

# Efficient reclassification of large libraries

Kai Eckert and Magnus Pfeffer  
{lastname}@bib.uni-mannheim.de

University Library  
University of Mannheim, Germany

**Abstract.** In this paper, we describe a software application that was developed and is currently applied at the Mannheim University Library to manually reclassify about 1 million books in an efficient manner by supporting various different working strategies and by using information from several sources.

**Revised author copy of “An Application to Support Reclassification of Large Libraries”, published in S. Gradmann et al. (Eds.): TPDL 2011, LNCS 6966, pp. 461–464, 2011.**

## 1 Background

A common example for the reclassification of large library collections is a switch to the Library of Congress Classification scheme. Using the classification numbers and cutter information provided by OCLC significantly reduces processing time for newly acquired titles and is a logical next step for libraries that use OCLC services for their catalogues. The situation in Germany, however, is different. Historically, no single institution with an authority comparable to the Library of Congress or the British Library emerged. As a consequence, libraries formed union catalogues to facilitate collaboration and reuse, but classification systems only slowly gained acceptance, with many libraries using self-designed or adapted classification systems.

In recent years there has been an increase in the construction of new library buildings that are designed to house the merged contents of several smaller libraries. Having multiple call number systems in an integrated library building severely obstructs the use of materials, and a switch to a unified classification system is deemed necessary. The decision on which classification system to choose is often based on the proportion of already classified titles in the union catalogs. Many libraries choose the Regensburg union classification (abbreviated as RVK) which in recent years has gained some prevalence in Germany [4], so that classification information contributed by other libraries can be reused. But still, RVK class numbers are available for only about 50% of the titles in German union catalogs. This is an additional handicap on top of the already challenging task of reclassifying large libraries as described by recent reports [1, 3].

In this demo we introduce an application that is intended to assist library staff with the reclassification task, i.e. the application of class numbers to individual titles. It is actively developed at the library of the university of Mannheim, Germany and is cur-

rently applied by several librarians who are working on the ongoing reclassification of the roughly 1 million books available in open access areas<sup>1</sup>.

## 2 The application

The main paradigm of the classification application is to support the users in all possible ways while still keeping the whole process strictly intellectual and under the full control of the users.

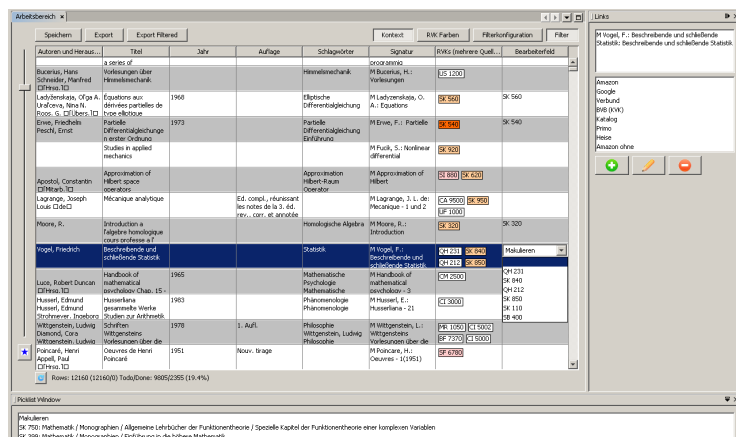


Fig. 1: The main window

**Title List.** Figure 1 shows a screenshot of the application's main window. Most prominent is the central tabular view of all the titles that are to be reclassified. The table view is very fast, even with large lists<sup>2</sup>, and its contents can be sorted by any column. Columns can be hidden or shown and their order rearranged according to individual preferences and needs. In the sample screenshot the following columns are displayed: from left to right "Authors and Editors", "Title", "Year", "Edition", "Subject Headings", "Call Number", "Available Classification Numbers" and "Assigned Classification Numbers". The last field contains the decision of the user about the book represented by the table row. It can be filled in various ways:

- The user can click on one of the classes in the "Available Classification numbers" field.
- The user can select one of the values available from the drop down list. It contains the class numbers from the "Available Classifications" field, plus all recently entered class numbers.

<sup>1</sup> The books are located in four libraries in different buildings on the campus. The closed stacks of the library hold another 1 million books.

<sup>2</sup> Unlike common spreadsheet programs, there is no arbitrary limit of rows. The application works flawlessly with lists containing up to 300.000 titles on an older PC with 2GB of RAM.

- The user can choose a value from the *Picklist*.
- The user can assign a value to all visible rows by *Mass Assignment*.
- Finally, the user can enter arbitrary text, e.g. any class number that is not suggested or a note.

Throughout the application, all the class numbers can be selected to show a pop-up tooltip with the full class definition. So, the users can make themselves familiar with class numbers outside their main field of expertise.

**Picklist.** The *Picklist* is shown in the bottom area of the main window. The users can add any text values here to assign them quickly with a double click to the currently selected row in the table. Typical values in the picklist are a set of classes that are currently assigned to the filtered selection of books plus typical notes for actions like “Maculation” or “Presentation of the book”.

**Links.** The *Links* window is visible on the right-hand side in the main window. It contains links to web resources that provide additional information for the currently selected title in the tabular view. The user can adapt and extend these links. Currently, per default, the following links are implemented: Amazon Book Search, Google (Books), the union catalog, a link to the local OPAC and a link to the local discovery system (Primo).

**Color Highlighting.** Highlighting of class numbers can be configured in various ways. As illustrated in Figure 2a, the user can chose to highlight class numbers that share a common prefix.

**Filters.** As shown in Figure 2b, sophisticated filtering options for all columns are available and can be combined in various ways.

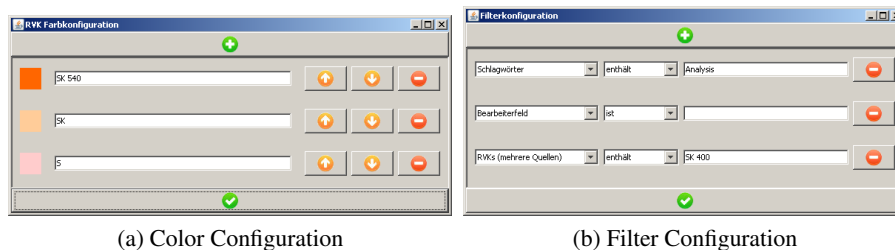


Fig. 2: Adjusting the color highlighting and the filters

### 3 Scenarios

In the ongoing reclassification project, we considered several common scenarios and designed the features of the application to optimally support them.

**Mapping class numbers.** In some cases, a class from the previously used classification system can be mapped congruently to a class from the system used for reclassification, i.e. every book with a certain class number in the old system will get the same

class number in the new classification scheme. For common classification systems, such mapping sets have already been compiled [2] and can be used to speed up the reclassification process. In the application, this is implemented by first filtering the list using the "call number starts with" combination and then using *Mass Assignment* to assign the corresponding class to all selected titles. The same steps can be used for several-to-one mappings. In the case of a one-to-several mapping, the same filter is combined with the *Picklist* to quickly assign the alternative class numbers.

**Reusing available class numbers.** As the classification data in the union catalog is the result of a cooperative effort of the member libraries, different codes of practice and, of course, also mistakes can be found. Therefore, our librarians feel the need to double-check the available class numbers before reusing them. By filtering the list to a single class number and highlighting potentially troublesome class prefixes, a homogenous group of titles can be produced. In this list, titles with implausible and erroneous class numbers are very conspicuous and can easily be identified for a closer inspection.

**User-generated clusters.** In the ongoing use of the application in our library, we noted that humans simply excel at pattern recognition. The librarians were quick to identify authors, terms or term combinations that highly correlate with a given class number and would filter and mass assign accordingly<sup>3</sup>.

## 4 Experimental results and outlook

The application has been in use for several weeks now. As an additional feature, it collects usage statistics to help with the analysis of possible problems and to provide an insight into the usefulness of features. For each discipline a single list of titles is created from the library database and preloaded into the application for the use of the respective subject specialist. Most users started with existing mapping sets to reclassify the "easy" titles and switched to clustering the titles. In this phase of the reclassification process, we saw lasting performance rates of 600 titles/hour with peaks of 1000 titles/hour depending on the cluster size. Our librarians reported that working with the application is efficient and fun, but also tiring as there is no more dead time. The next step in development will be the inclusion of a cutter number generator, so that the application can create complete call numbers for the reclassified titles.

## References

1. Lewis, N., Seago, K.: An Automated Reclassification Project at the University of Kentucky. *Cataloging & Classification Quarterly* 28(4), 117–134 (May 2000)
2. Scott, M.L.: *Conversion Tables: Set- Dewey-LC (volume 2), LC-Dewey (volume 1), Subject Headings, LC and Dewey (volume 3) (v. 1-3)*. Libraries Unlimited (2005)
3. Weaver, M., Stanning, M.: Reclassification project at St Martin's College: a case study. *Library Review* 56(1), 61–72 (2007)
4. Werr, N., Ball, R.: Die "neue" Regensburger Verbundklassifikation (RVK) oder die Zukunft eines Erfolgsmodells. *Bibliotheksdienst* 43(8/9), 845–853 (2009)

<sup>3</sup> As a rather unintended side effect, our librarians used the application to find titles that should be disposed of or moved to the closed stacks and just assigned an appropriate note.