32nd Annual Conference of the German Classification Society
Joint Conference with the British Classification Society and
the Dutch/Flemish Classification Society –
AG BIB:
Subject Indexing 2008: Accept Progress!
Hamburg, July 16 – 17, 2008

Computer-aided Assignment of DDC Numbers

(Computergestütze Zuweisung von DDC-Notationen)

July 17, 2008

Ulrike Reiner
Verbundzentrale des GBV (VZG), Göttingen

Contents

Preface

Assignment of DDC* numbers to

- Bibliographic title records
- Non-DDC terms
- Atomic/molecular DDC numbers

Computer aid: DDC Search System

- State of the Art
- Continuation

[* DDC: Dewey Decimal Classification]

Preface (1)

AG BIB:
Subject Indexing 2008: Accept Progress!

"14.00 Uhr Werkstattbericht: DDC automatisieren
Dr. Ulrike Reiner, Göttingen (angefragt)"

[http://www.ub.uni-dortmund.de/listen/inetbib/msg36527.html]

Computer-aided Assignment of DDC Numbers

[http://www.ub.uni-dortmund.de/listen/inetbib/msg36838.html]

VZG Project Colibri/DDC (Research and Development)

025.4310285 Dewey Decimal Classification--
Data processing Computer applications
025.431072 Dewey Decimal Classification--
Research; statistical methods

Preface (2)

VZG Project Colibri/DDC

Quebec, QC, Canada to Kingston, ON, Canada - Google Maps - Mozilla Firefox

Datei Bearbeiten Ansicht Chronik Lesezeichen Extras Hilfe

http://maps.google.de/maps?hl=en&tab=wl

Web Images Maps News Shopping Mail more ▼ Sign in | Help

Google Maps Deutschland

Start address e.g., "Erfurt": quebec city, ca

End address e.g., "Leipzig": kingston, ca

Get Directions

Search the map | Find businesses | Get directions

Search Results | My Maps

Print | Send | Link to this page

Avoid highways | Get reverse directions

From: Quebec, QC Canada — Edit

Drive: 541 km – about 5 hours 43 mins

1.	Head **west** on **Côte D'Abraham** toward **Escalier de la Chapelle**	0.3 km
2.	Turn **right** at **Rue de la Couronne**	0.2 km
3.	Turn **left** at **Boulevard Charest E**	4.8 km
4.	Continue on **AUT-440 O**	2.1 km
5.	Take exit **12** for **Aut. Henri IV S./AUT-73 S**	1.4 km
6.	Merge onto **AUT-73 S**	6.8 km
7.	Take exit **131-O** to merge onto **AUT-20 O/ Autoroute Transcanadienne O** toward **Montréal**	223 km
8.	Continue on **AUT-25 N/Autoroute Transcanadienne O** (signs for **AUT-40/ Tunnel Louis-H.-La Fontaine/AUT-25 N/ Aéroport Mirabel/Montréal**)	7.8 km
9.	Take exit **8-O** on the **left** to merge onto **AUT-40 O/Autoroute Transcanadienne O**	48.3 km

More... | Map | Satellite | Terrain

Québec, Shawinigan, Grand-Remous, Trois-Rivières, Thetford Mines, St-Georges, Terrebonne, Montréal, Ottawa, Sherbrooke, Maine, Vermont, Brockville, South Frontenac, Kingston, Waterville, Lewiston, Watertown, 100 km, 50 mi

©2008 Google - Map data ©2008 NAVTEQ™ - Terms of Use

Suchen: 850 | Abwärts | Aufwärts | Hervorheben | Groß-/Kleinschreibung

Fertig

Start

Preface (3)

"World Library and Information Congress:
74th IFLA* General Conference and Council

Libraries without borders:
Navigating towards global understanding
10-14 August 2008, Québec, Canada"

[* International Federation of Library Associations and Institutions,
http://www.ifla.org/IV/ifla74/index.htm,
http://www.ifla.org/IV/ifla74/2008ifla_logo.jpg]

Preface (4)

Fundy National Parc
25 May 2008 by Flo

VZG Project Colibri/DDC

VZG

Preface (5)


Fundy National Parc
25 May 2008 by Flo

[http://research.cs.queensu.ca/uc07/img/poster_big.jpg]

Proceedings of the UC 2007 in Kingston, CA

Unconventional Computation
6th International Conference, UC 2007, **Kingston, Canada**, August 13 - 17, 2007, Proceedings
Series: Lecture Notes in Computer Science, Vol. 4618
Sublibrary: Theoretical Computer Science and General Issues
Akl, S.G.; Calude, C.S.; Dinneen, M.J.; Rozenberg, G.; Wareham, H.T. (eds.)
2007, X, 243 p. with online files/update, softcover
ISBN: 978-3-540-73553-3

Keywords:
DNA computing, algorithms, ant colony optimization, approximation, authentication, biomolecular computing, cellular automata, chaos, computational models, computing theory, genetic algorithms, heuristic algorithms, local search, mebrane computing, meta-algorithmics, natural computing, neural network, optimization, quantum computing, quantum key distribution, theoretical computer science, theoretical informatics, theory of computation

[http://www.springer.com/computer/foundations/book/978-3-540-73553-3]

Bibliographic Title Record of the UC 2007 (1) (MAB2 format)


```
001 984632514
002a20070614
...
026 DNB984632514
...
037beng
...
070 1145
070aDNB
070b9999
100bAkl, Selim G. ª[Hrsg.]ª
102a112655688
200bUC <6, 2007, Kingston, Ontario>
202a6517086-6
331 Unconventional computation
335 6th international conference ; proceedings
359 UC 2007, Kingston, Canada, August 13 - 17, 2007. Selim G. Akl ... (ed.)
410 Berlin ; Heidelberg ; New York
412 Springer
425 2007
425a2007
433 X, 241 S.
...
451 Lecture notes in computer science ; Vol. 4618
...
```

Identification number

Name of 1st additional person involved

Title proper in descriptive or hybrid form

Remainder of title

Place(s) of 1st publication, printing, etc.

Name of 1st publisher, printer, etc.

1st series title in descriptive form

Bibliographic Title Record of the UC 2007 (2) (MAB2 format)


```
...
501 Literaturangaben
540aISBN 978-3-540-73553-3 kart. : EUR 48.15 (freier Pr.), ...
540aISBN 3-540-73553-4 kart. : EUR 48.15 (freier Pr.), ...
551a12088782
553a9783540735533
568 07,N28,0093
574 07,A48,0090
700 |004ÎDNB
705a a006.3 c006.3 eDDC22ger
705a a004.0151 c004.0151 eDDC22ger
902s  4196735-5          Theoretische Informatik
902f11|Kongress
902g11|Kingston <Ontario, 2007>
907s  4455833-8          Soft Computing
907f11|Kongress
907g11|Kingston <Ontario, 2007>
912s  4611085-9          Bioinformatik
912f11|Kongress
912g11|Kingston <Ontario, 2007>
917s  4190671-8          Zellularer Automat
917f11|Kongress
917g11|Kingston <Ontario, 2007>
```

ISBN formally (technically) valid

Notation of a classification system

DDC (Dewey Decimal Classification) analytical (a = Full edition)

Chain link of 1st subject heading chain (s = topical heading, f = form heading, g = geographical / ethnographical heading)

Chain link of 2nd subject heading chain

Chain link of 3rd subject heading chain

Chain link of 4th subject heading chain

Objects

Example of

√ Bibliographic title record

➡ Non-DDC terms (notations, numbers or subject headings of other non-DDC classification systems)

- Atomic/molecular DDC numbers

Non-DDC Terms

	in PICA+
Basisklassifikation (BK)	[045Q]
Theoretische Informatik (54.10)	
Kettenglied einer RSWK-Kette	[041A]
Berechnung	
Einzelschlagwörter	[044K]
Zellularer Automat	
British Library Subject Headings (BLSH)	[044G]
Heuristic programming	
Library of Congress Subject Headings (LCSH)	[044A]
Heuristic programming	
Library of Congress Classification (LCC)	[045A]
QA76.9 (Computer programming)	

Objects

Example of

- √ Bibliographic title record
- √ Non-DDC terms (notations, numbers, or subject headings of other non-DDC classification systems)
- ➡ Atomic/molecular DDC numbers

Atomic/Molecular DDC Numbers

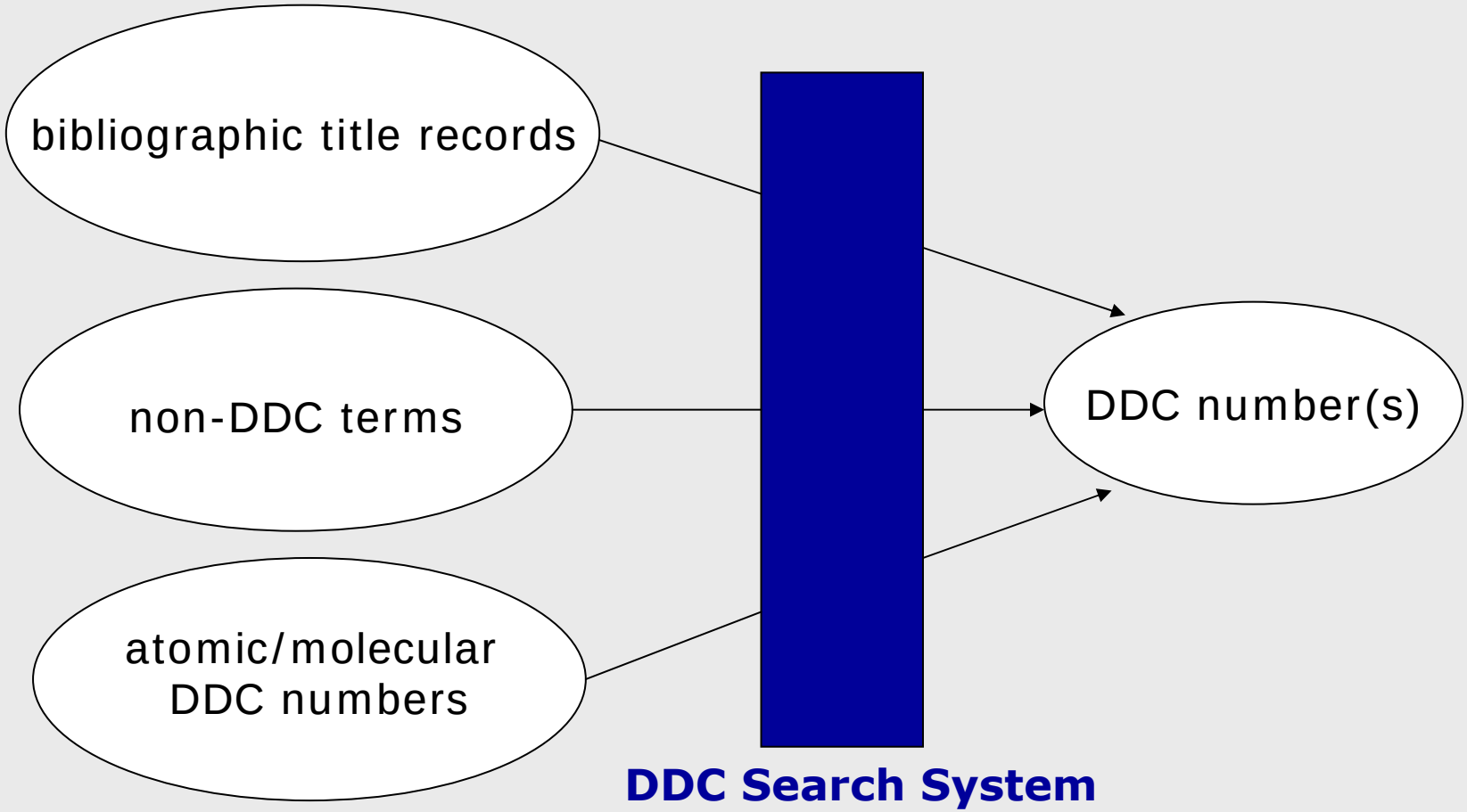
An **atomic DDC number** (**dno_atom**) is a semantically indecomposable string (of symbols) that represents a DDC class.

511.36 (Proof theory and constructive mathematics)

A **molecular DDC number** (**dno_mol**) is a string that is syntactically decomposable into atomic DDC numbers.

511.36028563 (Automatic theorem proving)
[511.36 (Proof theory ... mathematics),
T1--028563 (Artificial intelligence)]

Computer-aided Assignment of DDC Numbers



Assignment of DDC Numbers (1)

with the aid of the

DDC Search System vc_ds

"Unconventional Computation" (DNB984632514)

Genetic programming -- computer science QA76.623

511.36028563

511.36, T1--028563

classification

classification

question answering

analysis

synthesis

vc_dcl

vc_dqa

vc_day

vc_dsy

slides 34 & 35

slide 33

slide 29

slide 14

slide 24

slide 14

slide 28

{005.1[1], 006.3[1]}

{006[4]}

{004.35, 005.1, 006.31, 006.332}

{500, 510, 511, 511.3, 511.36, 511.36028563, T1—02, T1--028, T1—0285, T1--028563}

{511.36028563}

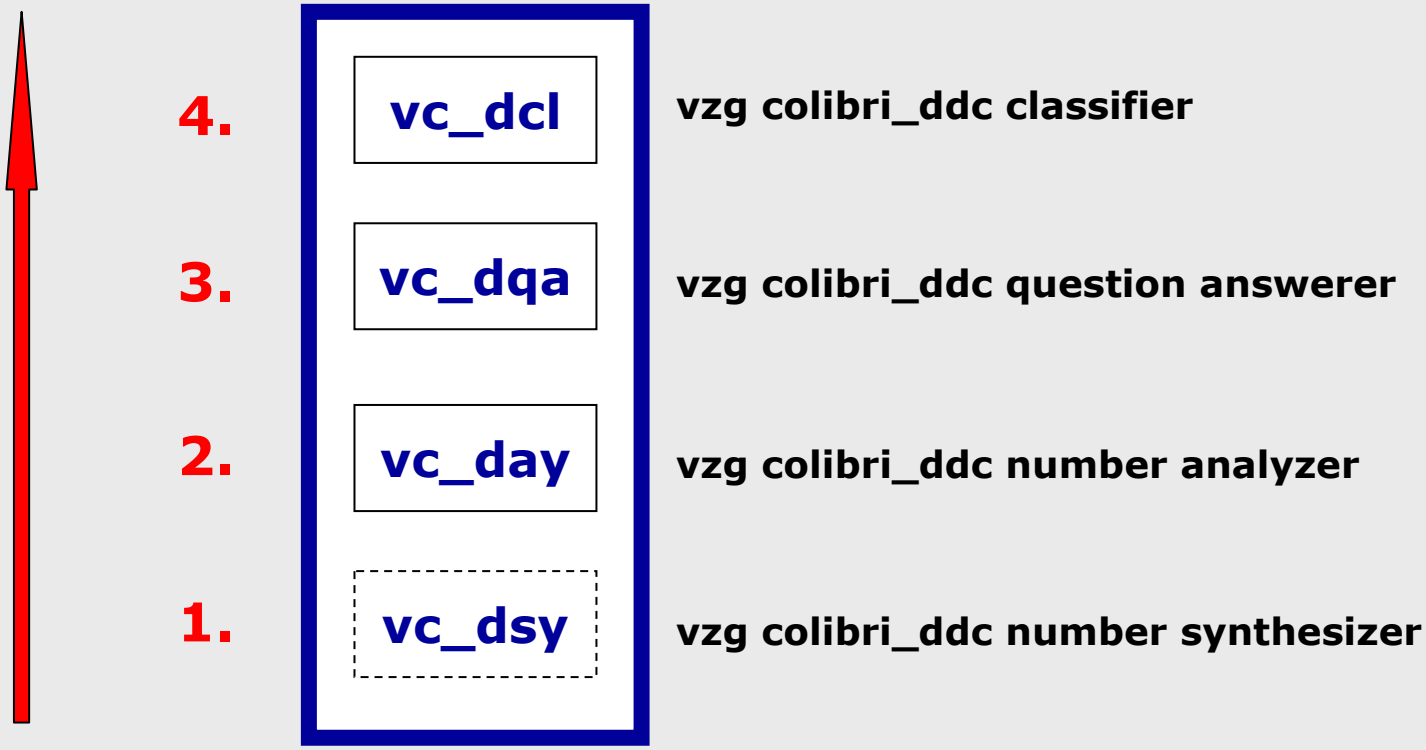
vc_dcl : vzg colibri_ddc classifier
vc_dqa : vzg colibri_ddc question answerer
vc_day : vzg colibri_ddc number analyzer
vc_dsy : vzg colibri_ddc number synthesizer

vc_ds components

Assignment of DDC Numbers (2)

with the aid of the

DDC Search System vc_ds



DDC Notational Synthesis (1)

"The number of devices for synthesis and instructions for their use are so large that no one knows how many million useful DDC numbers can be composed. From an enumerative scheme of limited scope it has grown to be a sophisticated machine for number synthesis." [1]

1,041,073,100 Potential classes (19th ed. of the DDC) [2]

7,705 Instructions (22nd ed. of the DDC) [3]

[1] M.P. Satija: The Theory and Practice of the Dewey Decimal Classification System. Chandos Publishing, Oxford, UK, 2007, p. 11

[2] Estimation by Francis L. Miksa in his book "The DDC, the Universe of Knowledge, and the Post-Modern Library". Forest Press, Albany, New York, 1998, p. 24

[3] [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_3024.pdf , p. 30]

DDC Notational Synthesis (2)

"WebDewey has many additional features, although so far there is no provision for any expert system for the automatic synthesis of numbers following *add to* instructions." [1]

[1] M.P. Satija: The Theory and Practice of the Dewey Decimal Classification System. Chandos Publishing, Oxford, UK, 2007, p. 16

State of the Art of vc_dsy (1)
vzg colibri_ddc number synthesizer

First considerations on a computer-aided number synthesis

DDC number synthesis with the aid of components of the DDC search system (vc_ds):

vc_dsy: user interface for DDC synthesis (not implemented yet)

vc_dqa: determining of dno_atoms (slides 26 & 27) and dno_mols (slide 28)

vc_day: testing the suitability or correctness of dno_mol(s) (slide 24)

State of the Art of vc_dsy (2)
vzg colibri_ddc number synthesizer

Automatic DDC notational synthesis?

Main Title: Wisdom and compassion = Śes rab dan̓ sniñ̓ rje' i rol pa : the sacred art of Tibet / Marylin M. Rhie, Robert A.F. Thurman ; photographs by John Bigelow Taylor.

Subjects: Art, Buddhist--China--Tibet.
Buddhist art and symbolism--China--Tibet.
Art, Tibetan

Note: ... an exhibition organized by the Asian Art Museum of San Francisco in conjunction with Tibet House, New York

=> Common knowledge and DDC knowledge

[http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&IKT=1016&SRT=YOP&TRM=ppn+113549423,
http://lccn.loc.gov/90046899]

State of the Art of vc_dsy (3)
vzg colibri_ddc number synthesizer

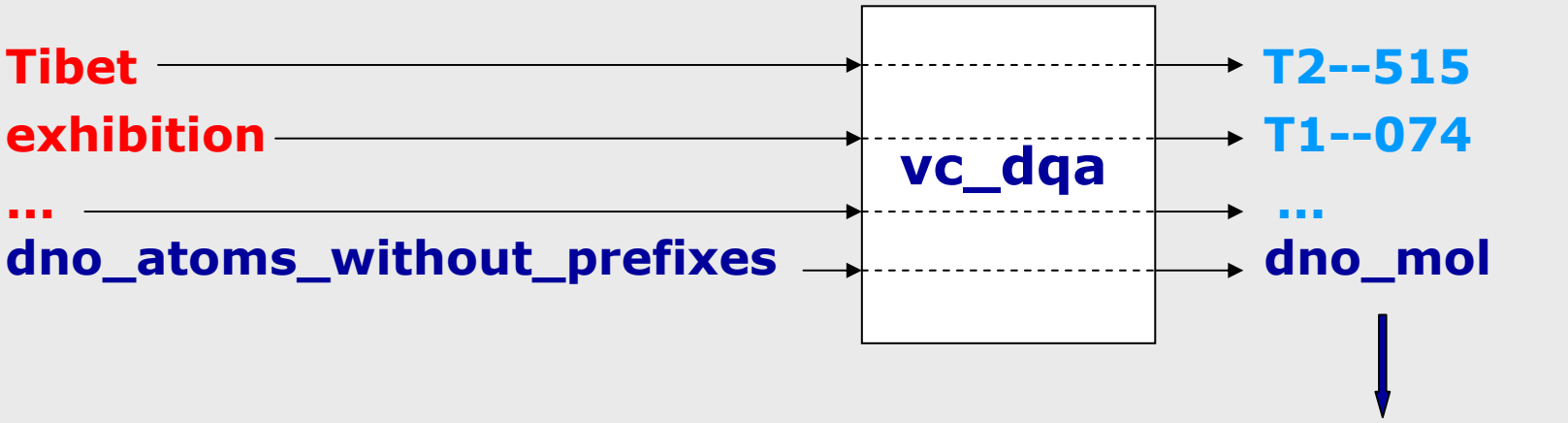
	dno_atoms
Tibet =>	
Historical, geographic, persons treatment	**T1--09**
Tibet Autonomous Region (Xizang Zizhiqu)	**T2--515**
exhibition =>	
Museums, collections, exhibits	**T1--074**
San Francisco, New York =>	
United States	**T2--73**
the sacred art of Tibet ... photographs =>	
Arts	**700**
Iconography	**704.9**
Other religions	**704.9489**
Tibetan =>	
Tibetan Buddhism (Lamaism)	**294.3923**

[red: terms of the bibliographic title record; turquoise: terms of the DDC System]

State of the Art of vc_dsy (4)
vzg colibri_ddc number synthesizer

dno_atoms := set of atomic DDC numbers
dno_mols := set of molecular DDC numbers

=> dno_atoms: {T1--09, T2--515, T1--074, T2--73, 704.9489, 294.3923}



State of the Art of vc_day (1)
vzg colibri_ddc number analyzer

Analysis Diagram of a DDC number (in_liu_417)

```
704.9489439230951507473 <liu_417_to_analyze; length: 23>
7---------------------- Arts & recreation <hatzen>
70--------------------- Arts <hatzen>
704-------------------- Special topics in fine and decorative arts <hat>
704.9------------------ Iconography <hat>
704.94----------------- Specific subjects <hat>
704.948---------------- Religion <hat>
704.9489--------------- Other religions <hat>
704.94894-------------- Indic religions--art representation <hatien>
704.948943------------- Buddhism--art representation <hatien>
---.----4-------------- Religions of Indic origin <na1r1:294>
---.----43------------- Buddhism <na1r1:294.3>
---.----439------------ Branches, sects, reform movements <na1r1:294.39>
---.----4392----------- Mahayana Buddhism (Northern Buddhism) <na1r1:294.392>
---.----43923---------- Tibetan Buddhism (Lamaism) <na1r1:294.3923>
---.---------09-------- Historical, geographic, persons treatment <T1--09>
---.---------095------- Treatment by specific continents, countries, localities;
                         extraterrestrial worlds <T1--095>
---.-----------5------- Asia  OrientFar East <ba4r2span:T1--093-T1--099:T2--5>
---.-----------51------ China and adjacent areas <ba4r2span:T1--093-T1--099:T2--51>
---.-----------515----- Tibet Autonomous Region (Xizang Zizhiqu)
                         <ba4r2span:T1--093-T1--099:T2--515>
---.--------------07--- Museums, collections, exhibits; collecting objects
                         <ba4r2span:T1--093-T1--099+07>
---.--------------074-- Museums, collections, exhibits <ba4r2span:T1--093-T1--099+074>
---.----------------7- North America <ba4r2span:T1--093-T1--099+074:na4r2:T2--7>
---.----------------73 United States <ba4r2span:T1--093-T1--099+074:na4r2:T2--73>
```

State of the Art of vc_day (2)
vzg colibri_ddc number analyzer

Presentation

GfKl 2007, Librarian Workshop, Freiburg

[http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2835.pdf]

Publication

Automatic Analysis of Dewey Decimal Classification Notations

[http://www.springerlink.com/content/l044082243v1l7u6/, pp. 697-704]

State of the Art of vc_dqa (1)
vzg colibri_ddc question answerer

Searching for dno_atoms (1)

```
mysql> select * from dno_kb where (dno like "t%" and descr_val
   like "%tibet%");
+----------+-------+-------------------------------------------------------------+
| dno      | descr | descr_val                                                   |
+----------+-------+-------------------------------------------------------------+
| t2--515  | <hat> | tibet autonomous region (xizang zizhiqu)                    |
| t5--954  | <hat> | tibetans                                                    |
| t6--9541 | <hat> | tibetan                                                     |
| t6--954  | <hat> | tibeto-burman languages                                     |
| t6--95   | <hat> | languages of east and southeast asia    sino-tibetan languages |
+----------+-------+-------------------------------------------------------------+
5 rows in set (0.07 sec)
```

```
mysql> select * from dno_kb where (dno like "t%" and descr_val
   like "%exhibit%");
+----------------------+---------+-----------------------------------------------------------------------------+
| dno                  | descr   | descr_val                                                                   |
+----------------------+---------+-----------------------------------------------------------------------------+
| t1--074              | <hat>   | museums, collections, exhibits                                              |
| t1--0753-t1--0755    | <hat>   | [organizing and preparing collections and exhibits, service to patrons]     |
| t1--08+074           | <ba5>   | museums, collections, exhibits|<na4r2>|074||t2--4-t2--9|                  |
| t1--0901-t1--0905+074| <ba5>   | museums, collections, exhibits|<na4r2>|074||t2--4-t2--9|                  |
| t1--0901-t1--0905+07 | <ba5>   | museums, collections, exhibits                                              |
| t1--093-t1--099+074  | <ba4r2> | museums, collections, exhibits|<na4r2>|074||t2--4-t2--9|                  |
| t1--093-t1--099+07   | <ba4r2> | museums, collections, exhibits                                              |
+----------------------+---------+-----------------------------------------------------------------------------+
7 rows in set (0.07 sec)
```

[for dno_kb (DDC knowledge base), <hat>, <ba5>, ... see p. 29 and p. 32
http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_3024.pdf]

State of the Art of vc_dqa (2)
vzg colibri_ddc question answerer

Searching for dno_atoms (2)

dno_atoms : {T1--09, T2--515, T1--074, T2--73, 704.9489, 294.3923} (slide 22-24 & 26)

```
mysql> select * from dno_kb where dno="704.9489";
+----------+---------+---------------------------+
| dno      | descr   | descr_val                 |
+----------+---------+---------------------------+
| 704.9489 | <hat>   | other religions           |
| 704.9489 | <na1r1> | |704.9489|292-299|29||||| |
| 704.9489 | <ri>    | hell                      |
+----------+---------+---------------------------+
3 rows in set (0.07 sec)
```

stands for

"Add to base number 704.9489 the numbers following 29 in 292-299"

cross-check by

```
mysql> select * from dno_kb where dno="294.3923";
```

=> no rules!

[for the knowledge representation of DDC facts and rules, see p. 32
http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_3024.pdf]

State of the Art of vc_dqa (3)
vzg colibri_ddc question answerer

Searching for dno_mols*

```
mysql> select * from dno_db where (dno like "%73%" and dno like
"%515%" and dno like "%43923%" and dno like "704.9489%" and dno
like "%09%" and dno like "%074%");
```

dno	descr	descr_val
704.9489439230951507473	<001A>	02003
704.9489439230951507473	<003@>	0113549423
704.9489439230951507473	<021A>	compassion
704.9489439230951507473	<021A>	wisdom
704.9489439230951507473	<022A>	france
704.9489439230951507473	<022A>	south
704.9489439230951507473	<028A_da>	marylin m.#rhie
704.9489439230951507473	<028C>	john bigelow#taylor
704.9489439230951507473	<028C>	robert a. f.#thurman
704.9489439230951507473	<033A>	<033A>-royal academy of arts@london
704.9489439230951507473	<044A_a>	art, buddhist
704.9489439230951507473	<044A_a>	art, tibetan
704.9489439230951507473	<044A_a>	buddhist art and symbolism
704.9489439230951507473	<044A_z>	china
704.9489439230951507473	<044A_z>	tibet
704.9489439230951507473	<045A>	N8193.T5
704.9489439230951507473	<145Z_a>	lh 65990
704.9489439230951507473	liu_417_to	length: 23

18 rows in set (0.02 sec)

[Pica+ tags: http://www.gbv.de/vgm/info/mitglieder/02Verbund/01Erschliessung/02Richtlinien/01KatRicht/inhalt.shtml,
* unique dno_mols in GVK: 590,120 (January 2008), 466,134 (July 2004); approx. 3,000 dno_mols/month]

State of the Art of vc_dqa (4)
vzg colibri_ddc question answerer

Searching for dno* with Non-DDC terms

```
mysql> select * from dno_db where descr like "<044A%" and descr_val like "genetic
programming%computer science";
+------------+----------+---------------------------------------+
| dno        | descr    | descr_val                             |
+------------+----------+---------------------------------------+
| 004.1      | <044A_s> | genetic programming computer science |
| 004.35     | <044A_a> | genetic programming computer science |
| 005.1      | <044A_a> | genetic programming computer science |
| 006.31     | <044A_a> | genetic programming computer science |
| 006.31     | <044A_s> | genetic programming computer science |
| 006.332    | <044A_s> | genetic programming computer science |
| 006.3      | <044A_a> | genetic programming computer science |
| 006.3      | <044A_s> | genetic programming computer science |
| 332.015118 | <044A_s> | genetic programming computer science |
| 621.381    | <044A_s> | genetic programming computer science |
+------------+----------+---------------------------------------+
10 rows in set (0.00 sec)
```

LCSH (Library of Congress Subject Headings)

```
mysql> select * from dno_db where descr like "<045A%" and descr_val="qa76.623";
+---------+--------+-----------+
| dno     | descr  | descr_val |
+---------+--------+-----------+
| 004.35  | <045A> | QA76.623  |
| 005.1   | <045A> | QA76.623  |
| 006.31  | <045A> | QA76.623  |
| 006.332 | <045A> | QA76.623  |
+---------+--------+-----------+
4 rows in set (0.00 sec)
```

LCC (Library of Congress Classification)

* dno: DDC number

State of the Art of vc_dcl (1)
vzg colibri_ddc number classifier

Automatic classification

- DDC database **vc_DB**
- DDC knowledge base **vc_KB**

} vc_DB_PLUS (intellectual basis)

- vector product*
  as <u>similarity measure</u>

$$s_{uc} = \sum_{i=1}^{l} u_i c_i$$

between the terms of

DDC-unclassified title records ***u*** (elements of in_dnb_ABH**)
and DDC-classified title records ***c*** (elements of vc_DB_PLUS)

- two heuristic functions: **cutoff_val_dyn**, **cutoff_val_stat**
- two methods to calculate DDC class candidates: **calc1**, **calc2**

[* Gerard Salton: Automatic Information Organization and Retrieval. McGraw-Hill, New York, 1968, p. 237
** input test data (25,653 bibliographic title records) from the German National Library (Deutsche Nationalbibliothek DNB)]

State of the Art of vc_dcl (2)
vzg colibri_ddc number classifier

Evaluation of the automatic classification

- automatic evaluation

two (automatic) correlation measures:

correlation pattern CP, e.g., 110.xxx xxx xxx xxx

correlation number CN ∈ [0,1] , e.g., 0.666667, 1

- (first) intellectual evaluation of vc_dcl (July 3, 2008)

- 11 DNB experts evaluated Sample1*
- "Colibri ist derzeit für eine automatische Klassifizierung von Publikationen insbesondere Netzpublikationen nicht einsetzbar ... Eine Modifizierung des Colibri-Systems vor einem erneuten Testlauf wäre wünschenswert." **

[* 1,000 automatic DDC-classified title records of in_dnb_ABH by vc_dcl)
** Yvonne Jahns; Elisabeth Mödden: Colibri-Test Juni 2008. Automatisches Klassifizieren_DDC_Colibri. L2/F2 AG Netzpublikationen. 3. Juli 2008]

State of the Art of vc_dcl (3)
vzg colibri_ddc number classifier

Accuracy of vc_dcl (cf. slide 28)

```
dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_ppn_113549423
number of ddc-classified title:     2
identifier (dno,schedno):           113549423 (704.9489439230951507473,704.948943)
Data of LoC:                        704.9489439230951507473
calculated cutoff value:            40
title:                              Wisdom and compassionfÃ¢Ses rab dan sÃ¤nin rjeÂ®i
    rol padthe sacred art of TibethMarylin M. Rhie; Robert A.F. Thurman
considered descriptor values:       |12| {<028A_da>-marylin m.#rhie[6], <028C>-robert
    f.#thurman[16], <028C>-john bigelow#taylor[24], <021A>-wisdom[1830], <021A>-
    compassion[274], <033A>-<033A>-royal academy arts@london[47], <044A_a>-tibet[786],
    <044A_a>-buddhist art symbolism[40], <044A_a>-china[4756], <044A_a>-art
    tibetan[20], <044A_a>-art buddhist[44], <045A>-N8193.T5[2]}
matched descriptor values:          |6| {N8193.T5, marylin m.#rhie, john
    bigelow#taylor, art tibetan, robert f.#thurman, buddhist art symbolism}
max. match value of matched descriptor values: |6|
calculated1 ddc classes (subdiv):   |1| {704.948943}
calculated1 ddc classes (sections): |1| {704}
calculated1 ddc classes (main):     |1| {700}
calculated2 ddc classes (subdiv):   {704.948943[1]}
calculated2 ddc classes (sections): {704[1]}
calculated2 ddc classes (divisions):{700[1]}
calculated2 ddc classes (main):     {700[1]}
correlation (113549423,704.948943): 111.111 111 xxx xxx (1)
```

Sorted list of frequencies

111 [FUCUTVAL] S_list: 2-6-16-20-24-40-44-47-274-786-1830-4756

State of the Art of vc_dcl (4)
vzg colibri_ddc number classifier

Assignment of DDC numbers to non-DDC terms by vc_dcl

```
dl385-122 colibri/ul-test> vc_dcl_cli < vc_DB/in_ul_gfkl08_FICTIVE
number of ddc-classified title:     1
identifier (dno,schedno):           FICTIVE (XXX,X)
Data of ???:                        XXX
calculated cutoff value:            11
title:                              Unconventional computation
considered descriptor values:       |1|{<331>-genetic programming computer
   science[11]}
matched descriptor values:          |1|{genetic programming computer science}
max. match value of matched descriptor values: |1|
calculated1 ddc classes (subdiv):   |11| {004.1, 004.35, 005.1, 006.3,
   006.31, 006.32, 006.332, 332, 620.00113, 621.3, 621.381}
calculated1 ddc classes (sections): |3| {004, 005, 006}
calculated1 ddc classes (main):     |1| {000}
calculated2 ddc classes (subdiv):   {004.1[1], 004.35[1], 005.1[1],
   006.31[1], 006.32[1], 006.332[1], 006.3[1]}
calculated2 ddc classes (sections): {006[4]}
calculated2 ddc classes (divisions):{000[7]}
calculated2 ddc classes (main):     {000[7]}
correlation (FICTIVE,X): xxx.xxx xxx xxx xxx (X)
```

State of the Art of vc_dcl (5)
vzg colibri_ddc number classifier

Classification of "Unconventional computation" by vc_dcl (1)

```
number of ddc-classified title:     9131
identifier (dno,schedno):           DNB0984632514 (004.0151,004.0151)
DNB DDC notation (MAB2 field 700)   {004}
calculated cutoff value:            393
title:                              Unconventional computation
title (remainder):                  6th international conference ; proceedings
title (series):                     Lecture notes in computer science ; Vol. 4618
considered descriptor values:       |18| {<100b>-selim#akl[0], <331>-
unconventional[393], <331>-computation[306], <335>-proceedings[12003], <335>-6th[801],
<335>-international[12911], <335>-conference[6787], <412@410>-<033A>-
springer@berlin[750], <451>-computer[2601], <451>-science[7828], <451>-vol.[3], <451>-
lecture[1875], <451>-notes[4125], <540a>-3-540-73553-4[0], <902s>-theoretische
informatik[62], <902g1>-902g11|kingston <ontario 2007>[0], <907s>-soft computing[74],
<907g>-907g11|kingston <ontario 2007>[0]}
matched descriptor values:          |4| {theoretische informatik, unconventional, soft
computing, computation}
max. match value of matched descriptor values: |4|
calculated1 ddc classes (subdiv):   |4| {004, 005.1, 006.3, 511.3}
calculated1 ddc classes (sections): |3| {004, 005, 006}
calculated1 ddc classes (main):     |1| {000}
calculated2 ddc classes (subdiv):   {005.1[1], 006.3[1]}
calculated2 ddc classes (sections): {004[1], 005[1], 006[1]}
calculated2 ddc classes (divisions):{000[3]}
calculated2 ddc classes (main):     {000[3]}
correlation (DNB0984632514,004.0151: 111.000 0xx xxx xxx (0.428571)
```

Programming

Artificial Intelligence

Special computer methods

Computer programming, programs, data

Data processing Computer science

Computer science, information, general works

State of the Art of vc_dcl (6)
vzg colibri_ddc number classifier

Classification of "Unconventional computation" by vc_dcl (2)

```
number of ddc-classified title:     9506
identifier (dno,schedno):           DNB0984632514 (006.3,006.3)
DNB DDC notation (MAB2 field 700):  {004}
DDC notation (MAB2 field 705):      {006.3,004.0151}
calculated cutoff value:            260
title:                              Unconventional computation
title (remainder):                  6th international conference ; proceedings
title (series):                     Lecture notes in computer science ; Vol. 4618
considered descriptor values:       |19| {<100b>-selim#akl[0], <331>-
unconventional[394], <331>-computation[311], <335>-proceedings[12058], <335>-6th[808],
<335>-international[12948], <335>-conference[6840], <412@410>-<033A>-
springer@berlin[750], <451>-computer[2627], <451>-science[7870], <451>-vol.[3], <451>-
lecture[1883], <451>-notes[4163], <540a>-3-540-73553-4[0], <912g11>-kingston[260],
<902s>-theoretische informatik[62], <912s>-bioinformatik[51], <907s>-soft
computing[74], <917s>-zellularer automat[21]}
matched descriptor values:          |5| {kingston, theoretische informatik, soft
computing, bioinformatik, zellularer automat}
max. match value of matched descriptor values: |4|
calculated1 ddc classes (subdiv):   |3| {004, 006.3, 511.3}
calculated1 ddc classes (sections): |2| {004, 006}
calculated1 ddc classes (main):     |1| {000}
calculated2 ddc classes (subdiv):   {006.3[1]}
calculated2 ddc classes (sections): {004[1], 006[1]}
calculated2 ddc classes (divisions):{000[2]}
calculated2 ddc classes (main):     {000[2]}
correlation (DNB0984632514,006.3): 111.1xx xxx xxx xxx (1)
```

State of the Art of vc_dcl (7)
vzg colibri_ddc number classifier

Automatic classification obstacles (1) (extract of *)

- **"A subject may occur in almost any discipline. ... Thus there is theoretically no single class number for any subject."**

- **"Determination of the specific subject of a given document is an art which machines cannot do."**

- "The specific subject of a document may be determined by ... title, subtitle, blurb, preface and table of contents, and scanning through the text ... The institutional affiliation of the author, the index at the back of the book, the series and cited references ... published reviews or consult some reference tools ... a subject expert."

[* M.P. Satija: The Theory and Practice of the Dewey Decimal Classification System. Chandos Publishing, Oxford, UK, 2007, pp. 39; 51; 52]

State of the Art of vc_dcl (8)
vzg colibri_ddc number classifier

Automatic classification obstacles (2) (following *)

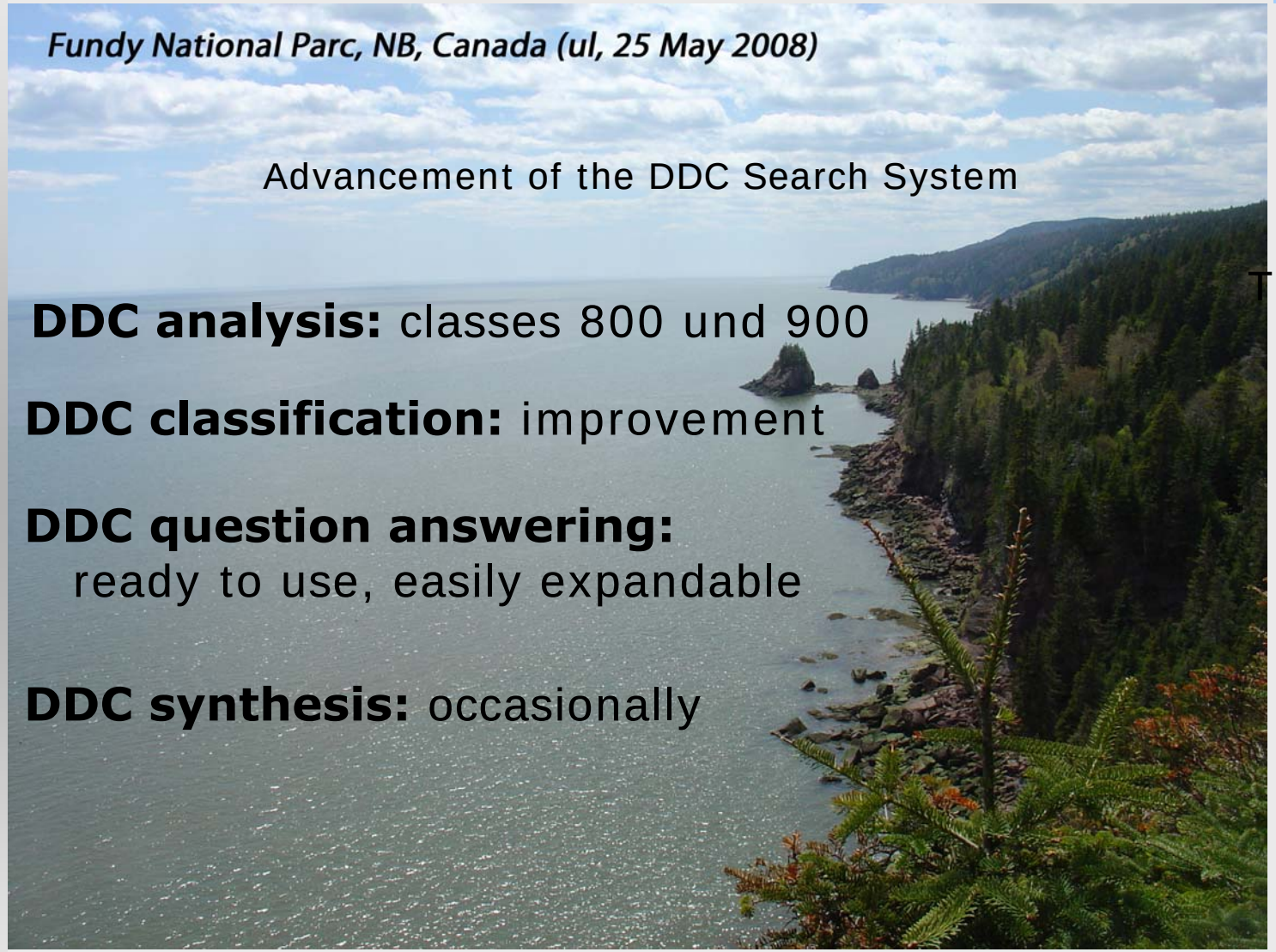
- "... there is too much ambiguity and complexity in the world of publishing and the DDC ..."

- a title could be incomplete, fanciful, or vague
- a title could contain redundant words
- it could be a new subject that has not yet been given a place in the DDC
- obscure subjects, e.g., "travels in transoxiana"
- different terminology, e.g., "German Bundestag", "US Congress"
- the terminology for some concepts differs even between countries with the same language, e.g., gas station (US) - petrol pump (UK)

[* M.P. Satija: The Theory and Practice of the Dewey Decimal Classification System. Chandos Publishing, Oxford, UK, 2007, pp. 59]

Continuation

25 May 2008 by Flo



Advancement of the DDC Search System

DDC analysis: classes 800 und 900

DDC classification: improvement

DDC question answering:
ready to use, easily expandable

DDC synthesis: occasionally

Thank you for your interest in the VZG Project Colibri/DDC!