hat>
362.19----- Services to patients with specific conditions <hat>
362.196----- Specific conditions <hatrspan:362.196-362.198:362.196>
362.1969----- Communicable diseases--humans,... <hatien>
362.196994----- Cancer--humans,... <hatien>
362.1969942----- Respiratory tract diseases--humans--cancer,... <hatien>
---.--6----- Diseases <balrspan:362.196-362.198:616>
---.--69----- Other diseases <balrspan:362.196-362.198:616.9>
---.--699----- Tumors and miscellaneous communicable diseases
      <balrspan:362.196-362.198:616.99>
---.--6994----- Cancers <balrspan:362.196-362.198:616.994>
---.--69942----- Respiratory tract diseases--humans--cancer--medicine,...
      <balrspan:362.196-362.198:616.9942>
---.--699422----- Laryngeal neoplasms--medicine,...
      <balrspan:362.196-362.198:616.99422>
---.--6994----- Cancers <nfa->618.1-618.8>
---.-----2----- Respiratory organs <nalrspan:616.9942-616.9949:611.2>
---.-----22----- Larynx <nalrspan:616.9942-616.9949:611.22>
---.-----009- Historical, geographic, persons treatment
      <nalrspan:616.9942-616.9949:618.2+009>
---.-----0092 Persons <nalrspan:616.9942-616.9949:618.2+0092>

```

Fig. 8 Analysis result of *in_dnbt1_4* as *vc_daygram*

²⁸ [MARC21] MARC PROPOSAL NO. 2008-01: Representation of the Dewey Decimal Classification (DDC) System in MARC 21 formats [<http://www.loc.gov/marc/marbi/2008/2008-01.html>].


```

300;Social sciences
360;Social problems & social services
362;Social welfare problems and services
362.1;Physical illness
362.19;Services to patients with specific conditions
362.196;Specific conditions
362.1969;Communicable diseases--humans,...
362.196994;Cancer--humans,...
362.1969942;Respiratory tract diseases--humans--cancer,...
616;Diseases
616.9;Other diseases
616.99;Tumors and miscellaneous communicable diseases
616.994;Cancers
616.9942;Respiratory tract diseases--humans--cancer--medicine,...
616.99422;Laryngeal neoplasms--medicine,...
611.2;Respiratory organs
611.22;Larynx
618.2+009;Historical, geographic, persons treatment
618.2+0092;Persons

```

Fig. 9 Analysis result of *in_dnbtl_4* as *vc_dayset_dno_cap*

```

082 0# $8 1 $a 362.196994220092 $2 22
085 0# $8 1.1 $b 362.19 $a 362.196 $c 362.198 $r 61 $s 6994
085 0# $8 1.2 $p 362.196694 $b 616.994 $a 616.9942 $c 616.9949 $r 611 $s 22
085 0# $8 1.3 $p 362.19699422 $a 616.9942 $c616.9949 $w 618.1 $c 618.8
      $t 0092

```

Fig. 10 Analysis result of *in_dnbtl_4* as *vc_dayset_marc2l*

```

number of ddc-classified title:      1
identifier (dno,schedno):            DNB985469137 (297.2,297.2)
DNB DDC notation (MAB2 field 700)   {290}
calculated cutoff value:             104
title:                               Muslims as actors
title (remainder):                  Islamic meanings and Muslim
interpretations in the perspective of the study of religions
title (series):                      Religion and reason ; Vol. 46
considered descriptor values:        |17| {<100>-jean#waardenburg[0], <331>-
muslims[467], <331>-actors[539], <335>-muslim[646], <335>-
interpretations[657], <335>-perspective[3851], <335>-study[9307], <335>-
religions[760], <335>-islamic[778], <335>-meanings[417], <412@410>-<033A>-
^ade^a gruyter@berlin[0], <451>-vol.[3], <451>-religion[4117], <451>-
reason[882], <540a>-3-11-019142-3[0], <902s1>-religionswissenschaft[104],
<902s2>-islamwissenschaft[17]}
matched descriptor values:           |2| {religionswissenschaft,
                                       islamwissenschaft}
max. match value of matched descriptor values: |2|
calculated1 ddc classes (subdiv):    |3| {200.92, 290, 297.2}
calculated1 ddc classes (sections):  |3| {200, 290, 297}
calculated1 ddc classes (main):      |1| {200}
calculated2 ddc classes (subdiv):    {200.92[1], 297.2[1]}
calculated2 ddc classes (sections):  {200[1], 290[1], 297[1]}
calculated2 ddc classes (divisions): {290[2]}
calculated2 ddc classes (main):      {200[3]}
correlation (DNB985469137,297.2):  111.1xx xxx xxx xxx (1)

```

Fig. 11 One *vc_dcl* classification result (of input file *in_dnb_DNB985469137*)

With the information given above, we are now able to answer questions Q1.-Q8. The answer to Q1. is “yes”, to Q2. “Diseases”, to Q3. “616”, and to Q4. “Social sciences, Social

problems & social services, Social welfare problems and services, ..., Persons". For Q5., the most specific *bno* comparable with *dno* T6--94387 (Tatar) that can be found in the "Basis-klassifikation"²⁹ is "18.87" (Türkische Sprachen und Literaturen). In the "Regensburger Verbundklassifikation"³⁰, the *rno(s)* "AH 64800" (Tatarisch/Wörterbuch), "EH 6440" (Tatarisch (Wolga-, Kasan-Türkisch)), KU 8773 (Tatarka, Dominik), and LS 38900 (Tataren) are most comparable. Q6. is answered with WebDewey, where the *lcs* "Cancer" is given as one LC Subject Heading of the "Terms" at 616.994. Browsing the Relative Index of WebDewey with the *blsh* "Islam" leads to *dno* 297, the answer to Q7. The answer to Q8. "{200.92[1], 297.2[1]}" can be derived from the output line "calculated2 ddc classes (subdiv):" of Fig. 11. For the model and algorithm of the automatic classification process, we refer to [Reiner2007], pp. 79-96. All title records of *in_dnb_ABH* have been automatically DDC-classified with *vc_dcl*³¹ – 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)³² 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). Here are some examples of last (correlation) lines in DDC-classified title records:

```
correlation (DNB970909225,733.30938): 000.000 00x xxx xxx (0)
correlation (DNB973114088,613):      110.xxx xxx xxx xxx (0.666667)
correlation (DNB985469137,297.2):   111.1xx xxx xxx xxx (1)
correlation (DNB985215968,X):       xxx.xxx xxx xxx xxx (X)
```

In each "correlation line", the DNB identifier and the *dno_i* are contained in parentheses. What follows is a "correlation pattern" "CP" and then a "correlation number" "CN" (or an "x", if there is no *dno_i*), also enclosed in parentheses. In CP, a digit-for-digit comparison between *dno_i* (e.g., 297.2) and *dno(s)_a* (e.g., 200.92 and 297.2) from left to right is made. Possible values of CP at each position are "1" (same digit), "0" (different digit), and "x" (end of *dno_i*). 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". Depending on the length L of *dno_i* with CN value 1, the distribution is: 5% (L1), 52% (L2), 13% (L3), 12% (L4), 7% (L5), 7% (L6), 2% (L7), 1% (L8), 1% (L9). This outline of the results of *vc_dcl* and *vc_day* should be discussed.

Prospects. This paper is dedicated to Magda and her work on the DDC. Here, we presented a model of a DDC-based search and the current results of the *vzg colibri_dewey* search system *vc_ds*. We will further enhance and improve *vc_ds* in Magda's sense.

Acknowledgements. I am grateful to Reiner Diedrichs (for his support and trust in my work), Siegfried Kalb (for the latest DDC data of the GVK), the network team Heinz Köhntopp, Günter Becker, and Dietrich Sahlbach (for his prompt upgrade of the computer hardware due to an urgent need for the automatic DDC classification), Magda Heiner-Freiling (for introducing me to the DDC community), Dr. Heidrun Alex, Claudia Werner (for the data of the German National Bibliography), Prof. Erhard Konrad (for his continuous scientific support and for proofreading this text), Martin Fischer (for the English proofreading of this text), and Alfred Vogelbacher.

²⁹ Basisklassifikation 3., erweiterte Ausgabe 2000 [http://www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/05Sacherschliessung/Sacherschliessung_0448.pdf].

³⁰ RVK-Online -- WWW-Version (als Baum) [http://www.bibliothek.uni-regensburg.de/rvko_neu/].

³¹ Hardware platform: HP Proliant DL585 G1, 4x AMD Opteron 275, 2.2 GHz, 16 GB memory. Software: SuSE Linux Enterprise 10, GNU awk 3.1.5, *vc_dcl* server and client (1230 gawk lines of code). 24,642 title records are classified in 74 min. (2 min. reading time for the inverted database *vc_IDB* into the main memory).

³² DDC-like number: subject group based on the divisions of the DDC.