

Workshop on  
"The Use of the DDC in Europe:  
Recent Developments and Future Perspectives"

# Automatic Analysis of DDC Notations

Ulrike Reiner  
VZG Project Colibri/DDC

Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG)  
Platz der Göttinger Sieben 1, 37073 Göttingen



[ <http://www.arthurgrosset.com/sabirds/photos/eupmac4792.jpg> ]

# A "Colibri Flight" Into the DDC

- Colibri/DDC – hummingbird
- 025.431023
- The use of the DDC: **Coding** and **Decoding**

The Classifier: Number building (notational synthesis)

Program: Decomposition of DDC synthesized numbers  
(Liu 1993)

Program: Analysis of DDC notations  
(Reiner 2005 et seqq.)

- Results & Applications



## VZG Project Colibri/DDC ...

... named after a former Pica Project Colibri:

"The Colibri project (Context generation and Linguistic tools for Bibliographic Retrieval Interfaces) aims at the development and operation of a subject-oriented search facility built by Pica and based on a knowledge database of search terms and their mutual relations. ..."

**VZG Project Colibri/DDC: based on the DDC**

Colibri (French) - Kolibri (German) - Hummingbird (English)

[ Pica. Pica Jaarverslag 97 Annual report. ISSN 0168-882, ISBN 90-70311-87-9. Leiden, 1998.  
Online: <http://oclc-pica.org/content/1098/pdf/annualreport97.pdf> ]



"025.431023" ("DDC--Subject as Profession, Occupation, Hobby")

## Widely used DDC System

- Long-standing tradition and still up to date
- Human knowledge organization system
- High complexity
- Notational elegance



# Coding (Synthesis) (1)

**"Froggy went a-courtin'**

**by Kevin O'Malley**

**Language:** English **Type:** Book : Juvenile audience

**Publisher:** New York : Stewart, Tabori & Chang, 1992.

**ISBN:** 1556702604 : 9781556702600 | **OCLC:** 25007770 | Cite this Item

**Subjects:** Folk songs, English -- United States -- Texts. | Folk songs -- United States." (1)

- "Songs" } "Folk music-- ("782.42162")
- "Folk music" } songs" } "American
- "People of United States" } ("782.4216213")
- "Words and other vocal sounds to be sung or  
recited with music" (2)

[ (1) <http://www.worldcat.org/oclc/25007770>

(2) <http://connexion.oclc.org/WebZ/QUERY?sessionid=...&:tdbname=DeweyDB> ]



# Coding (Synthesis) (2)

Class Number: 782.42

Caption: \*Songs

← WebDewey (extract)

## Main Classes

700 Arts & recreation  
 780 Music  
 781-788 Principles, forms, ensembles, voices, instruments  
 782 Vocal music  
 782.1-782.4 Vocal forms  
 782.4 \*Secular forms  
**782.42 \*Songs**  
 782.42079 Festivals--songs, . . .  
 782.42083 Children's songs  
 . . . . .

## Notes

Class here comprehensive works on songs

For sacred songs , see 782.25.

\* **Add as instructed under 782.1-782.4 .**

Segmentation Instruction

=> **782.42**



# Coding (Synthesis) (3)

Centered Entry: 782.1-782.4

Caption: Vocal forms

← WebDewey (extract)

## Main Classes

```

700      Arts & recreation
...
782.1-782.4  Vocal forms
782.1      *Dramatic vocal forms          Operas
...
  
```

## Notes

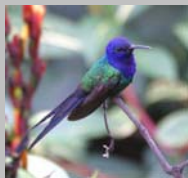
...  
Add to each subdivision identified by \* as follows:

```

01-09 Standard subdivisions
      Notation from Table 1 as modified under 780.1-780.9, ...
1      General principles and musical forms
11-17 General principles
      Add to 1 the numbers following 781 in 781.1-781.7, ...
18     Musical forms
      Add to 18, ... e.g., da capo form 1822, ...
  
```

...

=> **782.421**



# Coding (Synthesis) (4)

Class Number: 781.621-781.629 ← WebDewey (extract)  
 Segmented Number: 781.62/1-781.62/9  
 Caption: Folk music of specific ethnic and national groups ↓

## Main Classes

700 Arts & recreation  
 ...  
 781.6 \*Traditions of music  
 781.62 **Folk music**  
 781.621-781.629 **Folk music of specific ethnic and national groups**  
 ...

## Notes

Add to base number 781.62 notation T5--1-T5--9 from Table 5, ...;  
 then add further as follows:

001-008 Standard subdivisions  
 Notation from Table 1 as modified under 780.1-780.9, ...  
 ...

...

=> 782.42162





# Coding (Synthesis) (5)

Table Number: T5--13 ← WebDewey (extract)  
 Caption: People of United States ("Americans") ↓

## Tables

T5--0	Table 5. Ethnic and National Groups
T5--05-T5--9	Ethnic groups, ...
T5--1-T5--9	Specific ethnic and national groups
T5--1	North Americans
<b>T5--13</b>	<b>People of United States ("Americans")</b>

## Notes

...

=> 782.4216213



# Coding (Synthesis) (6)

Class Number: 780.268

Segmented Number: 780.26/8

Caption: Words and other vocal sounds to be sung or recited with music

← WebDewey (extract)

## Main Classes

700 Arts & recreation

780 Music

780 Music

780.2 Miscellany

780.26 Texts; treatises on music scores and recordings

**780.268 Words and other vocal sounds to be sung or recited with music**

## Notes

...

=> **782.421621300268** (built or synthesized number)  
**"the froggy number"**



# Decoding (1)

## to decompose

Liu 1993

"to divide sth into smaller parts" (1)

## to analyze

Reiner 2005 et seqq.

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

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



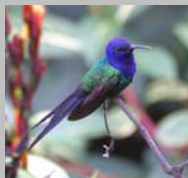
## Decoding (2)

"Hierarchy in the DDC is expressed through structure and notation. ... Structural hierarchy means that **all topics** (aside from the ten main classes) **are part of all the broader topics above them**. The corollary is also true: **whatever is true of the whole is true of the parts**. This important concept is called hierarchical force. ..." (1)

"... the full meaning of a DDC number depends on its position in the hierarchy to which it belongs. Thus, **in many cases, a particular caption can only be understood within the context of its superordinate numbers**." (2)

[ (1) DDC Introduction. Online: <http://www.oclc.org/dewey/versions/ddc22/intro.pdf>

(2) Lois Mai Chan; Joan S. Mitchell: Dewey Decimal Classification: Principles and Application. Third Edition. OCLC Online Computer Library Center, Inc., Dublin, Ohio, 2003, p. 18 ]



# Decoding (Decomposition) (3)

## Liu's Dewey Number Decomposer (DND)

"782.421621300268 has been decomposed as follows:

782.42: \*Songs

11-17: General principles

781.621-781.629: Folk music of specific racial, ethnic, national groups

13: People of the United States ("Americans")

001-008: Standard subdivisions

780.268: Words and other vocal sounds to be sung or recited with music

The title of this book is:

245 10 #aFroggy went a-courtin'#cwritten and illustrated by Kevin

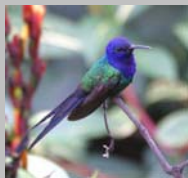
The subject headings for this book are:

#aFolk songs, English#zUnited States#xTexts.

#aFolk songs#zUnited States." (1)(2)

[ (1) Songqiao Liu: The Automatic Decomposition of DDC Synthesized Numbers. Diss., University of California, Los Angeles, 1993, p. 174

(2) Results from "The Classifier" coloured in red by ul ]



# Decoding (Analysis) (4)

## Analysis Diagram in English (vc\_daygram\_eng)

```

782.421621300268 <liu_355_to_analyze; length: 16>
7----- Arts & recreation <hatzen>
78----- Music <hatzen>
782----- Vocal music <hat>
782.4----- Secular forms <hat>
782.42----- Songs <hat>
782.42162----- Folk music--songs,... <hatien> (cf. page 5)
782.4216213----- American folk songs <hatien> (cf. page 5)
782.42----- Songs <nfa->782.1-782.4>
---.---1----- General principles and musical forms <ba5span:782.1-782.4+1>
---.---16----- General principles <ba5span:782.1-782.4+11-17>
---.---6----- Traditions of music <ba5span:782.1-782.4+11-17:na1r5:781.6>
---.---62----- Folk music <ba5span:782.1-782.4+11-17:na1r5:781.62>
---.---621----- Folk music of specific ethnic and national groups <ba5span:782.1-782.4+11-17:na1r5:781.621>
---.---6213----- American folk music <ba5span:782.1-782.4+11-17:na1r5:781.6213>
---.---1----- North Americans <ba4r2span:781.621-781.629:T5--1>
---.---13----- People of United States ("Americans") <ba4r2span:781.621-781.629:T5--13>
---.---002----- Standard subdivisions <ba4r2span:781.621-781.629+001-008>
---.---02----- Miscellany <nst->780.2>
---.---026----- Texts; treatises on music scores and recordings <nst->780.26>
---.---0268----- Words and other vocal sounds to be sung or recited with music <nst->780.268>

```



# Decoding (Analysis) (5)

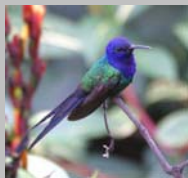
vc\_daygram\_eng ("The Classifier" / Liu's DND)

```

782.421621300268 <liu_355_to_analyze; length: 16>
7----- Arts & recreation <hatzen>
78----- Music <hatzen>
782----- Vocal music <hat>
782.4----- Secular forms <hat>
782.42----- Songs <hat>
782.42162----- Folk music--songs,... <hatien>
782.4216213----- American folk songs <hatien>
782.42----- Songs <nfa->782.1-782.4>
---.---1----- General principles and musical forms <ba5span:782.1-782.4+1>
---.---16----- General principles <ba5span:782.1-782.4+11-17>
---.---6----- Traditions of music <ba5span:782.1-782.4+11-17:na1r5:781.6>
---.---62----- Folk music <ba5span:782.1-782.4+11-17:na1r5:781.62>
---.---621----- Folk music of specific ethnic and national groups <ba5span:782.1-782.4+11-17:na1r5:781.621>
---.---6213----- American folk music <ba5span:782.1-782.4+11-17:na1r5:781.6213>
---.---1----- North Americans <ba4r2span:781.621-781.629:T5--1>
---.---13----- People of United States ("Americans") <ba4r2span:781.621-781.629:T5--13>
---.---002----- Standard subdivisions <ba4r2span:781.621-781.629+001-008>
---.---02----- Miscellany <nst->780.2>
---.---026----- Texts; treatises on music scores and recordings <nst->780.26>
---.---0268----- Words and other vocal sounds to be sung or recited with music <nst->780.268>

```

cf. Decoding (2), page 12



# Decoding (Analysis) (6)

## Liu's DND

782.421621300268 has been decomposed as follows:

782.42: \*Songs

11-17: General principles

781.621-781.629: Folk music of specific racial, ethnic, national groups

13: People of the United States ("Americans")

001-008: Standard subdivisions

780.268: Words and other vocal sounds to be sung or recited with music

## vc\_daygram\_eng (results from Liu's DND)

782.421621300268 <liu\_355\_to\_analyze; length: 16>

782.42----- Songs

---.--16----- General principles <...+11-17>

---.---621----- Folk music of specific ethnic and national  
groups

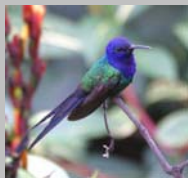
---.-----1----- North Americans <...:781.621-781.629:T5--1>

---.-----13----- People of United States ("Americans")

---.-----002-- Standard subdivisions <...+001-008>

---.-----0268 Words and other vocal sounds to be sung or  
recited with music <nst->780.268>





# Decoding (Analysis) (7)

## Analysis Diagram in German (vc\_daygram\_ger)

```

782.421621300268 <liu_355_zu_analysieren; Laenge: 16>
7----- Kuenste und Unterhaltung <hatzen>
78----- Musik <hatzen>
782----- Vokalmusik <hat>
782.4----- Weltliche Formen <hat>
782.42----- Lieder <hat>
782.42162----- Volksmusik--Lieder,... <hatien>
782.4216213----- Amerikanische Volkslieder <hatien>
782.42----- Lieder <nfa->782.1-782.4>
---.-1----- Allgemeine Prinzipien und musikalische Formen <ba5span:782.1-782.4+1>
---.-16----- Allgemeine Prinzipien <ba5span:782.1-782.4+11-17>
---.-6----- Musiktraditionen <ba5span:782.1-782.4+11-17:na1r5:781.6>
---.-62----- Volksmusik <ba5span:782.1-782.4+11-17:na1r5:781.62>
---.-621----- Volksmusik einzelner ethnischer und nationaler Gruppen <ba5span:782.1-782.4+11-17:na1r5:781.621>
---.-6213----- Amerikanische Volksmusik <ba5span:782.1-782.4+11-17:na1r5:781.6213>
---.-1----- Nordamerikaner <ba4r2span:781.621-781.629:T5--1>
---.-13----- US-Amerikaner <ba4r2span:781.621-781.629:T5--13>
---.-002-- Standardschluessel <ba4r2span:781.621-781.629+001-008>
---.-02-- Verschiedenes <nst->780.2>
---.-026- Texte; Abhandlungen zu Partituren und Musikaufzeichnungen <nst->780.26>
---.-0268 Woerter und andere Laute, die gesungen oder mit Musik vorgetragen werden <nst->780.268>

```



# Decoding (Analysis) (8)

## Another Example

The title of this book is:

#aVoices in architectural education: #bcultural politics and

The subject headings for this book are:

#aArchitecture #xStudy and teaching #zUnited States.

#aArchitecture and state #zUnited States.

**vc\_daygram\_eng** (results from Liu's DND)

720.7073 <ul-liu/liu\_37\_to\_analyze; length: 8>

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

72----- Architecture <hatzen>

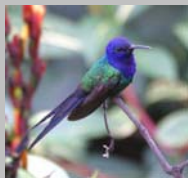
**720----- Architecture <hat>**

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

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

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

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



## Decoding (Analysis) (9)

### Analysis Result Set: DDC Notations & DDC Captions (vc\_dayset\_dno\_cap\_eng)

782.421621300268 <liu\_355\_to\_analyze>

7;Arts & recreation

78;Music

...

780.268;Words and other vocal sounds ...

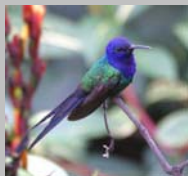
### Analysis Result Set: DDC Notations – Fine Analysis (vc\_dayset\_fine)

liu\_355\_to\_analyze:782.421621300268;7;78;782;782.4;

782.42;782.42162;782.4216213;782.4+1;782.4+16;

781.6;781.62;781.621;781.6213;T5--1;T5--13;

780.2;780.26;780.268



## Decoding (Analysis) (10)

### Analysis Result Set: MAB2 Format (vc\_dayset\_mab2)

1. 705\$a782.421621300268\$p782.42\$c781.62\$d782.4+1;\781.6213;780.268\$l113\$t782.4+16
2. 705\$a782.421621300268\$p782.42\$d782.4+1\$t782.4+16\\$c781.62\$d781.6213\$l113\$d780.268

### Subfields of MAB2 Field 705

\$a: dno            \$b: Item number    \$c: bano

\$d: schedno    \$e: edno

\$f: tabno1    \$g: tabno2            \$h: tabno3A

\$i: tabno3B    \$j: tabno3C            \$k: tabno4

\$l: tabno5    \$m: tabno6            \$p: preno

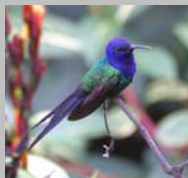
\$t: schedno\_add or tabno\_add

\$A: Source of dno (0: LoC; 1: DNB; 4: other)

dno:	DDC number	bano:	base number	edno:	edition number
schedno:	schedule number	tabno:	table number	preno:	prefix dno (ul)
schedno_add:	add table of schedno	tabno_add:	add table of tabno		

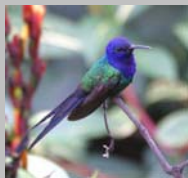
vc\_dayset\_marc21, ...

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



# Analysis of DDC Notations (1)

- [Reiner2005]** Ulrike Reiner: DDC-Notationsanalyse und -synthese. Sept. 2004 - Feb. 2005. VZG-Colibri-Bericht 2/2004. Verbundzentrale des Gemeinsamen Bibliotheksverbundes (VZG), Göttingen, 2005. Online: <http://taipan.dyndns.org/~ul/colibri03-05-04-11.pdf>
- [Reiner2007a]** Ulrike Reiner: Automatische Analyse von Notationen der Dewey-Dezimalklassifikation. PowerPoint Presentation. 31st Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications – Librarian Workshop: Subject Indexing and Library Science. March 7 - 9, 2007, Freiburg i. Br., Germany. Online: [http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf\\_2835.pdf](http://www.gbv.de/vgm/info/biblio/01VZG/06Publikationen/2007/pdf/pdf_2835.pdf)
- [Reiner2007b]** Ulrike Reiner: Automatic Analysis of Dewey Decimal Classification Notations. Full paper of [Reiner2007a] in English. Online: <http://taipan.dyndns.org/~ul/colibri-07-05-22.pdf>



## Analysis of DDC Notations (2)

VZG Project **Colibri/DDC** research questions:

Q1: Is it possible to automatically decompose molecular DDC notations into atomic DDC notations?

Q2: Is it possible to improve automatic classification and retrieval by means of atomic DDC notations?



## Analysis of DDC Notations (3)

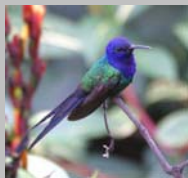
### Atomic DDC notation

An atomic DDC notation is a **semantically indecomposable** string that represents a DDC class.

### Molecular DDC notation

A molecular DDC notation is a string that is **syntactically decomposable** into atomic DDC notations.

[ Russell's Logical Atomism. Online: <http://plato.stanford.edu/entries/logical-atomism/> ]



# Analysis of DDC Notations (4)

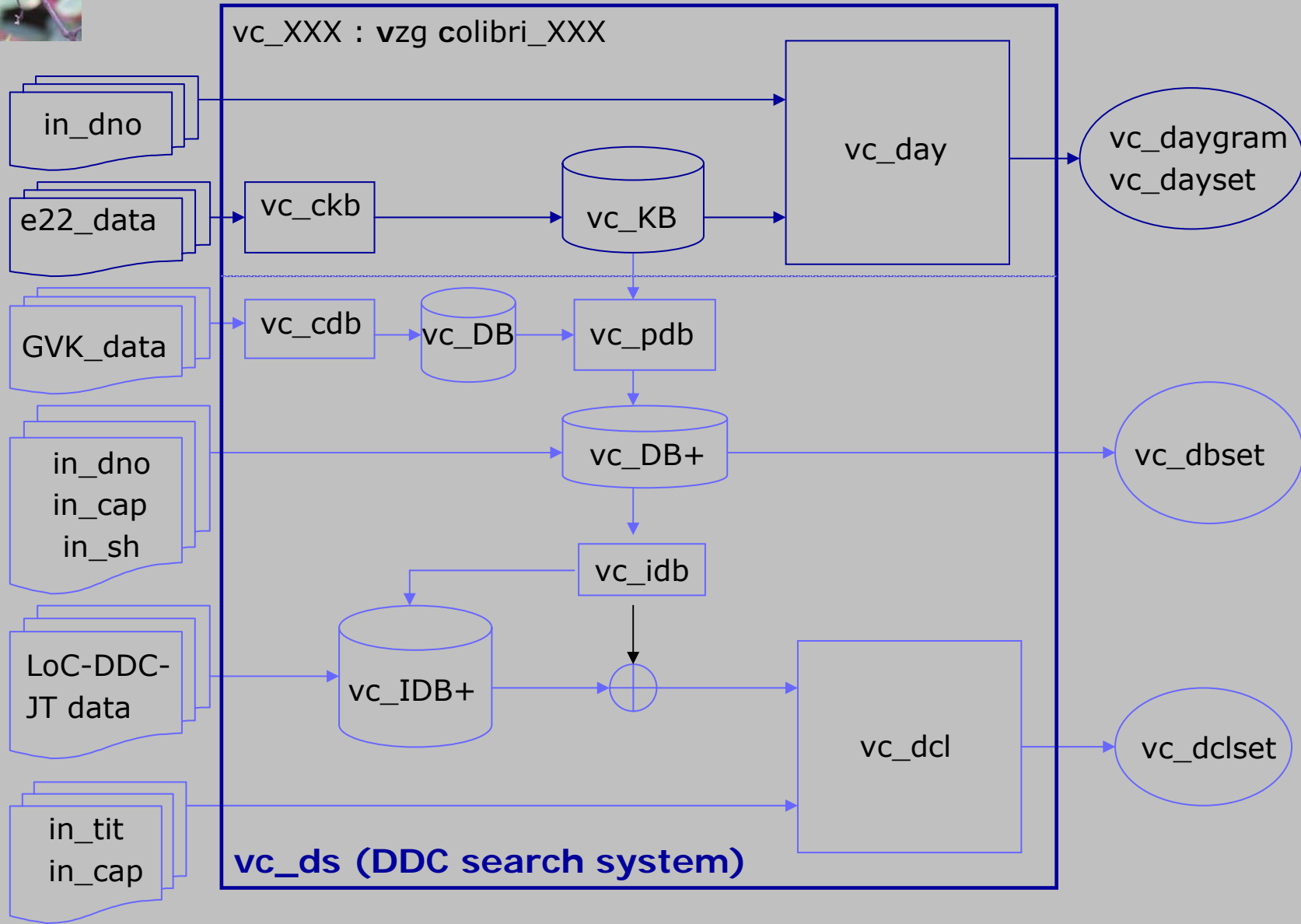
## Abbreviations <sup>(1)</sup>

<b>e22_data</b>	<b>data</b> of the electronic 22nd edition of the DDC system
<b>in_dno</b>	<b>input ddc notations</b>
<b>vc_KB</b>	vzg colibri_ddc <b>K</b> nowledge <b>B</b> ase
<b>vc_ckb</b>	vzg colibri_create ddc <b>k</b> nowledge <b>b</b> ase
<b>vc_day</b>	<b>vzg colibri_ddc number analyzer</b>
<b>vc_daygram</b>	vzg colibri_ddc <b>a</b> nalysis <b>d</b> iagram
<b>vc_dayset</b>	vzg colibri_ddc <b>a</b> nalysis <b>r</b> esult <b>s</b> et
<b>vc_dcl</b>	vzg colibri_ddc <b>c</b> lassification
<b>vc_ds</b>	<b>vzg colibri_dewey search system</b>

[ (1) For the other abbreviations, cf. Ulrike Reiner: Automatische Analyse von DDC-Notationen und DDC-Klassifizierung von GVK-Plus-Titeldatensätzen. Workshop on Dewey Decimal Classification "DDC-Einsichten und -Aussichten 2007". March 1, 2007, SUB Göttingen, Germany. Online: [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_2836.pdf) ]



# Analysis of DDC Notations (5)





# Analysis of DDC Notations (6)

Input Data: Input DDC Notations (in\_dno)  
 (e.g., in\_liu\_t, in\_gvk\_all, in\_ddbt2, in\_ddbH, in\_bow<sup>(1)</sup>, ...)

in_liu_t	in_gvk_all
in_liu_1	in_gvk_all_1
700.23	0
in_liu_2	in_gvk_all_2
700.90440747471	00
in_liu_3	in_gvk_all_3
700.942	000
in_liu_4	in_gvk_all_4
700.972910904	000.00
...	...
in_liu_600	in_gvk_all_466134
799.26092	999.23

[ (1) John H. Bowman: Essential Dewey. Facet Publ., London, 2005 ]

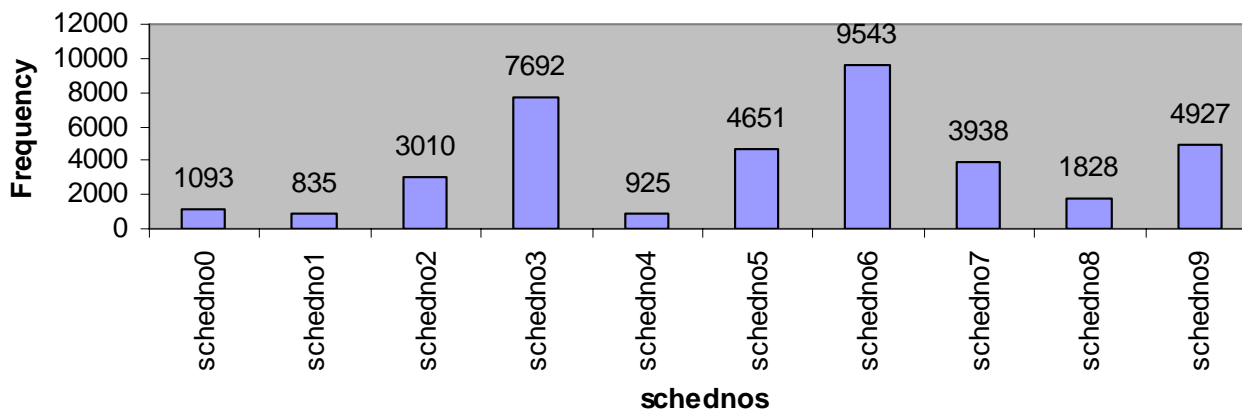




# Analysis of DDC Notations (8)

## Analysis of e22\_data

Frequency distribution of 38,442 Schedule Notations (schednos)



**schednos**

0: Computer science, information & general works

1: Philosophy & psychology

2: Religion

3: Social sciences

4: Language

5: Science

6: Technology

7: Arts & recreation

8: Literature

9: History & geography

**tabnos**

1: Standard Subdivisions

2: Geographic Areas, Historical Periods, Persons

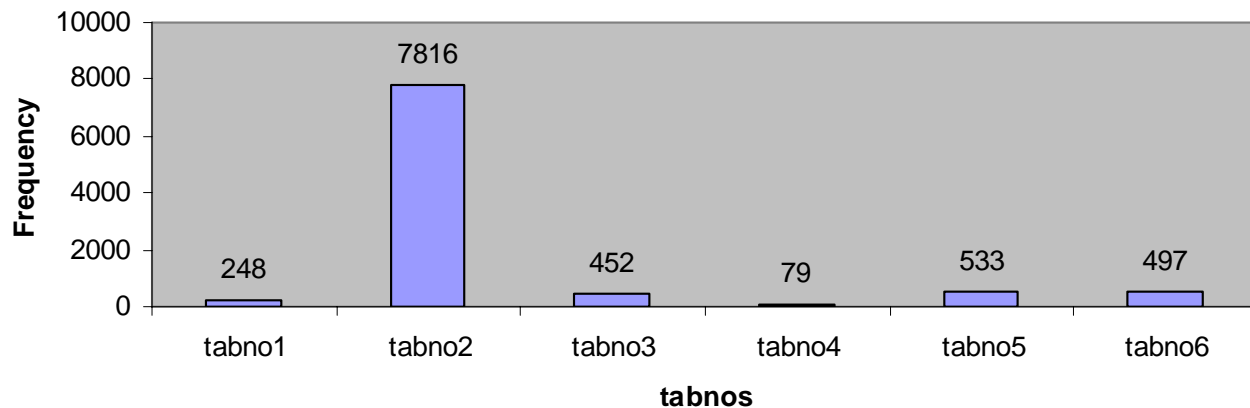
3: Subdivisions for the Arts, for Individual Literatures, for Specific Literary Forms

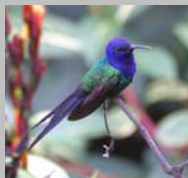
4: Subdivisions of Individual Languages and Language Families

5: Ethnic and National Groups

6: Languages

Frequency distribution of 9,625 Table Notations (tabnos)





# Analysis of DDC Notations (9)

Input Data: Data of the 22nd Edition of the DDC System (e22\_data)  
Approx. 48,000 DDC Notations & 8,000 Instructions

```

...
<ScheduleRecord lang="en" idR="ddc:782.42,, refC="ddc::700" span1="782.42">
...
<dtc><txt>1998-10-27T00:00:00.0-05:00</txt></dtc>
...
<hat><lnk refC="ddc::700" refR="ddc:782.42" span1="782.42"><txt>782.42</txt></lnk>
  <txt>Songs</txt></hat>
...
<nfa><txt>*Add as instructed under </txt><lnk refC="ddc::700" refR="ddc:782.1-.4
  span1="782.1" span2="782.4"><txt>782.1-782.4</txt></lnk><txt>.</txt></nfa>
...
<ien><lnk refC="ddc::700" refR="ddc:782.4216213" span1="782.4216213">
  <txt>782.4216213</txt></lnk></ien>
...
<hat><lnk refC="ddc::700" refR="ddc:782.4216213" span1="782.4216213">
  <txt>782.4216213</txt></lnk><txt>American folk songs</txt></hat>
...
<ba5><txt>Add to each subdivision identified by * as follows:</txt></ba5>
...
<reh><txt>General principles and musical forms</txt></reh>
<ren><txt>11-17</txt></ren>
...
<nal><txt>Add to 1 the numbers following 781 in </txt><lnk refC="ddc::700
  refR="ddc:781" span1="781.1" span2="781.7"><txt>781.1-781.7</txt></lnk><txt>,
  e.g., rock musicÂ 166, rehearsing rock musicÂ 166144</txt></nal>

```



# Analysis of DDC Notations (10)

## Instructions (of e22\_data)

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

Almost all instructions are obtained through the following regular expression:

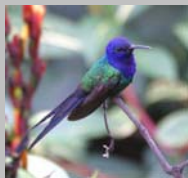
```
colibri/ul-test> egrep -i "add to base number.*(notation | the numbers following.*[notation]*) | add to each subdivision identified by.*(as follows | notation | the numbers following | as instructed (at|under)) | (add [to]* | add to base number).*as instructed (at | under) | (add.*notation.*from.*table.*[under]*) | add [to]*.*(the [historical period | period division]*numbers following | notation)" e22_data
```



# Analysis of DDC Notations (11)

## 26 Macro Rules

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



# Analysis of DDC Notations (12)

## Knowledge Base (vc\_KB)

<b>dno</b> (1)	<b>descriptor</b>	<b>descriptor value(s)</b>
<b>782.42</b>	<b>&lt;hat&gt;</b>	<b>Songs</b>
<b>782.4216213</b>	<b>&lt;hatien&gt;</b>	<b>American folk songs</b>
<b>782.1</b>	<b>&lt;hatspan:782.1-782.4&gt;</b>	<b>Vocal forms</b>
<b>782.42</b>	<b>&lt;nfa&gt;</b>	<b>782.1-782.4</b>
<b>782.1+16</b>	<b>&lt;ba5span:782.1-782.4+11-17&gt;</b>	<b>General principles &lt;na1r5&gt; 1 781.1-781.7 781</b>

### Facts (descriptors beginning with "<hat")

dno **782.42** has the caption **Songs**

dno **782.4216213** is a built number with the caption **American folk songs**

dno **782.1** (out of span of numbers **782.1-782.4**) has the caption **Vocal forms**

### Rules (descriptors not beginning with "<hat")

If dno is **782.42**, then add as instructed under **782.1-782.4**

If dno is **782.1+16** with caption **General principles**,

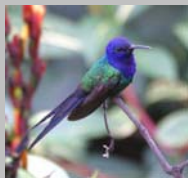
then add to **1** the numbers following **781** in **781.1-781.7**

**vc\_KB** : 48,067 facts, 8,033 rules (2 MB)

**vc\_KB\_spli** : 121,035 facts, 80,324 rules (15 MB)

[ (1) dno: DDC notation ]





# Analysis of DDC Notations (13)

## DDC Analyzer (vc\_day)

### Algorithm

**BEGIN**

1. FUNCTION `fu_init_days_variables`
2. FUNCTION `fu_read_vc_kb_spli`
3. WHILE (true) DO BEGIN
4. WHILE (port SRVPORT 7654 is connected) DO BEGIN
  - a. count `infile_lines`
  - b. determine `dno`, count number of (correct, incorrect) `dnos`
  - c. if (`vc_daygram_test_pri` is 1) then print `vc_daygram_test_pri_header`
  - d. `phase1`
    1. FUNCTION `fu_get_facts` (from left to right)
  - e. `phase2`
    1. initialization
    2. FUNCTION `fu_get_facts_via_rules` (from left to right)
  - f. FUNCTION `fu_chk_pri` (determine output format, print)
  - g. determine whether `dno` is totally complete or partially complete analyzed, count no. of `dnos`
  - h. print -- FINI -- on SRVPORT
  - i. FUNCTION `fu_init_days_variables`

**END**

print "client has closed the connection, classifying has finished"

close SRVPORT

PRINT statistics (number of [totally/partially analyzed `dnos`], ...)

**END**

**END**



# Results (1)

## 1. Decomposition / Analysis of Liu's Sample (600 DDC notations of class 700 ("Arts & recreation" ) )

### Liu 1993

- 1-phase decomposition
- right-to-left algorithm
- decomposition in chunks
- 600 (-27) decompositions of class 700

### Reiner 2005 et seqq.

- 2-phase analysis
- left-to-right algorithm
- decomposition as fine as possible: molecular DDC notations are decomposed into all their atomic DDC notations
- successful reproduction of Liu's results: analysis of all 600 DDC notations of his sample (minor differences: typing errors, different DDC editions)



## Results (2)

### 2. Analysis of GVK DDC notations (GVK data status: July 2004)

466,134 DDC notations analyzed (cf. page 27)  
 297,782 DDC notations totally analyzed (64%)  
 168,352 DDC notations partially analyzed (36%)

DDC class, notation beginning with	No. of DDC notations analyzed	No. of DDC notations totally analyzed
0	23,946	14,887 (62%)
1	4,596	4,095 (89%)
2	20,888	16,122 (77%)
3	189,246	119,378 (63%)
4	6,680	3,124 (47%)
5	33,649	27,706 (82%)
6	51,704	44,879 (87%)
7	52,632	43,378 (82%)
8	20,678	4,205 (20%)
9	62,115	21,036 (34%)

Errors are due to incorrect DDC synthesis or typing errors, to incorrect analysis (incorrect or incomplete analysis algorithm), and to the DDC system.



## Results (3)

### 3. Analysis of DDC notations of DNB file "0605all.ma2"

2,659 DDC notations analyzed

2,210 DDC notations totally analyzed (83%)

449 DDC notations partially analyzed (17%)

### 4. Several output formats

5a. Output as result analysis diagram: `vc_daygram`

5b. Output as result set: `vc_dayset`

### 5. The automatic analysis of DDC numbers seems to be accurate with high probability since it covers Liu's sample that was evaluated by experts (statistical inference).



# Applications

1. Atomic DDC notations could be used to improve recall and precision of information retrieval systems.
2. Atomic DDC notations could be used as search terms for multilingual retrieval.
3. On the basis of analysis diagrams, DDC tutorials or expert systems could be developed to support teaching of DDC number building or to control the quality of built DDC numbers.



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

Increasing the quantity of DDC notations totally analyzed  
will be the next step.

**My thanks go to:**

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

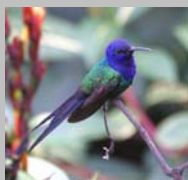
Prof. Dr. Erhard Konrad (retired 2005), Faculty of Electrical Engineering and Computer Science, TU Berlin

Dipl.-Inform. (FH) Alfred Vogelbacher, Network Support Engineer Solaris, Sun Microsystems GmbH, Berlin

M.A. Martin Fischer, Freelance Translator, Berlin



# Extras



# Froggy Went A Courtin'

Frog - gy went(a) court-in'(and) he did ride

Rol-ly-bob-b-rink-tum - ki - no And took Miss Mous-ey by his side

Rol-ly-bob-b-rink-tum - ki - no Ki - no - ka - ro - cap-tain - ka - ro

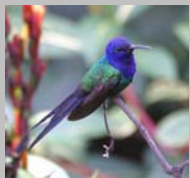
Bom-bin-ee-shee-ki-no Shim-a-rick-i-bom-bin-ick-i Rol-ly-bob-a-rink-tum-

ki - no

a: This note is sometimes sung as a B-flat in some verses.  
 b: This note is sung as a B-flat in some verses, and as a D-flat in others.

[ <http://maxhunter.missouristate.edu/0159/0159.gif> ]





## DDC Notation

```

001A  02003:04-12-01
...
002@   0Aackf
003@   0337713219
...
021A   a@Shining eyes, cruel fortune  dthe lives and loves of Italian Renaissance women
       poets  hlрма B. Jaffe with Gernando Colombardo
028A   dlрма B.  aJaffe
028C   dGernando  aColombardo
033A   pBronx, N.Y  nFordham University Press
...
037A   alncludes bibliographical references and index
044A   Sa  aWomen poets, Italian  Sy  a16th century
044A   Sa  aPoets, Italian  Sy  a16th century
044A   Sa  aWomen and literature  Sz  alItaly  Sx  aHistory  Sy  a16th century
044A   Sa  alItalian poetry  Sx  aWomen authors  Sx  aHistory and criticism
044A   Sa  alItalian poetry  Sy  a16th century  Sx  aHistory and criticism
045A   aPQ4063
045F   a851/.40999999999999999999999999999999287  aB

```

```

851.40999999999999999999999999999999287 (ppn 0337713219)
851.40999999999999999999999999999999287 <dno_gvk_all_397750_to_analyze; length: 31>
8----- Literature <hatzen>
85----- Italian, Romanian & related literatures <hatzen>
851----- Italian poetry <hat>
851.4----- Italian poetry--1542-1585,... <hatien>
while_FUGFA
*** -----> //dno_gvk_all_397750// 851.40999999999999999999999999999999287
      partially complete analyzed ***

```



# MARC 21

## 082 – Dewey Decimal Classification Number

The classification number is taken from  
 -(Abridged) Dewey Decimal Classification and Relative Index  
 - electronic updates to either edition

Through 1997, updates were issued in DC&: Dewey Decimal Classification Additions, Notes and Decisions.

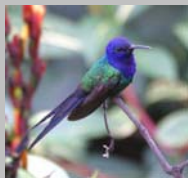
### Indicators

1st - Type of edition (0: Full edition; 1: Abridged edition)  
 2nd - Source of classification number (#: No information provided;  
 0: Assigned by LC, may be used by organizations transcribing from LC copy; 4 - Assigned by agency other than LC)

### Subfield Codes

\$a - Classification number (R)  
 \$b - Item number (NR)  
 \$2 - Edition number (NR)  
 \$6 - Linkage (NR) *See Control Subfields*  
 \$8 - Field link and sequence number (R) *See Control Subfields*

[ following <http://www.loc.gov/marc/bibliographic/ecbdclas.html#mrcb082> ]



## Errors due to the DDC System

"Date: Tue 25 Oct 2005 22:09

...

From: OCLC Customer Support

... We can confirm that it is a typographical error.

We will change 940.54434-940.544349 to

940.54434-940.54439 to correct the problem;

thanks again for letting us know of the mistake

..."



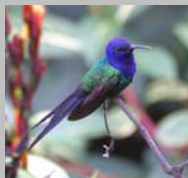
# Suggestion for Text Correction

"8.2 There are **four sources** of notation from which to build numbers: (A) Table 1 Standard Subdivisions; (B) Tables 2-6; (C) **other parts** of the schedules; and (D) add tables **in the schedules**." (1)

## Suggestion:

8.2 There are three sources of notation from which to build numbers: (A) Schedules; (B) Tables; and (C) Internal Tables.

[ (1) <http://www.oclc.org/dewey/versions/ddc22print/intro.pdf> ]



## Liu's Conclusion

"Since the success rate of the decomposition is more than 95% (actually 100%), it is concluded that synthesized DDC numbers can be accurately decomposed by computer. Although the study covers only DDC synthesized class numbers in the main class Arts (700), I strongly believe that synthesized numbers in all the main classes can be accurately decomposed automatically, on the ground that the synthesizing rules for the Arts (700) are representative of the rules throughout the schedule. ... thus, with some modification or augmentation, the decomposing rules that were defined by this study could be extended to all synthesized numbers in all DDC mains classes.

Nevertheless, **it seems a logical next step to extend and apply the methodology to other DDC schedules.** Future research on automatic decomposition of synthesized numbers in the other nine main DDC classes would validate and supplement the findings and conclusions of the present study."

[ (1) Songqiao Liu: The Automatic Decomposition of DDC Synthesized Numbers. Diss., University of California, Los Angeles, 1993, p. 65]