

Discovering Dewey in Göttingen

EDUG Symposium 2016, 16-04-25

Automatic Analysis of DDC Numbers

based on MARC21

Dr.-Ing. Ulrike Reiner, VZG

Egg or DDC numbers?



333.333

555.55

666.444

2

Content



- **VZG Project Colibri/DDC (ongoing since 2003)**
- **DDC Search System**

Automatic Analysis of DDC Numbers

- » past 2004-2007 (DDC proprietary XML tags)
- » present 2013- (DDC MARC21 XML tags)
- » future ? ... bf:Classification ... (BIBFRAME)

3



*C*ontext generation and
*L*inguistic tools for
*B*ibliographic
*R*etrieval
*I*nterfaces



000 – 020 – 510 – 511.3

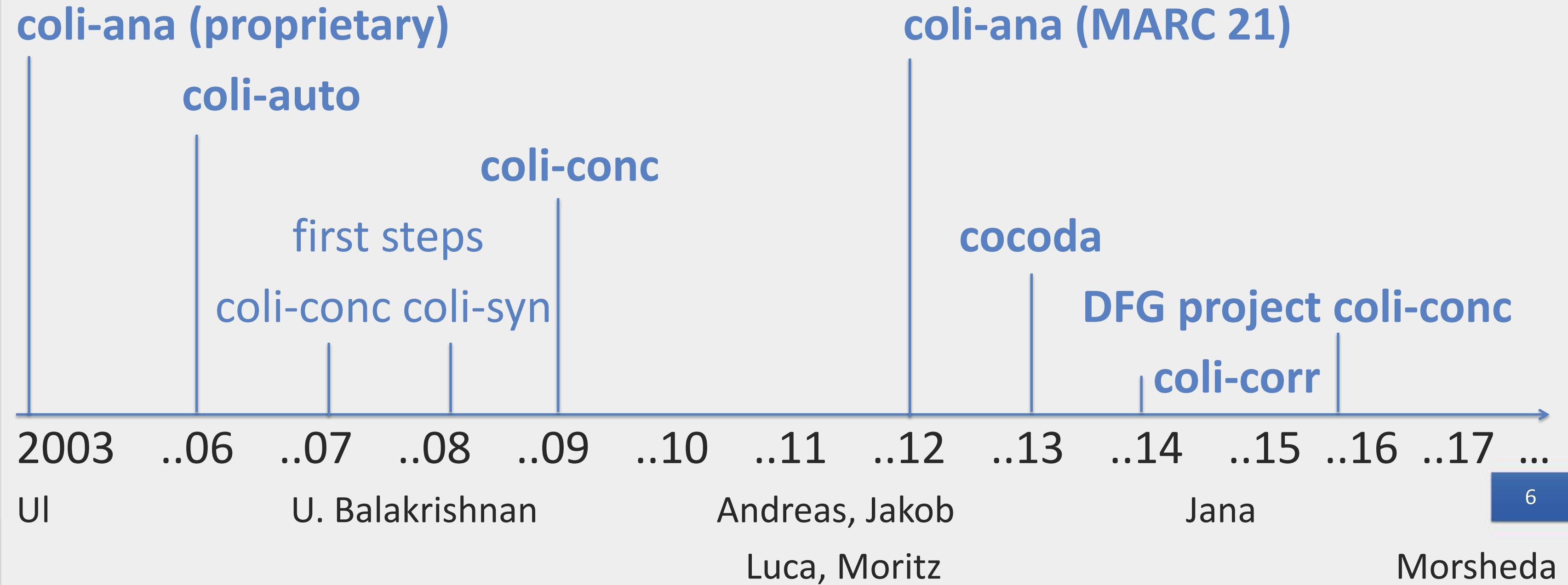
Project Colibri/DDC – subprojects (1)



- automatic **analysis** and **synthesis** of DDC numbers 1
- automatic DDC classification 2
- semi-automatic creation of DDC-EZB, DDC-RVK, ...
concordances 3
- automatic checking of the **correctness** of DDC
numbers 4 5

1 **coli-ana, coli-syn** * 2 **coli-auto** * 3 **coli-conc** * 4 **coli-corr**

Project Colibri/DDC – subprojects (2)



6



DDC search system (vc_ds) - components

DDC based queries → DDC search system → DDC based answers

vzg colibri_ddc number analyzer

vc_day

vzg colibri_ddc number synthesizer

vc_dsy

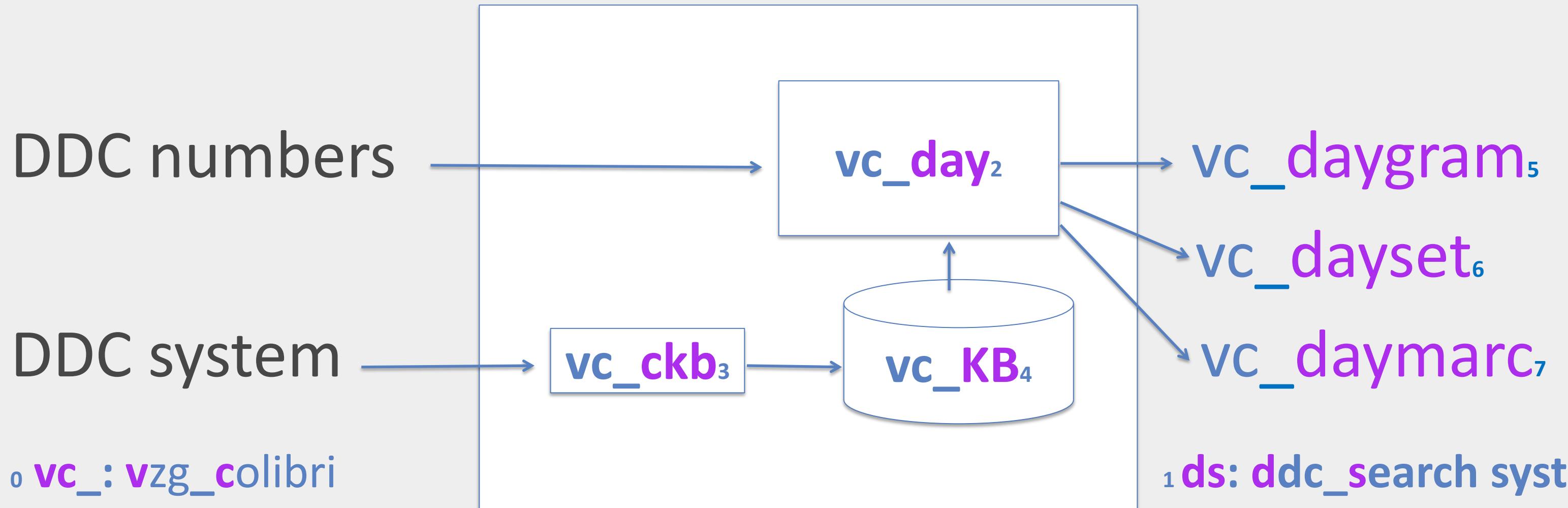
vzg colibri_ddc classifier

vc_dcl

vzg colibri_ddc Q&A component

vc_dqa

DDC search system (**vc₀_ds₁**) - architecture



₀ **vc** : **vzg_colibri**

₂ **day**: **ddc number analyzer**

₅ **ddc analysis diagram**

₃ **ckb**: **create knowledge base**

₆ **ddc analysis result set**

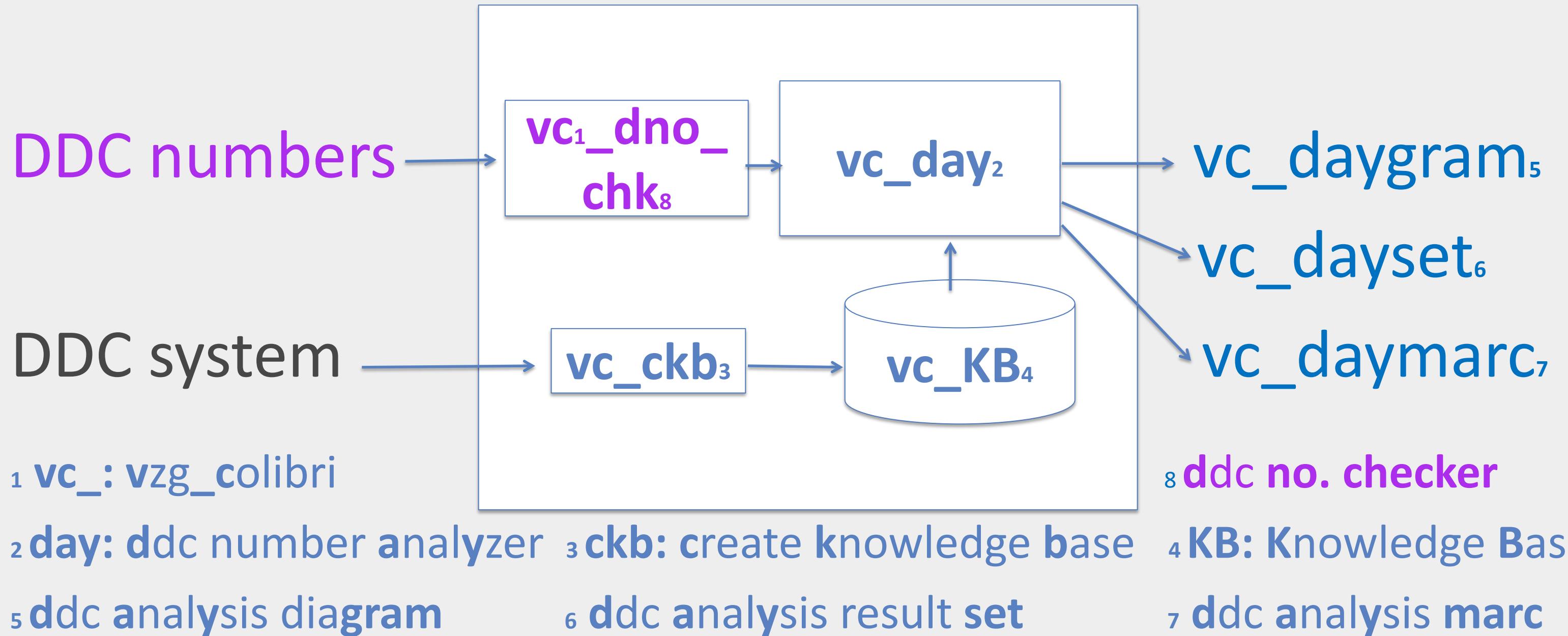
₁ **ds**: **ddc search system**

₄ **KB**: **Knowledge Base**

8

₇ **ddc analysis marc**

DDC search system (vc_ds) - architecture



DDC numbers ? (1)



Checking with ddc number checker ([vc_dno_chk](#))

1. 000 Computer science, information & general works
2. G:nl S:eg Z:30 Megalithic research in the Netherlands ...

GVK: [ppn_626796563](#)

3. 372(.243) Upper level (Optional number; prefer 373.236)
4. (370.973) Education--United States, . . .

GVK: [ppn_314802843](#)

DDC numbers (2)



5. 708.111 Arts & recreation -- Canada -- British Columbia
6. 331.890410263362014098133 (24-digits)
(Tore Brattli: <http://is.gd/nnAxsR>)
7. 704.03917104436109041074434323 (29-digits)
[\(http://classify.oclc.org/classify2/ClassifyDemo?wi=54425207\)](http://classify.oclc.org/classify2/ClassifyDemo?wi=54425207)

DDC numbers (3)



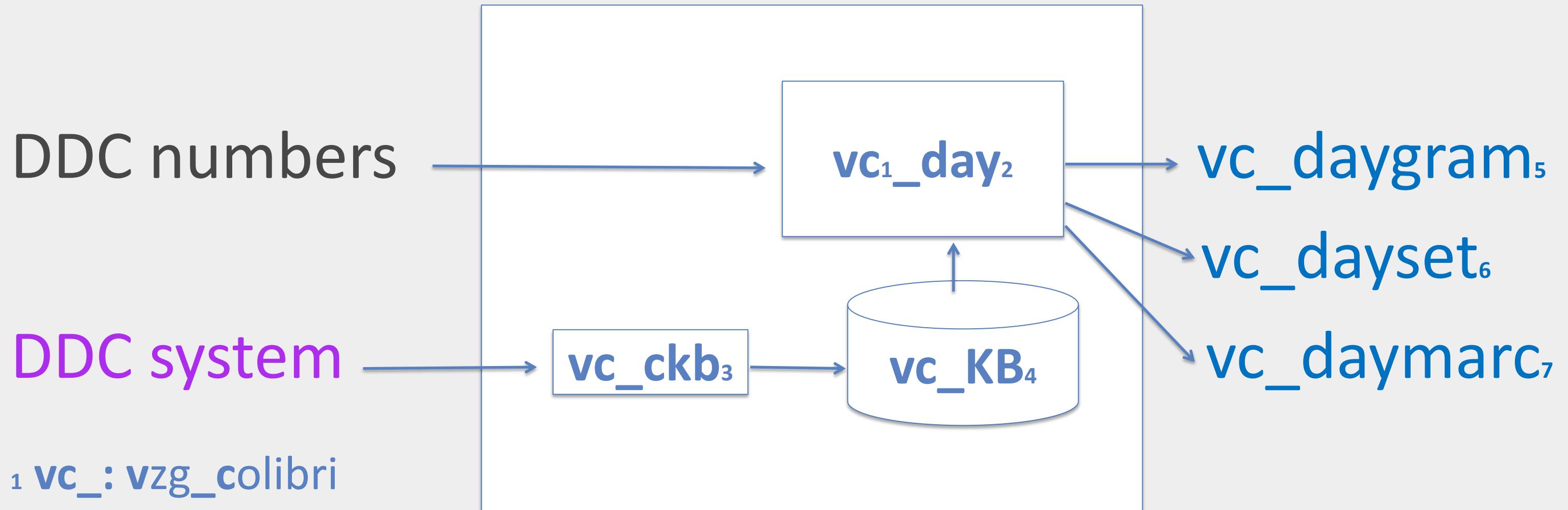
8. 709.0407609436582090511074436582 (31-digits)

[<http://experiment.worldcat.org/entity/work/data/900146701>]

9. 851.4099999999999999999999999287 (30-digits)

(GVK ppn 337713219)

DDC search system (vc_ds) - architecture



¹ vc_ : vzg_colibri

² day: ddc number analyzer

⁵ ddc analysis diagram

³ ckb: create knowledge base

⁶ ddc analysis result set

⁴ KB: Knowledge Base

⁷ ddc analysis marc



DDC System 23 - MARC21 XML file (1)

```
... <mx:record><mx:leader>00000nw a220000n 4500 </mx: leader> <mx:controlfield tag= "001">ocd00116680</mx:controlfi  
Eld><... tag= "153"> <mx:subfield code = "a">003.3</mx:subfi  
d>...<mx:subfield code = "j">Computer modeling and simulation  
</mx:subfield>...
```



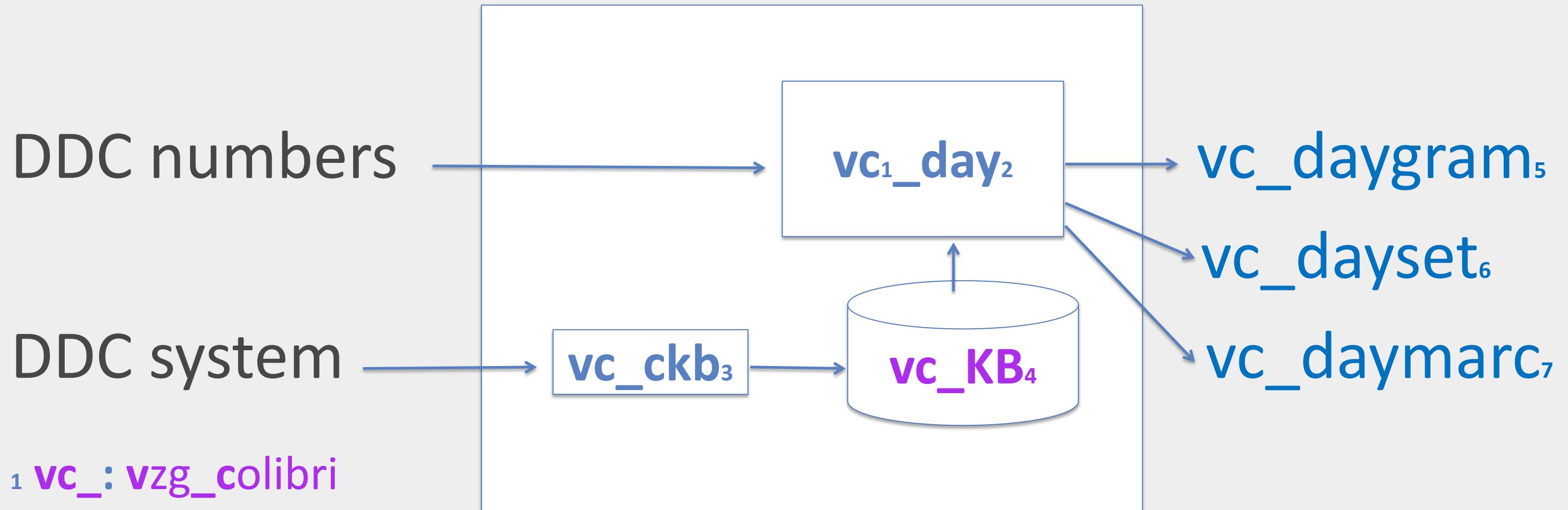
DDC System 23 - MARC21 XML file (2)

```
... tag=“761“><mx:subfield code =“i“>Add to base number</mx:  
subfield><mx:subfield code=“b“>003.3</mx:subfield><mx:subfi  
eld code=“i“>the numbers following</mx:subfield><mx:subfield  
code=“r“>00</mx:subfield><mx:subfield code=“i“>in</mx:subfi  
eld><mx:subfield code = “d“>004</mx:subfield><mx:subfield co  
de=“c“>006</mx:subfield>...
```

15

converted into triplets of the OAV model for the Knowledge Base

DDC search system (vc_ds) - architecture



¹ vc_ : vzg_colibri

² day: ddc number analyzer

⁵ ddc analysis diagram

³ ckb: create knowledge base

⁶ ddc analysis result set

⁴ KB: Knowledge Base

⁷ ddc analysis marc

Knowledge Base – fact parts



003.3, Computer modeling and simulation

<dn0_main>, <dn0_div>, <dn0_sec>, <dn0_sub>, <dn0_bui>
<RI> (MARC21 tag=„750“ ... code=„a“ ... code=„2“ >ddcri<...)

dno: DDC number (no.) ... main, ... division, ... section,
... subdivision, ... built, relative index term (RI)

17

Knowledge Base – rule parts - examples



r1: >... add to base number ...<

r2: >the numbers following<

r3: >in notation<

r7: >in<

r20: >Add to< | Add to <

r21: >Add the numbers following<

r54: >from the period table under<

r59: >. In all cases use one<



Knowledge Base – patterns (1)

examples:

R1: $r1.*r2.*r3.*[r4|r17|r31|r35.*[r38|r51|r36.*r7]]$

-> p1_2_3, p1_2_3_35_38, ...

R2: $r1.*r6.*[r28|r30]$

-> p1_6_28, p1_6_30

R3: $r20.*[r5|[r2.*[r3|r7]]|r23.*r24|r34|r45.*r46]$

-> p20_5, p20_2_3, p20_2_7



Knowledge Base – patterns (2)

cardinalities of rule patterns in the **vc_day** analysis of Liu's sample
(600 DDC numbers)

- | 275| **card_p1_5** add to base number...notation
- | 208| **card_p20_5** add to...notation
- | 173| **card_p1_5_30** add to base number...notation...add further as follows
- | 110| **card_p9** add...instructed under →
- | 74| **card_p1_2_7** add to base number...the numbers following...in
- | 69| **card_p20_2_3** add to...the numbers following...in notation
- | 35| **card_p12** table T1 as modified under



Knowledge Base – patterns (3)

cardinalities of rule patterns in the **vc_day** analysis of Liu's sample
(600 DDC numbers)

- | 24| **card_p1_2_3** add to base number...the numbers following...in notation
- | 23| **card_p20_2_7** add to...the numbers following...in
- | 5| **card_p10_2_7** add to each subdivision identified by...the numbers following...in
- | 3| **card_p1_2_3_17** add to base number...the numbers following...in notation...then add further as instructed under

Knowledge Base – facts & rules (1)



Object - Attribute - Value (OAV model)
dno # descr # descr_val

003.3#<dno_sub>#Computer modeling and simulation

003.3#<dno_seg>#003/.3

003.3#<RI>#Computer modeling

003.3#<RI>#Computer simulation dno_span7_761-

bano

preno2

dno_span8_761

003.3#<r1>#003.3#<r2>#00#<r7>#004-006

Knowledge Base – facts & rules (2)



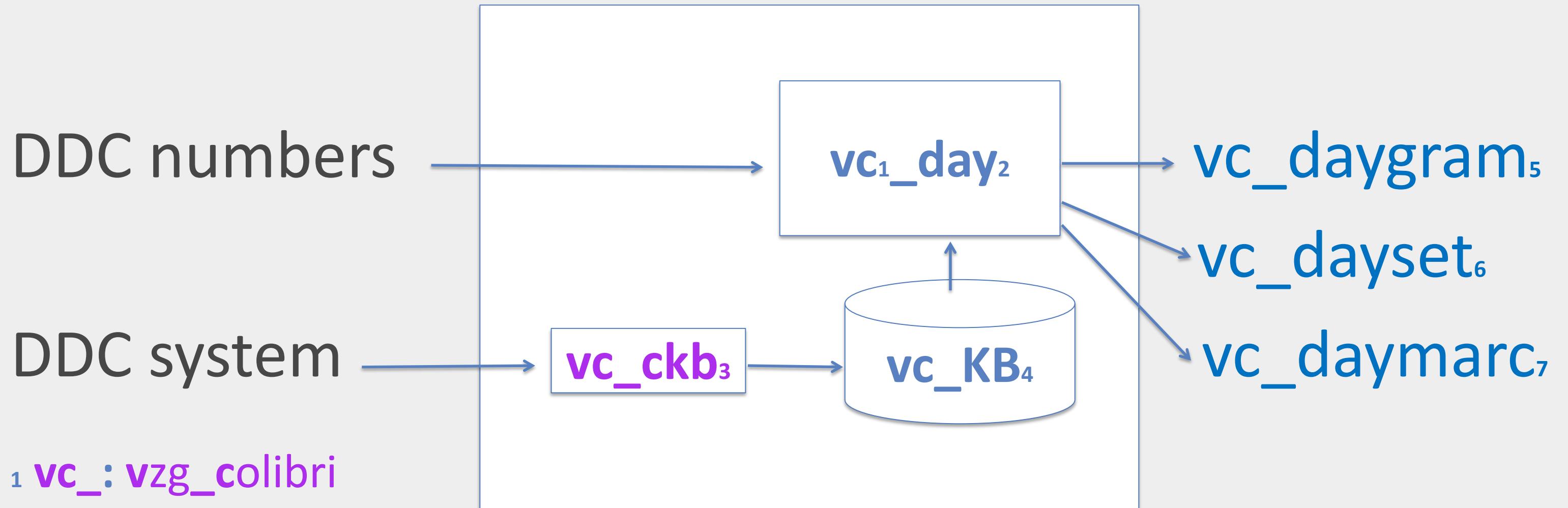
Knowledge Base (vc_KB) – span of DDC numbers

as DDC spans (**k**) / as enumerated DDC numbers (**K**)

facts 197,117 / 333,000

rules 9,210 / 67,530

DDC search system (vc_ds) - architecture



¹ vc_ : vzg_colibri

² day: ddc number analyzer

⁵ ddc analysis diagram

³ ckb: create knowledge base

⁶ ddc analysis result set

⁴ KB: Knowledge Base

⁷ ddc analysis marc

create knowledge base (vc_ckb) algorithm

BEGIN

fu_init_vars()

 IF (a DDC system record is given in several lines) THEN (join them into “e23_line_join”) ELSE (is “e23_line”)

fu_split(e23_line)

fu_pri()

fu_init_vars()

fu_split(e23_line_join)

 ...

FUNCTION fu_init_vars()

FUNCTION fu_split(line)

WHILE (tag is not empty) DO BEGIN

 ...

IF (tag == 153) { determine dno, dno_span1_153, dno_span2_153, dno_add_span1, dno_add_span2, add_tab, cap, descr, tabno_head_153, symbol1_153, symbol2_153, is_dno_bui, ... }

IF (tag == 683) { determine option, addto_r22, dno_span1_683, dno_span2_683, tabno_head_r57, nota_r57, ... }

IF (tag == 761) { determine, bano, addto_r20, preno2, dno_span7_761, dno_span8_761, ... }

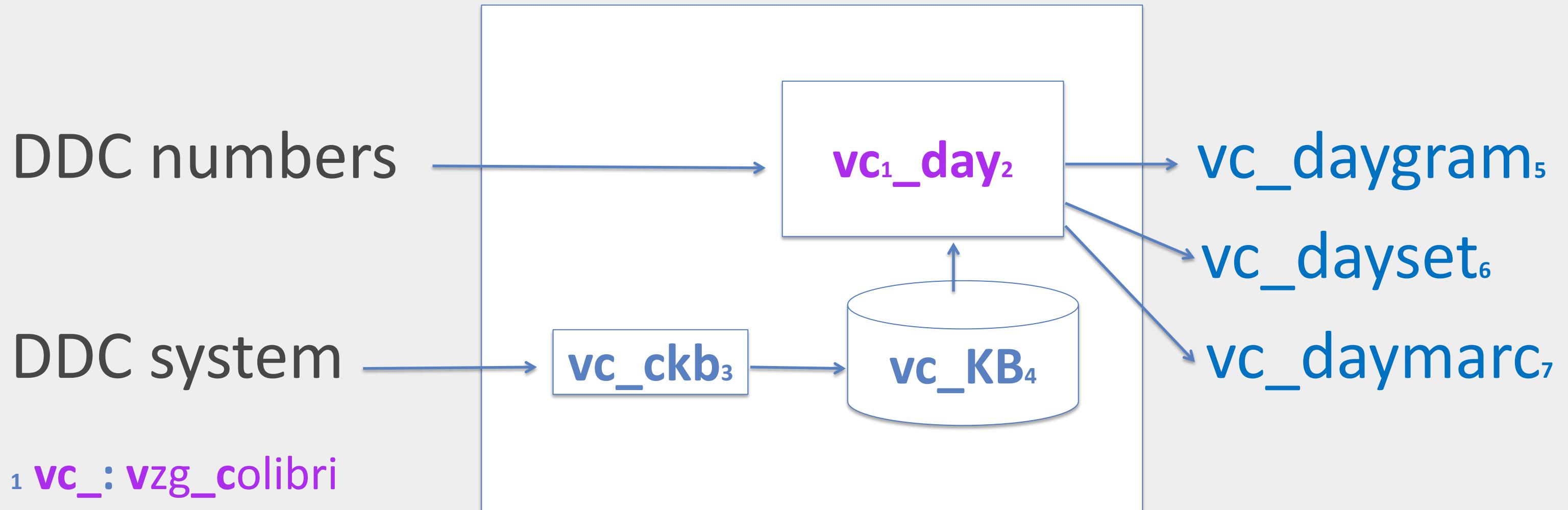
 ...

END

END



DDC search system (vc_ds) - architecture



¹ vc_ : vzg_colibri

² day: ddc number analyzer

⁵ ddc analysis diagram

³ ckb: create knowledge base

⁶ ddc analysis result set

⁴ KB: Knowledge Base

⁷ ddc analysis marc

DDC number analyzer (**vc_day**) algorithm (1)



```
BEGIN
    fu_init()
    fu_read()  (K: 397.532 lines read in 5 sec on colibri2.gbv.de)
    WHILE (true) DO BEGIN
        WHILE (port SRVPORT 5555 is connected) DO BEGIN
            fu_init()
            count infile_lines
            determine dno, count number of (correct, incorrect) dnos
            fu_get_facts ()
            fu_get_facts_via_rules (in_dno)
            fu_pri_765()
            determine whether dno analysis is totally complete or partially complete
            print -- FINI -- on SRVPORT
        END
        print "client has closed the connection, analysis has been finished"
        close SRVPORT
        PRINT statistics (number of [totally/partially analyzed dnos], ...)
    END
END
```

Client-server-application

phase 1

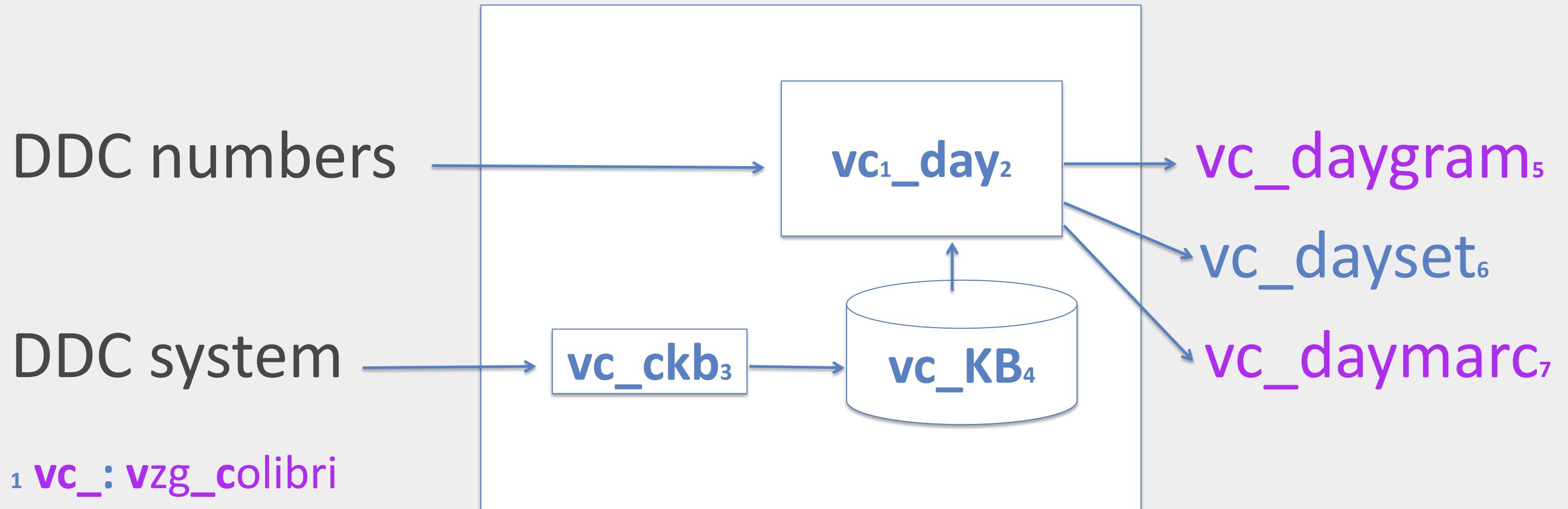
phase 2

DDC number analyzer (**vc_day**) algorithm (2)



```
FUNCTION fu_get_facts_via_rules (in_dno) {
    fu_select_rule()
    fu_chk_rules(dno_head_ok,has_rule)
    ...
    fu_app_p1_5()
    ...
    fu_p1_2_7_17()
    ...
    fu_p1_5()
    ...
    fu_chk_lr_T1();
    ...
}
FUNCTION fu_p1_5() {
    fu_get_tabno_head_and_spans(nota_r5,TS)
    ...
    fu_get_facts2(nota)
}
```

DDC search system (vc_ds) - architecture



¹ vc_ : vzg_colibri

² day: ddc number analyzer

⁵ ddc analysis diagram

³ ckb: create knowledge base

⁶ ddc analysis result set

⁴ KB: Knowledge Base

⁷ ddc analysis marc



Picture see References



708.111 (vc_daygram)

708.111 <765_26_to_analyze; length: 7>
7----- **Arts & recreation** <dno_main> 1 //765_26//
70----- **Arts** <dno_div> 1 //765_26//
708---- **Galleries, museums, private collections of fine and decorative arts** <dno_sec> 1 //765_26//
708.1-- **Geographic treatment** <dno_sub_span> 1 //765_26//
708.1-- **North America** <dno_sub> 2 //765_26//
708.11- **Canada** <dno_sub> 1 //765_26//
---.---1 **British Columbia** <p1_2_3->tabno:T2--711> 6a //765_26//
---.---1 **Canadian Cordillera** <p1_2_3->RI:T2--711> 6a //765_26//

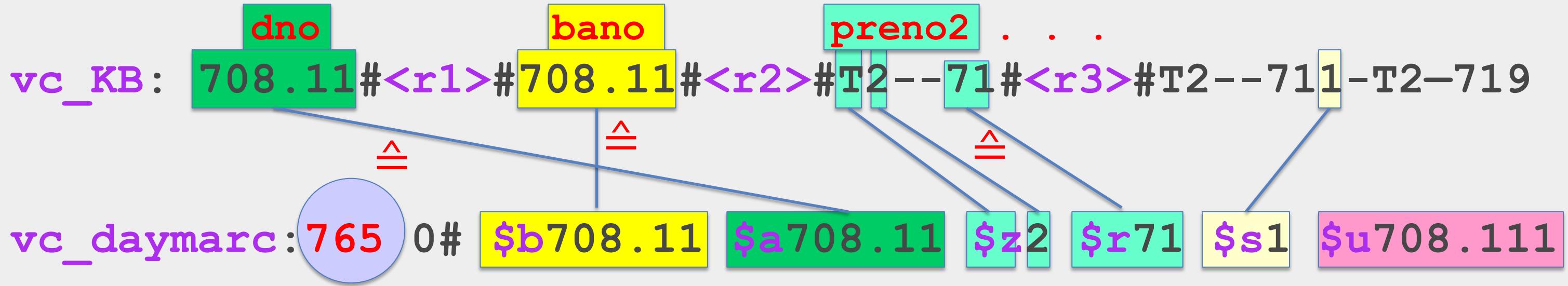
dno_atom: semantically smallest unit that represents a DDC class

dno_mol: via rule(s) synthesized (built) DDC number

30

708.111

(vc_KB, vc_daymarc)



\$a: number for instruction, **\$b**: base number, **\$r**: root number,
\$s: digit/s being added, **\$u**: number being analyzed

\$c: ending number of span, **\$w**: internal table identification
\$y: table sequence number, **\$z**: external table identification

333.333

DDC number!



```
333.333 <ul1_to_analyze; length: 7>
3----- Social sciences <dno_main> 1 //ul1//
33----- Economics <dno_div> 1 //ul1//
333---- Economics of land and energy <dno_sec> 1 //ul1//
333.3-- Ownership of land <dno_sub_span> 1 //ul1//
333.3-- Private ownership of land <dno_sub> 2 //ul1//
333.33- Transfer of possession and of right to use <dno_sub> 1
         //ul1//
333.333 Sale and gift <dno_sub> 1 //ul1//
```

555.55

DDC number!



```
555.55 <ul2_to_analyze; length: 6>
5---- Science <dno_main> 1 //ul2//
55---- Earth sciences & geology <dno_div> 1 //ul2//
555--- Earth sciences by specific continents, countries, localities in modern world;
       extraterrestrial worlds <dno_sub_span> 1 //ul2//
555--- Earth sciences of Asia, . . . <dno_bui> 4 //ul2//
555--- Stratifications (Rock formations) <RI_bui> 5 //ul2//

--5--- Asia <p1_5->tabno:T2--5> 6b //ul2//
--5--- Modern world; extraterrestrial worlds <p1_5->tabno_span:T2--4-T2--9> 6b
      //ul2//
--5--- Modern world <p1_5->RI:T2--5> 6b //ul2//
--5.5- Iran <p1_5->tabno:T2--55> 6b //ul2//
--5.55 Hamadān province, Kordestān province, Kermānshāhān province, Īlām province
      <p1_5->tabno:T2--555> 6b //ul2//
```

666.444

DDC number!



```
666.444 <ul3_to_analyze; length: 7>
6----- Technology <dno_main> 1 //ul3//
66----- Chemical engineering <dno_div> 1 //ul3//
666---- Ceramic and allied technologies <dno_sec> 1 //ul3//
666.4-- Pottery materials, equipment, processes <dno_sub> 1 //ul3//
666.44- Techniques and procedures <dno_sub> 1 //ul3//
666.444 Glazing--technology <RI_bui> 1 //ul3//
666.444 Glazing--pottery, . . . <dno_bui> 4 //ul3//
---.-4- Techniques and procedures <p1_2_7->dno_sub:738.14> 6a //ul3//
---.-44 Glazing <p1_2_7->dno_sub:738.144> 6a //ul3//
---.-44 Glazing--arts <p1_2_7->RI:738.144> 6a //ul3//
```

331.890410263362014098133 (Tore Brattli)



```
331.890410263362014098133 <ul63_to_analyze; length: 25>
3----- Social sciences <dno_main> 1 //ul63//
33----- Economics <dno_div> 1 //ul63//
331----- Labor economics <dno_sec> 1 //ul63//
331.8----- Labor unions, labor-management bargaining and disputes <dno_sub> 1 //ul63//
331.89----- Labor-management bargaining and disputes <dno_sub> 1 //ul63//
331.8904----- Labor-management bargaining and disputes by industry and occupation <dno_sub> 1 //ul63//
331.89041----- Labor-management bargaining and disputes in industries and occupations other than extractive, manufacturing, construction <dno_sub> 1 //ul63//
331.89041026----- Subdivisions for labor-management bargaining and disputes in industries and occupations other than extractive, manufacturing, construction <dno_sub_span> 1 //ul63//
----0----- Computer science, information & general works <p1_5->331.89041001-331.89041999->dno_main:0> 6b //ul63//
----02----- Library & information sciences <p1_5->331.89041001-331.89041999->dno_div:02> 6b //ul63//
----026----- Special libraries <p1_5->331.89041001-331.89041999->RI:026> 6b //ul63//
----026----- Libraries, archives, information centers devoted to specific subjects <p1_5->331.89041001-331.89041999->dno_sec:026> 6b //ul63//
----026----- Specific kinds of institutions <p1_5->331.89041001-331.89041999->dno_sub_span:026> 6b //ul63//
----02633----- Economics libraries <p1_5->331.89041001-331.89041999->RI_bui:026.33> 6b //ul63//
----02633----- Economics libraries, . . . <p1_5->331.89041001-331.89041999->dno_bui:026.33> 6b //ul63//
----026336----- Specific subjects <p1_5->331.89041001-331.89041999->dno_sub_span:026.336> 6b //ul63//
----3----- Social sciences <p1_5->331.89041001-331.89041999->dno_main:3> 6b //ul63//
----33----- Economics <p1_5->331.89041001-331.89041999->dno_div:33> 6b //ul63//
----336----- Public finance <p1_5->331.89041001-331.89041999->RI:336> 6b //ul63//
----336----- Public finance <p1_5->331.89041001-331.89041999->dno_sec:336> 6b //ul63//
----3362----- Internal revenue <p1_5->331.89041001-331.89041999->RI:336.2> 6b //ul63//
----3361----- Taxes <p1_5->331.89041001-331.89041999->dno_sub:336.2> 6b //ul63//
----3362----- Specific forms of revenue <p1_5->331.89041001-331.89041999->dno_sub_span:336.2> 6b //ul63//
----336201----- Taxes by governmental level <p1_5->331.89041001-331.89041999->dno_sub:336.201> 6b //ul63//
----336201----- General topics of taxes <p1_5->331.89041001-331.89041999->dno_sub_span:336.201> 6b //ul63//
----3362014----- Local taxation--public finance <p1_5->331.89041001-331.89041999->RI:336.2014> 6b //ul63//
----3362014----- Local taxes <p1_5->331.89041001-331.89041999->dno_sub:336.2014> 6b //ul63//
----3362014098--- Specific continents, countries, localities <p1_5->331.89041001-331.89041999->dno_sub_span:336.2014098> 6b //ul63//
----0----- Table 1. Standard Subdivisions <tabno:T1--0> 8 //ul63//
----09----- History, geographic treatment, biography <tabno:T1--09> 8 //ul63//
----09----- Historical treatment <RI:T1--09> 8 //ul63//
----098--- Specific continents, countries, localities; extraterrestrial worlds <tabno_span:T1--093-T1--099:098> 8 //ul63//
----8--- South America <p20_5->tabno:T2--8> 6b //ul63//
----8--- Modern world; extraterrestrial worlds <p20_5->tabno_span:T2--4-T2--9> 6b //ul63//
----8--- Modern world <p20_5->RI:T2--8> 6b //ul63//
----81--- Brazil <p20_5->tabno:T2--81> 6b //ul63//
----813- Northeastern region of Brazil <p20_5->tabno:T2--813> 6b //ul63//
----8133 Paraiba state <p20_5->tabno:T2--8133> 6b //ul63//
```

Winners in the Dewey Palindrome Challenge
(The Dewey blog: Rebecca, February 12, 2016)

704.03917104436109041074434323



704.03917104436109041074434323 <ul67a_to_analyze; length: 30>
7----- Arts & recreation <dno_main> 1 //ul67a//
70----- Arts <dno_div> 1 //ul67a//
704----- Special topics in fine and decorative arts <dno_sec> 1 //ul67a//
704.03----- Groups of people <dno_sub_span> 1 //ul67a//
704.03----- Ethnic and national groups <dno_sub> 2 //ul67a//
704.039----- Specific ethnic and national groups <dno_sub_span> 1 //ul67a//
----.9----- Other ethnic and national groups <p1_5->tabno:T5--9> 6b //ul67a//
----.9----- Specific ethnic and national groups <p1_5->tabno_span:T5--1-T5--9> 6b //ul67a//
----.91----- Other Indo-European peoples <p1_5->tabno:T5--91> 6b //ul67a//
----.917----- East Slavs <p1_5->tabno:T5--917> 6b //ul67a//
----.917----- Soviets (People) <p1_5->RI:T5--917> 6b //ul67a//
----.9171----- Russians <p1_5->tabno:T5--9171> 6b //ul67a//
----.9171----- Russians <p1_5->RI:T5--9171> 6b //ul67a//
----.0----- Table 1. Standard Subdivisions <tabno:T1--0> 8 //ul67a//
----.09----- History, geographic treatment, biography <tabno:T1--09> 8 //ul67a//
----.09----- Historical treatment <RI:T1--09> 8 //ul67a//
----.0904----- *20th century, 1900-1999 <tabno:T1--0904> 8 //ul67a//
----.0904----- Historical periods <tabno_span:T1--0901-T1--0905:0904> 8 //ul67a//
----.0904----- Twentieth century <RI:T1--0904> 8 //ul67a//
----.09041----- *1900-1919 <tabno:T1--09041> 8 //ul67a//
----.09041----- World War I, 1914-1918 <RI:T1--09041> 8 //ul67a//
----.07----- Museums, collections, exhibits; collecting objects <p9->tabno_span_1:T1--0901-T1--0905:07> 6b //ul67a//
----.074----- Museums, collections, exhibits <p9->tabno_span_1:T1--0901-T1--0905:074> 6b //ul67a//
----.4----- Europe <p20_5->tabno:T2--4> 6b //ul67a//
----.4----- Modern world; extraterrestrial worlds <p20_5->tabno_span:T2--4-T2--9> 6b //ul67a//
----.4----- Modern world <p20_5->RI:T2--4> 6b //ul67a//
----.43----- Germany and neighboring central European countries <p20_5->tabno:T2--43> 6b //ul67a//
----.434--- Southwestern Germany <p20_5->tabno:T2--434> 6b //ul67a//
----.434--- Germany <p20_5->tabno_span:T2--431-T2--435> 6b //ul67a//
----.4343-- Rhineland-Palatinate (Rheinland-Pfalz) <p20_5->tabno:T2--4343> 6b //ul67a//
----.43432- Northern Rhineland-Palatinate <p20_5->tabno:T2--43432> 6b //ul67a//
----.434323 Koblenz <p20_5->tabno:T2--434323> 6b //ul67a//

[http://www.worldcat.org/
oclc/260050165](http://www.worldcat.org/oclc/260050165)

Treffpunkt Paris! : Russlands
Künstler zwischen
Cezannismus und lyrischer
Abstraktion ; Ludwig Museum
im Deutschherrenhaus
Koblenz, 5. September bis 23.
November 2003

704.039171094436109041074434323



704.039171094436109041074434323 <ul67b_to_analyze; length: 31>
7----- Arts & recreation <dno_main> 1 //ul67b//
70----- Arts <dno_div> 1 //ul67b//
704----- Special topics in fine and decorative arts <dno_sec> 1 //ul67b//
704.03----- Groups of people <dno_sub_span> 1 //ul67b//
704.03----- Ethnic and national groups <dno_sub> 2 //ul67b//
704.039----- Specific ethnic and national groups <dno_sub_span> 1 //ul67b//
----9----- Other ethnic and national groups <p1_5->tabno:T5--9> 6b //ul67b//
----9----- Specific ethnic and national groups <p1_5->tabno_span:T5--1-T5--9> 6b //ul67b//
----91----- Other Indo-European peoples <p1_5->tabno:T5--91> 6b //ul67b//
----917----- East Slavs <p1_5->tabno:T5--917> 6b //ul67b//
----917----- Soviets (People) <p1_5->RI:T5--917> 6b //ul67b//
----9171----- Russians <p1_5->tabno:T5--9171> 6b //ul67b//
----9171----- Russians <p1_5->RI:T5--9171> 6b //ul67b//
----0----- Table 1. Standard Subdivisions <tabno:T1--0> 8 //ul67b//
----09----- History, geographic treatment, biography <tabno:T1--09> 8 //ul67b//
----09----- Historical treatment <RI:T1--09> 8 //ul67b//
----094----- Specific continents, countries, localities; extraterrestrial worlds <tabno_span:T1--093-T1--099:094> 8 //ul67b//
----4----- Europe <p20_5->tabno:T2--4> 6b //ul67b//
----4----- Modern world; extraterrestrial worlds <p20_5->tabno_span:T2--4-T2--9> 6b //ul67b//
----4----- Modern world <p20_5->RI:T2--4> 6b //ul67b//
----44----- France and Monaco <p20_5->tabno:T2--44> 6b //ul67b//
----443----- Champagne-Ardenne, Ile-de-France, Lorraine, Alsace <p20_5->tabno:T2--443> 6b //ul67b//
----4436----- Ile-de-France <p20_5->tabno:T2--4436> 6b //ul67b//
----44361----- Paris department (Paris) <p20_5->tabno:T2--44361> 6b //ul67b//
----09----- History and geographic treatment <p1_5_30->tabno_span_1:T1--093-T1--099:09> 6b //ul67b//
----0904----- Historical periods <p1_5_30->tabno_span_1:T1--093-T1--099:0901-0905> 6b //ul67b//
----0904----- *20th century, 1900-1999 <p20_2_3->tabno:T1--0904> 6b //ul67b//
----0904----- Historical periods <p20_2_3->tabno_span:T1--0901-T1--0905> 6b //ul67b//
----0904----- Twentieth century <p20_2_3->RI:T1--0904> 6b //ul67b//
----09041----- *1900-1919 <p20_2_3->tabno:T1--09041> 6b //ul67b//
----09041----- World War I, 1914-1918 <p20_2_3->RI:T1--09041> 6b //ul67b//
----07----- Museums, collections, exhibits; collecting objects <p9->tabno_span_1:T1--0901-T1--0905:07> 6b //ul67b//
----074----- Museums, collections, exhibits <p9->tabno_span_1:T1--0901-T1--0905:074> 6b //ul67b//
----4----- Europe <p20_5->tabno:T2--4> 6b //ul67b//
----4----- Modern world; extraterrestrial worlds <p20_5->tabno_span:T2--4-T2--9> 6b //ul67b//
----4----- Modern world <p20_5->RI:T2--4> 6b //ul67b//
----43----- Germany and neighboring central European countries <p20_5->tabno:T2--43> 6b //ul67b//
----434----- Southwestern Germany <p20_5->tabno:T2--434> 6b //ul67b//
----434----- Germany <p20_5->tabno_span:T2--431-T2--435> 6b //ul67b//
----4343----- Rhineland-Palatinate (Rheinland-Pfalz) <p20_5->tabno:T2--4343> 6b //ul67b//
----43432----- Northern Rhineland-Palatinate <p20_5->tabno:T2--43432> 6b //ul67b//
----434323----- Koblenz <p20_5->tabno:T2--434323> 6b //ul67b//

<http://www.worldcat.org/oclc/260050165>
Treffpunkt Paris! :
Russlands Künstler
zwischen Cezannismus
und lyrischer Abstraktion
; Ludwig Museum im
Deutschherrenhaus
Koblenz, 5. September
bis 23. November 2003

Causes of analysis failures



- DDC number incorrectly constructed (wrong / missing digit/s, extra digit/s) -> **input failure**
- Wrong/incomplete **DDC system**
- Wrong/incomplete **DDC Knowledge Base (vc_KB)**
- Wrong/incomplete **implementation (vc_day)**

38

Incomplete DDC system, e. g.



```
< ... >  
<mx:controlfield tag =“001“>ocd00154566</mx:controlfield>  
< ... >  
<mx:subfield code=“d“>616.1</mx:subfield>  
<mx:subfield code=“c“>616.9</mx:subfield>  
<mx:subfield code=“y“>1</mx:subfield>  
< ... >
```

Incomplete Knowledge Base (vc_KB)

if “ul-e23-added” is not contained in K



----- ul-added [beg] -----

T7--06#<tabno>#[Persons by miscellaneous social characteristics]#!ul-added from e21!

T7--06#<r1>#T7--06<9f><r2>#T1-086#<r7>#T1--0862-T1-0869#!ul-added from e21!

T7--792#<tabno>#[Stage presentations]#!ul-added from e21!

T7--7921#<tabno>#[Drama]#!ul-added from e21!

T2--43658#<tabno>#Suedwestliche Steiermark#!ul-added from German e23!

T2--436582#<tabno>#Bezirk Leibnitz#!ul-added from German e23!

005.369#<dno_sub>#[Specific programs]#!ul-added from e21!

721-729:042#<dno_sub>#[Architecture for persons with disabilities]#!ul-added from e21!

781.284#<r9>#781.2-781.8#!ul-added from e21!

790.0135#<dno_sub>#[Effective use of leisure]#!ul-added from e21!

...

----- ul-added [end] -----

Incomplete implementation (vc_day)



```
printf("# FUp1_5_11 ##### Not implemented yet #####\n");
printf("# FUp1_5_17 ##### Not implemented yet #####\n");
printf("# FUp9_19 ##### Not implemented yet #####\n");
printf("# FUp20_2 ##### Not implemented yet #####\n");
printf("# FUp20_23_24 ##### Not implemented yet #####\n");
printf("# FUp21_7 ##### Not implemented yet #####\n");
printf("# FUp22_5 ##### Not implemented yet #####\n");
printf("# FUp25 ##### Not implemented yet #####\n");
printf("# FUp26 ##### Not implemented yet #####\n");
printf("# FUp27_7 ##### Not implemented yet #####\n");
printf("# FUp29 ##### Not implemented yet #####\n");
printf("# FUp55_19 ##### Not implemented yet #####\n");
printf("# FUp8_18_17 ##### Not implemented yet #####\n");
```

DDC search system (vc_ds) – size & complexity



number of	2009	2016
programs & versions (vc_ds)	80	173
functions (vc_day)	78	91
lines (vc_day)	7,728	9,913
conditional statements (vc_day)	878	1,445

Result of the automatic analysis of 824,735 unique DDC numbers



DDC class analyzed (all)	last dno_id	completely analyzed	
DDC class 0	35,980	35980	25,782 72%
DDC class 1	9,896	45876	8,156 82%
DDC class 2	37,689	83565	33,093 88%
DDC class 3	334,556	418121	277,308 83%
DDC class 4	13,362	431483	11,330 85%
DDC class 5	62,134	493617	52,064 83%
DDC class 6	99,189	592806	88,561 89%
DDC class 7	99,083	691889	85,708 87%
DDC class 8	34,666	726555	21,477 62%
DDC class 9	98,180	824735	68,316 70%



Reaching up to further Colibri tasks

Live demonstration ?

Thank you for your interest in the VZG project

COLIBRI / DDC

My thanks to

Dipl.-Kfm. Reiner Diedrichs, Executive Director of the VZG, Göttingen

44



Appendix

45



Revision of the Analysis of the DDC Number

704.0391710443610904107443433

(instead of the slides 36 & 37)

<http://www.worldcat.org/oclc/260050165>
Treffpunkt Paris! :
Russlands Künstler
zwischen Cezannismus
und lyrischer
Abstraktion ; Ludwig
Museum im
Deutschherrenhaus
Koblenz, 5. September
bis 23. November
2003

704.03917104436109041074434323 (1)



704.03917104436109041074434323 <ul67a_to_analyze; length: 30>
7----- Arts & recreation <dno_main> 1 //ul67a//
70----- Arts <dno_div> 1 //ul67a//
704----- Special topics in fine and decorative arts <dno_sec> 1 //ul67a//
704.03----- Groups of people <dno_sub_span:704.03-704.08> 1 //ul67a//
704.03----- Ethnic and national groups <dno_sub> 2 //ul67a//
704.039----- Specific ethnic and national groups <dno_sub_span:704.031-704.039> 1 //ul67a//
---.9----- Other ethnic and national groups <p1_5->tabno:T5--9> 6b //ul67a//
---.9----- Specific ethnic and national groups <p1_5->tabno_span:T5--1-T5--9> 6b //ul67a//
---.91----- Other Indo-European peoples <p1_5->tabno:T5--91> 6b //ul67a//
---.917----- East Slavs <p1_5->tabno:T5--917> 6b //ul67a//
---.917----- Soviets (People) <p1_5->RI:T5--917> 6b //ul67a//
---.9171----- Russians <p1_5->tabno:T5--9171> 6b //ul67a//
---.9171----- Russians <p1_5->RI:T5--9171> 6b //ul67a//
---.0----- <Facet Indicator> <p49_52_53->tabno:T5--9171:0> 14 //ul67a//
---.4----- Europe <p49_52_53->tabno:T2--4> 6b //ul67a//
---.4----- Modern world; extraterrestrial worlds <p49_52_53->tabno_span:T2--4-T2--9> 6b //ul67a//
---.4----- Modern world <p49_52_53->RI:T2--4> 6b //ul67a//
---.44----- France and Monaco <p49_52_53->tabno:T2--44> 6b //ul67a//
---.443----- Champagne-Ardenne, Ile-de-France, Lorraine, Alsace <p49_52_53->tabno:T2--443> 6b //ul67a//
---.4436----- Ile-de-France <p49_52_53->tabno:T2--4436> 6b //ul67a//
---.44361----- Paris department (Paris) <p49_52_53->tabno:T2--44361> 6b //ul67a//
---.0----- Table 1. Standard Subdivisions <tabno:T1--0> 8 //ul67a//
---.09----- History, geographic treatment, biography <tabno:T1--09> 8 //ul67a//
---.09----- Historical treatment <RI:T1--09> 8 //ul67a//
---.0904----- *20th century, 1900-1999 <tabno:T1--0904> 8 //ul67a//
---.0904----- Historical periods <tabno_span:T1--0901-T1--0905:0904> 8 //ul67a//
---.0904----- Twentieth century <RI:T1--0904> 8 //ul67a//
---.09041----- *1900-1919 <tabno:T1--09041> 8 //ul67a//
---.09041----- World War I, 1914-1918 <RI:T1--09041> 8 //ul67a//
---.07----- Museums, collections, exhibits; collecting objects <p9->tabno_span_1:T1--0901-T1--0905:07> 6b //ul67a//
---.074----- Museums, collections, exhibits <p9->tabno_span_1:T1--0901-T1--0905:074> 6b //ul67a//
... .

<http://www.worldcat.org/oclc/260050165>
Treffpunkt Paris! : Russlands Künstler zwischen Cezannismus und lyrischer Abstraktion ; Ludwig Museum im Deutschherrenhaus Koblenz, 5. September bis 23. November 2003

704.03917104436109041074434323 (2)



. . .
-----4---- Europe <p20_5->tabno:T2--4> 6b //u167a//
-----4---- Modern world; extraterrestrial worlds <p20_5->tabno_span:T2--4-T2--9> 6b //u167a//
-----4---- Modern world <p20_5->RI:T2--4> 6b //u167a//
---43--- Germany and neighboring central European countries <p20_5->tabno:T2--43> 6b //u167a//
---434--- Southwestern Germany <p20_5->tabno:T2--434> 6b //u167a//
---434--- Germany <p20_5->tabno_span:T2--431-T2--435> 6b //u167a//
---4343-- Rhineland-Palatinate (Rheinland-Pfalz) <p20_5->tabno:T2--4343> 6b //u167a//
---43432- Northern Rhineland-Palatinate <p20_5->tabno:T2--43432> 6b //u167a//
---434323 Koblenz <p20_5->tabno:T2--434323> 6b //u167a//

u167a[5]: 765 0# \$b704.03917104436109041074 \$z1 \$a0901 \$c0905 \$z1 \$w0901 \$c0905 \$y1 \$a074 \$z2 \$s434323 \$u704.03917104436109041074434323
u167a[4]: 765 0# \$b704.03917104436109041 \$z1 \$a09041 \$z1 \$a0901 \$c0905 \$z1 \$w0901 \$c0905 \$y1 \$t074 \$u704.03917104436109041074
u167a[3]: 765 0# \$b704.039171044361 \$z1 \$s09041 \$u704.03917104436109041
u167a[2]: 765 0# \$b704.039171 \$z5 \$a9171 \$f0 \$z2 \$s44361 \$u704.039171044361
u167a[1]: 765 0# \$b704.03 \$a704.031 \$c704.039 \$z5 \$s9171 \$u704.039171

Knowledge Base – rule parts –examples



r1: >... add to base number ...<

r2: >the numbers following<

r3: >in notation<

r5: >[Nn]otation<

r9: >Add as instructed under<

r20: >Add to< | >Add to |<

r30: >... add [further] as follows ...<

r26: >Add to base number as instructed at beginning of<

Most frequent rule patterns of the analysis of the 824.735 DDC numbers



classes	dno0	dno1	dno2	dno3	dno4	dno5	dno6	dno7	dno8	dno9
card_p1_2_3					16.41%				17.92%	
card_p1_5	60.62%	35.95%	31.42%	35.41%	61.24%		23.33%	27.25%		44.81%
card_p1_5_30										36.19%
card_p9						26.85%				
card_p20_5	18.15%	34.12%	22.61%	31.31%		28.10%	29.60%	27.00%		
card_p1_26									29.82%	
	78.77%	74.07%	54.03%	66.72%	77.65%	54.95%	52.93%	54.25%	47.74%	81.00%

503 palindromes in GVK, longest GVK palindrome is „338.47674833“ (11-digits): The 2005-2010 world outlook for veneer, plywood, and engineered wood product manufacturing (by Philip M. Parker) ([ppn_801230349](#), [oclc_17163419](#), [in_ul76](#))

Cardinalities of patterns in the DDC classes (1)



	dno0	dno1	dno2	dno3	dno4	dno5	dno6	dno7	dno8	dno9
analyzed dnos	35980	9896	37689	334556	13362	62134	99189	99083	34666	98180
cardinalities										
p1_2_3	1456	53	778	10033	1798	1622	1773	5296	10195	1188
p1_2_3_17	412	47	205	5862	42	1260	1117	587	414	507
p1_2_3_31	153	3	23	479	5	119	163	54	2460	382
p1_2_3_4	15	-	-	1	-	-	-	-	778	-
p1_2_7	1634	352	5636	23270	100	4780	15110	10946	939	417
p1_2_7_17	-	-	-	-	-	-	-	9	-	-
p1_2_7_30	-	-	130	-	-	-	-	14	-	-
p1_5	25818	2092	13351	156259	6712	11008	20325	42945	6477	57884
p1_5_30	2607	608	4548	36052	392	2839	5735	20886	4251	46745
p1_5_47	-	-	-	230	-	-	-	-	-	38
p1_5_35_36_7	800	-	-	2172	-	-	-	1	-	253
p1_5_35_38	-	-	254	2920	-	-	-	-	1	-
p1_5_47	-	1	1	-	-	-	3	41	-	-
p1_6	87	-	-	-	-	-	-	1	8166	-
p1_26	732	-	-	3	-	-	-	5	16961	-
p9	623	89	4750	46539	18	18817	13560	19351	397	13981
p10_2_7	-	-	-	-	-	-	108	-	-	-
p12	48	3	1	317	1040	2741	356	4435	-	1
p13	-	-	-	-	-	-	-	831	-	-
p20_2_3	471	281	1472	14942	59	512	2051	6750	120	371
p20_2_7	5	131	1738	4028	-	6695	1030	2890	-	13
p20_5	7731	1986	9608	138169	794	19698	25785	42570	5726	7402
p20_34	-	174	-	-	-	-	-	-	-	-
p26	-	-	-	2	-	-	-	-	-	-
p32	-	-	-	2	-	-	-	1	1	-

Cardinalities of patterns in the DDC classes (2)



occurrences in dnos	42592	5820	42495	441280	10960	70091	87116	157613	56886	129182
cardinalities	dno0	dno1	dno2	dno3	dno4	dno5	dno6	dno7	dno8	dno9
p1_2_3	3.4%	0.91%	1.83%	2.27%	16.41%	2.31%	2.03%	3.36%	17.92%	0.92%
p1_2_3_17	0.96%	0.81%	0.48%	1.33%	0.38%	1.80%	1.28%	0.37%	0.73%	0.39%
p1_2_3_31	0.36%	0.07%	0.05%	0.11%	0.05%	0.17%	0.19%	0.03%	4.32%	0.30%
p1_2_3_4	0.04%	-	-	0.00%	-	-	-	-	1.36%	-
p1_2_7	3.84%	6.05%	13.26%	5.27%	0.91%	6.82%	17.35%	6.94%	1.65%	0.32%
p1_2_7_17	-	-	-	-	-	-	-	0.00%	-	-
p1_2_7_30	-	-	0.31%	-	-	-	-	0.00%	-	-
p1_5	60.62%	35.95%	31.42%	35.41%	61.24%	15.71%	23.33%	27.25%	11.39%	44.81%
p1_5_30	6.12%	10.45%	10.70%	8.17%	3.58%	4.05%	6.58%	13.25%	7.47%	36.19%
p1_5_47	-	-	-	0.05%	-	-	-	-	-	0.03%
p1_5_35_36_7	1.88%	-	-	0.49%	-	-	-	0.00%	-	0.19%
p1_5_35_38	-	-	0.60%	0.66%	-	-	-	-	0.00%	-
p1_5_47	-	0.02%	0.01%	-	-	-	0.00%	0.03%	-	-
p1_6	0.2%	-	-	-	-	-	-	0.00%	14.36%	-
p1_26	1.72%	-	-	0.00%	-	-	-	0.00%	29.82%	-
p9	1.46%	1.53%	11.18%	10.55%	0.16%	26.85%	15.57%	12.28%	0.70%	10.82%
p10_2_7	-	-	-	-	-	-	0.12%	-	-	-
p12	0.11%	0.05%	0.01%	0.07%	9.49%	3.91%	0.41%	2.81%	-	0.00%
p13	-	-	-	-	-	-	-	0.53%	-	-
p20_2_3	1.11%	4.83%	3.46%	3.39%	0.54%	0.73%	2.35%	4.28%	0.21%	0.29%
p20_2_7	0.01%	2.25%	4.09%	0.91%	-	9.55%	1.18%	1.83%	-	0.01%
p20_5	18.15%	34.12%	22.61%	31.31%	7.24%	28.10%	29.60%	27.00%	10.06%	5.73%
p20_34	-	2.99%	-	-	-	-	-	-	-	-
p26	-	-	-	0.00%	-	-	-	-	-	-
p32	-	-	-	0.00%	-	-	-	0.00%	0.00%	-

DDC search system - (vc_) programs (1)



```
04-06-30 vc_make_pdl.awk
04-10-12 vc_cdb.awk
04-10-15 vc_day_srv.awk
04-10-15 vc_lib_KB.awk
04-10-17 vc_ckb.awk
05-02-19 vc_lib_DB.awk
05-02-19 vc_lib_DB_new.awk
05-02-26 vc_cin1.awk
05-06-01 vc_kbs
05-06-01 vc_kbs.awk
05-06-14 vc_cin2.awk
05-07-13 vc_pdl2in.awk
06-01-19 vc_make_in
06-01-19 vc_make_in.awk
06-01-25 vc_make_eval_sample
06-01-25 vc_make_eval_sample.awk
06-03-10 vc_make_ddc_summaries
06-03-10 vc_make_ddc_summaries.awk
06-04-12 vc_make_sample.awk
06-09-07 vc_dcl_srv
06-09-07 vc_idb
06-09-07 vc_idb.awk
06-09-28 vc_pdb.awk
06-09-28 vc_pin.awk
06-10-07 vc_idb2
06-10-07 vc_idb2.awk
06-10-07 vc_idb3.awk
06-10-08 vc_dcl_srv.awk
06-11-15 vc_idb3

vzg_colibri-make list with ppn no. & DDC values
vzg_colibri-create ddc_data_base (DDC-Basis) (standalone) for Pica data ddc0.dat ... ddc9.dat
vzg_colibri-ddc notation analyzer; V1
vzg_colibri-library
vzg_colibri-create knowledge base
vzg_colibri-library for ddc_base (DDC-Basis) (standalone)
vzg_colibri-library for ddc_base (DDC-Basis) (standalone)
vzg_colibri-create infiles for vc_days
vzg_colibri-knowledge base (range) splitter
vzg_colibri-knowledge base (range) splitter
vzg_colibri-create infiles for vc_days
vzg_colibri-convert file "pdl" (ppn_dno_list) into file "in"
vzg_colibri-make (test) in[file]
vzg_colibri-make (test) in[file]
vzg_colibri-make sample for evaluation
vzg_colibri-make sample for evaluation
vzg_colibri-make ddc summaries
vzg_colibri-make ddc summaries
vzg_colibri-make sample for GVK title records
vzg_colibri_ddc (automatic) classification
vzg_colibri-create inverted ddc_base
vzg_colibri-create inverted ddc_base
vzg_colibri-prepare ddc_base (standalone)
vzg_colibri-prepare input file to ddc (gvk data -> for vc programs)
vzg_colibri-create inverted ddc_base
vzg_colibri-create inverted DDC base
vzg_colibri-create inverted DDC base
vzg_colibri_ddc (automatic) classification
vzg_colibri-create inverted ddc_base
```

DDC search system - (vc_) programs (2)



```
06-11-15 vc_sdb
06-11-16 vc_low.awk
06-11-27 vc_ran.awk
06-11-30 vc_polc.awk
06-11-30 vc_polc2.awk
06-12-07 vc_psru.awk
06-12-09 vc_psam.awk
06-12-13 vc_idba.awk
06-12-25 vc_dat1.awk
06-12-27 vc_dat3.awk
06-12-31 vc_dat4.awk
07-01-05 vc_dat2.awk
07-01-06 vc_add.awk
07-01-06 vc_cli.awk
07-01-06 vc_day_srv
07-01-06 vc_dno_chk
07-01-06 vc_pgvk
07-01-06 vc_srv.awk
07-01-13 vc_dcl_cli.awk
07-01-17 vc_con.awk
07-02-01 vc_pbk.awk
07-06-27 vc_pgvk.awk
07-07-06 vc_cdbl.awka
07-11-27 vc_dcl_srv_logic_if
07-12-19 vc_cdb2.awk
08-02-10 vc_chk2_ksh_nohup
08-03-04 vc_idbu.awk
08-04-21 vc_psam2.awk
08-04-30 vc_psam3.awk

vzg_colibri_search DDC database (search for title records already in vc_DB)
vzg_colibri_tolower descr_val (convert letter to lower case)
vzg_colibri_random number generator
vzg_colibri-prepare olc data (standalone)
vzg_colibri-prepare olc data tolower and bring data in input format for vc_dcl (standalone)
vzg_colibri-prepare sru requests
vzg_colibri_prepare sample
vzg_colibri-create inverted ddc_base idno-added (f.e. WWW)
vzg_colibri-data analysis: card. of dno_all, dno_correct, dno_incorrect
vzg_colibri-data analysis: count number of dnos
vzg_colibri-data analysis: count number of synthesized dnos (dno_mol_cnt)
vzg_colibri-data analysis: card. of x-arity-dnos
vzg_colibri-add numbers (of loc, days, ...)
vzg_colibri-client
vzg_colibri-ddc notation analyzer; V1
vzg_colibri-ddc number checker (if correct)
vzg_colibri_prepare gvk (test) file for ddc-based search
vzg_colibri-server
vzg_colibri-ddc classification client
vzg_colibri-convert file1 into file2
vzg_colibri-prepare bk data for ****
vzg_colibri_prepare gvk (test) file for ddc-based search
vzg_colibri-create ddc_data_base_for_lingo (DDC-Basis) for Pica data ddc0.dat ... ddc9.dat
# 1. FUREAD ... # IF in_file_to_ddc_lines~"line" ...
vzg_colibri-create ddc_data_base2 (DDC-Basis2) (standalone) 2: for DNB data A0745ti2.dis, ..., H0711ti2.dis
call vc_bidno_ksh mit nohup
vzg_colibri-idb unique
vzg_colibri_prepare sample for in_dnb_ABH (data in MAB2 format)
vzg_colibri_prepare sample vc_DB/dnb_A_0, ... , vc_DB/dnb_H_9
```

DDC search system - (vc_) programs (3)



```
08-05-01 vc_psam4.awk
08-06-19 vc_tra.awk
08-06-25 vc_corr.awk
08-07-08 vc_dsy_704.9489439230951507473
08-07-13 vc_determine_dnos
08-10-06 vc_res_sample1_dno_only
08-10-06 vc_res_sample1_dno_only.awk
08-10-15 vc_res_dno_dno_like_cnt.awk
08-10-29 vc_dcl_srv_logic_fu
08-10-30 vc_sdb.awk
09-01-29 vc_res_mark
09-01-29 vc_res_mark.awk
09-02-03 vc_res_sample1_final_extract
09-02-15 vc_dno_length_distribution.awk
09-02-15 vc_res_sample1_final_extract.awk
09-03-13 vc_pdl.awk
09-03-19 vc_pdl2.awk
09-03-25 vc_pdma.awk
09-03-26 vc_pdma
09-04-19 vc_make_subset_by_random.awk
09-04-24 vc_ran_norm_from_21422.awk
09-04-29 vc_make_subset_by_XXX.awk
09-05-09 vc_dce.awk
10-02-23 vc_make_test_coll
10-02-24 vc_make_test_coll.awk
10-03-02 vc_dat5.awk
10-03-08 vc_dat5_ksh
10-04-06 vc_ckb_ger
10-04-06 vc_ckb_ger.awk

vzg_colibri_prepare sample dnb_A_0-20, dnb_A_1-20, ..., dnb_H_9-40
vzg_colibri_translate characters (for vc_dqa)
vzg_colibri-correlation determination
mysql> select * from dno_db where (dno like "%73%" and dno like "%515%" ...
vzg_colibri-determine dnos and cardinality
vzg_colibri-produce only the dno sample1 results
vzg_colibri-produce only the dno sample1 results
vzg_colibri_count the number of dnos and dnos_like
# 1. FUREADIDB, ...
vzg_colibri_search for title records that are already contained in vc_DB!
vzg_colibri-mark the res correlation line with dno, dno_like
vzg_colibri-mark the res correlation line with dno, dno_like
vzg_colibri-extract the sample1 results from the clean final result file
vzg_colibri-determine the distribution of the dno lengths
vzg_colibri-extract the sample1 results from the clean final result file
vzg_colibri-prepare file ppn-dno list (using idno information)
vzg_colibri-prepare file ppn-dno list (using title records)
vzg_colibri-ppn dno molecular atomar list for GVK
vzg_colibri-ppn dno molecular atomar list for GVK
vzg_colibri-make subset from result set by random file(s)
vzg_colibri_random normalizing between [0,1] from 21422
vzg_colibri-make subset from result set by XXX
vzg_colibri-ddc classification evaluation
vzg_colibri-make test collection (title records not in vc_DB)
vzg_colibri-make test collection (title records not in vc_DB)
vzg_colibri-data analysis: count number of fields <001A>, <007G>
vc_dat5.awk-vc_DB-09-12-11/ddc0 vc_DB-09-12-11/ddc1 ...
vzg_colibri-create knowledge base_german
vzg_colibri-create knowledge base_german
```

DDC search system - (vc_) programs (4)



```
10-04-10 vc_kbs2
10-04-10 vc_kbs2.awk
10-05-06 vc_ckb_legend
10-05-25 vc_ckb2_ger
10-05-25 vc_ckb2_ger.awk
10-05-25 vc_ckb_ger_hatzen
10-05-25 vc_ckb_ger_hatzen.awk
10-05-27 vc_ckb3
10-05-27 vc_ckb3.awk
10-06-02 vc_tra2.awk
10-07-06 vc_cdbl
10-07-14 vc_cdbl2.awk
10-10-05 vc_chk
10-10-05 vc_chk2
10-11-16 vc_chk3.awk
11-05-25 vc_idbl
11-05-25 vc_idbl.awk
11-05-25 vc_idbn
11-05-25 vc_idbn.awk
11-09-20 vc_tst_chk2
11-10-19 vc_chk2t.awk
11-11-03 vc_chk2_ksh
11-11-03 vc_chk2_sum
11-11-03 vc_chk2_sum.awk
11-11-16 vc_ainb
11-11-16 vc_ainb.awk
11-11-21 vc_chkd_ksh
11-11-23 vc_cdb2_ksh
11-11-28 vc_del.awk

vzg_colibri-knowledge base (range) splitter2 splitting program for vc_ckb5.awk
vzg_colibri-knowledge base (range) splitter2 splitting program for vc_ckb5.awk
hat: hierarchy at class ... mcf: MARC21 classification format field
vzg_colibri-create knowledge base_german with <hatzen>
vzg_colibri-create knowledge base_german with <hatzen>
vzg_colibri-create knowledge base_german hatzen
vzg_colibri-create knowledge base_german hatzen
vzg_colibri-create knowledge base with infile "importlog.xml"
vzg_colibri-create knowledge base with infile "importlog.xml"
vzg colibri translate characters (e.g., for vc_days)
vzg colibri-create ddc_data_base_for_lingo (DDC-Basis) for Pica data ddc0.dat ... ddc9.dat
vzg colibri-create ddc_data_base_for_lingo 2: for DNB data A0745ti2.dis, ...
vzg_colibri_check whether title record (to classify) is contained in the DDC data / knowledge base
vzg_colibri-check whether title record (to classify) is contained in the DB/KB database
vzg_colibri-check which subfields are contained in tit_021A supplement to vc_chk2.awk
vzg_colibri-create inverted DDC base from lingo for vc_IDB
vzg_colibri-create inverted DDC base from lingo for vc_IDB
vzg_colibri-create inverted DDC base normalized
vzg_colibri-create inverted DDC base normalized
.../vc_cdb -f .../vc_lib_DB.awk .../vc_DB/pnn_0547516916 > dno5 ...
vzg_colibri-check test whether title record (to classify) is contained in the DB / KB
vzg_colibri-check which title records are contained in vc_DB
vzg_colibri-check summation of all single results of vc_chk2_ksh
vzg_colibri-check summation of all results of vc_chk2_ksh
vzg_colibri-is a in b contained?
vzg_colibri-is a in b contained?
egrep "in file_chkd" in_dnb_A0745 | wc -l ... egrep "in file_chkd" in_dnb_H1010 | wc -l
vzg_colibri-create ddc_data_base_for_lingo
vzg_colibri-delete lines with !!!!!!
```

DDC search system - (vc_) programs (5)



```
11-11-28 vc_del2.awk
11-11-28 vc_del_ksh
12-01-09 vc_pdce
12-01-10 vc_pdce.awk
12-02-16 vc_ckb4
12-02-16 vc_ckb4.awk
12-02-21 vc_conc_dre.awk
12-02-22 vc_conc_dre
12-03-12 vc_pconc_dr.awk
12-03-12 vc_tra3.awk
12-03-19 vc_cnt2
12-03-19 vc_cnt2.awk
12-07-18 vc_ckb5.awk
12-08-30 vc_dat6.awk
13-01-28 vc_day_cli.awk
13-01-28 vc_day_srv.awk
13-03-06 vc_tra4.awk
13-03-21 vc_pact.awk
13-03-25 vc_dce2.awk
13-03-25 vc_pact2.awk
13-03-25 vc_pact3
13-03-25 vc_pact3.awk
13-03-27 vc_pdce2
13-03-27 vc_pdce2.awk
13-03-29 vc_dce3.awk
13-03-29 vc_dce4.awk
13-03-29 vc_pact4
13-03-29 vc_pact4.awk
13-03-29 vc_pdce3.awk

vzg_colibri-delete "<026>" -lines without "^705_line"
vc_del vc_DB/in_dnb_A0745 > vc_DB/in_dnb_A0745_without_contained, ...
vzg_colibri-prepare ddc classification evaluation for lingo data
vzg_colibri-prepare ddc classification evaluation for lingo data
vzg_colibri-create knowledge base with infile "rvk_nota_cap_flo"
vzg_colibri-create knowledge base with infile "rvk_nota_cap_flo"
vzg_colibri-concordances ddc rvk evaluation
vzg_colibri-concordances ddc rvk evaluation
vzg_colibri-prepare conc ddc rvk data (for edug12)
vzg_colibri translate characters (for vc_conc_dr)
vzg_colibri-count dno rvk pairs (for edug12)
vzg_colibri-count dno rvk pairs (for edug12)
vzg_colibri-create knowledge base5 (with e23)
vzg_colibri-data analysis: count number of title records with holding information with corresponding dnos, etc.
vzg_colibri-ddc analyzer client; V3
vzg_colibri-ddc number analyzer; V3
vzg_colibri_translate characters (for vc_cdb char)
vzg_colibri-prepare data for act := the automatic classification toolbox
vzg_colibri-ddc classification evaluation 2
vzg_colibri-prepare data for act := the automatic classification toolbox
vzg_colibri-prepare data for act := the automatic classification toolbox
vzg_colibri-prepare vc_dcl_res_data for dce2
vzg_colibri-prepare vc_dcl_res_data for dce2
vzg_colibri-ddc classification evaluation 3
vzg_colibri-ddc classification evaluation 4
vzg_colibri-prepare data for act := the automatic classification toolbox
vzg_colibri-prepare data for act := the automatic classification toolbox
vzg_colibri-prepare vc_dcl_res_data for vc_dce2 take only 1 dno_a
```

DDC search system - (vc_) programs (6)



```
13-04-23 vc_rvk2marc.awk
13-08-06 vc_ddc_rvk_stat.awk
14-04-09 vc_dno_chk.awk
14-05-14 vc_cnt_dno_classes_ksh
15-02-20 vc_day_srv_FUp
15-05-08 vc_day_rules
16-01-12 vc_pal_chk.awk
16-01-12 vc_pal_make.awk
16-01-22 vc_dno_rev
16-01-22 vc_dno_rev.awk
16-01-29 vc_dsy.awk
16-03-16 vc_cnt3
16-03-16 vc_cnt3.awk
16-03-21 vc_make_pdl2.awk
16-03-21 vc_make_pdl3
16-03-21 vc_make_pdl3.awk
16-03-21 vc_make_pdl4.awk
16-03-22 vc_make_gvk_all
16-03-22 vc_make_gvk_all.awk
16-03-29 vc_cli_ksh
16-03-29 vc_day_ksh
16-04-06 vc_day_srv_logic_fu

vzg_colibri-rvk to MARC format
# vzg_colibri-DDC-RVK-concordance statistics
vzg_colibri-ddc number checker (if correct) for Pica data ddc0.dat ... ddc9.dat
echo "----- [0a] dno_wrong: char/s et al. -- ...
function fu_p10_2_7() ... function fu_p9_19()
r1, r2, ...
vzg_colibri-palindrome checker for input data, i.e., in_gvk_all; in_ul50, ...
vzg_colibri-palindrome maker
vzg_colibri-dno reverser
vzg_colibri-dno reverser
vzg_colibri-dno synthesizer, input data: in_dnosyn, ...
vzg_colibri-count patterns in vc_daygrams
vzg_colibri-count patterns in vc_daygrams
vzg_colibri-make ppn-dno list
vzg_colibri-make ppn-dno list
vzg_colibri-make ppn-dno list
vzg_colibri-make ppn-dno list
vzg_colibri-make set of all dnos in gvk
vzg_colibri-make set of all dnos in gvk
vc_day_cli -v vc_day_port=5550 < in_gvk_all_0-15-02-24 > ou_gvk_all_0-16-03-24 & ...
vc_day_srv K 5550 > ou_gvk_all_0-16-03-24n_log & ...
# 1. FUI, ...
```



References

- Ulrike Reiner: *Computer-Aided Assignment of DDC Numbers*. 32nd Annual Conference of the German Classification Society – AG BIB: Subject Indexing 2008: Accept Progress!, Hamburg, July 17, 2008
[http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend\[id\]=5668](http://www.qucosa.de/recherche/frontdoor/?tx_slubopus4frontend[id]=5668)
- Ulrike Reiner: *Automatische DDC-Klassifizierung bibliographischer Titeldatensätze der Deutschen Nationalbibliografie*. In: Dialog mit Bibliotheken 2010/1, Hrsg. Deutsche Nationalbiblio-thek, ISSN 0936-1138, 22. Jahrgang, S. 23-29. <http://d-nb.info/100980460X/34>
- Top left picture on slide 30 http://option.canada.pagesperso-orange.fr/images/iconographie%20Canada/lake_moraine.jpg