DCPAPERS
Proc. Int'l Conf. on Dublin Core and Metadata Applications 2016

Presentation

Data-Driven Development of the Dewey Decimal Classification

Rebecca Green
OCLC Online Computer Library
greenre@oclc.org

Description

Changes involved in maintaining the Dewey Decimal Classification (DDC), a general classification system, have derived in the past from many distinct sources. These include (but are not limited to) questions/ideas/complaints from end users, classifiers, translators, or members of the Decimal Classification Editorial Policy Committee (EPC); mappings of other knowledge organization systems to the DDC; and personal awareness of events, emerging issues, and trends. On the one hand, these phenomena may bring to light ambiguity or redundancy in the current system. On the other hand, they may bring to the attention of the editorial team new topics needing provision within the system.

Without disregarding these sources, the DDC editorial team is also considering data-driven methods of (1) identifying existing areas of the DDC warranting further development or (2) identifying topics with sufficient literary warrant to justify explicit inclusion in the DDC. The use of two sources of data is under investigation.

The first data source reflects the assignment of recently created Library of Congress Subject Headings (LCSHs) to resources described in WorldCat records (i.e., LCSHs added within the past 5 years). Identifiable sets of headings typically not mapped to the DDC (e.g., personal, family, and corporate names) are filtered out; headings are further restricted to those appearing in at least 10 WorldCat records. For these we gather the number of records to which they are assigned, corresponding holdings data, and any numbers from the current full edition of the DDC that are assigned to the same records. Sorted by number of records or holdings, such a headings list helps prioritize development of the DDC by topic. Further massaging of the data in conjunction with the DDC's expressive notation isolates areas of the classification most associated with emerging topics and thereby helps prioritize development by area of the system.

The second data source reflects the assignment of numbers from the current full edition of Dewey to WorldCat records. For each DDC number, we compute a value that favors these conditions:

- The DDC number and built numbers for which the original number is the base number are assigned relatively more often than other DDC numbers.
- The DDC number has been assigned relatively more often than its subordinate numbers.
- The DDC number has been assigned relatively more often than built numbers (including standard subdivisions) for which it is the base number.

Sorting the list of DDC numbers by the computed value helps identify areas within the schedule which are receiving extensive use, but are not well developed.

The topics and schedule areas identified through these means require investigation to ascertain if they are viable candidates for further development. Preliminary work with these data sources reveals that the strategies hold promise.

https://doi.org/10.23106/dcmi.952137674