

EDUG 2016 Symposium, Discovering Dewey in Göttingen, 25 April 2016

# Data-Driven Development

**Rebecca Green**

Senior editor, Dewey Decimal Classification

# The future ≠ the past

- In the past . . .
  - 7-year cycle of print editions
  - Each subject area reviewed during cycle
- In the future . . .
  - Focus on areas needing attention
  - Use data to drive development efforts

# Raw data: Dewey numbers

- On a WorldCat record-by-record basis:
  - Filters
    - 082 \$2 = 23\* (e.g., 23/fre)
    - 082/ind1 = 0
  - Data collected
    - 082 \$a
    - Holdings data
    - 082/ind2 (differentiate between LC, non-LC assignment)

# Computed data: schedule numbers (1)

- Computed value favors the following conditions:
  - Schedule number, built numbers in which schedule number is base number assigned many times
  - DDC notation has relatively few subordinate numbers / is not well developed
  - DDC notation lacks instruction to add further notation or the ability to add further notation has not been used often

# From computed data to priority areas

- Identified top 500+ DDC schedule numbers (based on computed value)
- Deleted numbers typically used for form rather than subject (e.g., 791.4372 Single films)
- Grouped together numbers in same parts of schedule

# Top DDC-number-based priority areas

DDC #	Caption	Value	Totnum	Tree size	Atnum (ratio)	SSnum (ratio)	Builtnum (ratio)
158.1	Personal improvement ...	0.87	9599	10519	9088 (0.95)	506 (0.05)	5 (0.00)
006.3	Artificial intelligence	0.67	8151	12593	8086 (0.99)	65 (0.01)	0 (0.00)
005.1	(Computer) Programming	0.46	7379	15880	6272 (0.85)	1102 (0.15)	5 (0.00)
248.4	Christian life and practice	0.45	5022	5840	4852 (0.97)	166 (0.03)	4 (0.00)
658.4092	Executive leadership	0.42	4464	4464	4240 (0.95)	221 (0.05)	3 (0.00)

# 158.1 Personal improvement and analysis

## .1 Personal improvement and analysis

Class here works intended to make one a better person or to stave off failure, to solve problems or to adjust to a life that does not meet one's expectations; works on specific systems and schools of applied psychology written for people who wish to be improved or analyzed

## .12 Personal improvement and analysis through meditation

### .125 Transcendental meditation

### .128 Meditations

# LCSHs with works classed at 158.1

- Aspects to improve: Interpersonal relations, Job satisfaction, Life skills, Self-control, Self-esteem, Spiritual life, Stress management
- Techniques: Meditation, Diaries, Mindfulness-based cognitive therapy, Acceptance and commitment therapy, Client-centered psychotherapy



# 006.3 Artificial intelligence

- .3     **\*Artificial intelligence**
  - Class here computational intelligence, intelligent agents, multi-agent systems, question-answering systems, comprehensive works on artificial intelligence and cognitive science
  
- .31    **\*Machine learning**
  - Including genetic algorithms
  
- .32    **\*Neural nets (Neural networks)**
  - Including perceptrons
  - Class here connectionism, neural computers
  
- .33    **\*Knowledge-based systems**
  - Class here expert systems
  
- .35    **\*Natural language processing**
  - Class here computational linguistics |
  
- .37    **\*Computer vision**

# ACM Computing Classification System

- Artificial intelligence
  - Natural language processing
  - Knowledge representation and reasoning
  - **Planning and scheduling**
  - **Search methodologies**
  - **Control methods**
  - Philosophical/theoretical foundations of artificial intelligence
  - **Distributed artificial intelligence**
  - Computer vision

# 005.1 (Computer) Programming

- .1    **\*Programming**
  - Class here application frameworks, application programming, computer algorithms, integrated development environments, software engineering
- .11   †\*Special programming techniques
- .12   †\*Software systems analysis and design
  - Class here analysis of a user's problem preparatory to developing a software system to solve it
- .13   †\*Programming languages
  - Including application generators, nonprocedural (declarative) languages, text editors specially designed to assist in coding programs
  - Class here coding of programs
- .14   †\*Verification, testing, measurement, debugging
  - Including software metrics
- .15   †\*Preparation of program documentation
  - Including development of online help
  - Class here preparation of software documentation
- .16   †\*Program maintenance
  - Class here software maintenance
- .18   †\*Microprogramming and microprograms
  - Including firmware viewed as microprograms, firmware development, microassembly languages, microcode

# ACM Computing Classification System

- Software creation and management
  - Designing software
  - **Software development process management**
  - **Software development techniques**
  - Software verification and validation
  - **Software post-development issues**
  - Collaboration in software development

# 248.4 Christian life and practice

## .4 Christian life and practice

Class here Christian marriage and family

## .46 Individual observances

Including ceremonial and ritual observances, observance of restrictions and limitations

*For asceticism, see 248.47; for pilgrimages, see 263.041*

## .47 Asceticism

Attitudes and practices aside from and beyond normal moral duties adopted as aids to moral and spiritual development

Including practice of celibacy, fasting and abstinence, poverty, solitude, other physical austerities, e.g., flagellation

*For clerical celibacy, see 253.25; for practices of religious congregations and orders, see 255*

## .48 Guides to Christian life by or for adherents of specific branches, denominations, sects

# Topics with literary warrant at 248.4

- Emotions
- Communication
- Interpersonal relations
- Marriage
- Family
- Discipling
- Spiritual warfare

# 658.4092 Executive leadership

- Of top 500+ numbers, 7.5% are under 658, including 15 under 658.4:

<b>658.4</b>	Executive management
<b>658.4012</b>	Planning and policy making
<b>658.4013</b>	Control and quality management
<b>658.4022</b>	Work teams
<b>658.403</b>	Decision making and information management
<b>658.4038</b>	Information management
<b>658.404</b>	Project management
<b>658.406</b>	Managing change
<b>658.4063</b>	Innovation by management
<b>658.408</b>	Social responsibility of executive management
<b>658.4083</b>	Protection of environment
<b>658.4092</b>	Executive leadership
<b>658.421</b>	Entrepreneurial management
<b>658.45</b>	Communication
<b>658.452</b>	Oral communication

# Available data: LCSH usage

- New LCSHs (exact or with subdivisions) from most recent 5 years
- Assigned to 10+ WorldCat records
- Associated Dewey numbers
  - 082/ind1 = 0
  - 082 \$2 = 23\*
- Holdings



# DDC-LCSH data sample

(1)	(2)	(3)	(4)	(5)	(6)	(7)
Indie pop music	101	744	10	418	3	782.42166 (8), 782.421644 (1), 782.42164 (1)

- (1) LCSH being investigated
- (2) Number of WCBIB records (1) assigned to
- (3) Number of holdings corresponding to (2)
- (4) Number of WCBIB records to which (1) assigned and to which DDC has been assigned
- (5) Number of holdings corresponding to (4)
- (6) Number of different DDCs assigned
- (7) DDCs assigned (number of times each has been assigned with [1])

# Analysis of DDC-LCSH data

- Sort by number-of-records/holdings column; map LCSHs
  - Independent mapping effort
  - Supplement DDC-based priority area development
- Identify other priority areas
  - Multiple new LCSHs in same hierarchy

# Interim results: Additional priority areas (1)

- 004.16 Personal computers
- 005.276 Programming for distributed computing
- 306.7 Sexual relations
- 341.4 Jurisdiction over physical space; human rights
- 341.6 Law of war
- 345.02 Criminal offenses

## Interim results: Additional priority areas (2)

- 363.3 Other aspects of public safety (e.g., terrorism, disasters)
- 364.1 Criminal offenses
- 572.8 Biochemical genetics
- 616.8 Diseases of nervous system and mental disorders
- 618.92 Pediatrics and geriatrics
- 741.5 Comic books, graphic novels, fotonovelas, cartoons, caricatures, comic strips

# Interim results: Additional priority areas (3)

- 782.42166 Rock (Rock 'n' roll) songs
- 791.43 Motion pictures
- 809.933 Literature dealing with specific themes and subjects
- 813.6 American fiction in English, 2000—
- 899.2 Literature of Austronesian languages
- 958.1 Afghanistan

# Other ideas for . . . ?

- Data to gather
- Ways to analyze existing data

# Thank you!

**Rebecca Green**

Senior editor, Dewey Decimal Classification

[greenre@oclc.org](mailto:greenre@oclc.org)

**Because what is known must be shared.<sup>SM</sup>**