638

Knowl. Org. 44(2017)No.8
M. Lardera, C. Gnoli, C. Rolandi, M. Trzmielewski. Developing SciGator, a *DDC*-based Library Browsing Tool

Developing SciGator,
a *DDC*-Based Library Browsing Tool †

Marco Lardera*, Claudio Gnoli**, Clara Rolandi***, Marcin Trzmielewski****

University of Pavia, Library Service, via Ferrata 1, Pavia, Italy 27100,
*<larderamarco@hotmail.com>, **<claudio.gnoli@unipv.it>,
<clara.rolandi@unipv.it>, *< marcin.trzmielewski@gmail.com>

Marco Lardera has a master's degree in philosophy. He is working at the scientific libraries in the University of Pavia, where he has developed a strong interest in knowledge organization issues. He is contributing to the development and promotion of the online *DDC*-based tool SciGator.

Claudio Gnoli has been an academic librarian since 1994, currently at the Science and Technology Library, University of Pavia, Italy. He has taught various courses and lessons on classification and knowledge organization. He is a specialist in classification and facet analysis on both the theoretical plane and its application to the development of classification systems. He is co-author of books, author of numerous papers in academic journals, and web editor of the ISKO Encyclopedia of Knowledge Organization. He tweets on KO topics as @scritur.

Clara Rolandi studied and researched in the field of musicology (libretti of the nineteenth-century Italian opera). She then worked as a librarian in collaboration with the Brera National Library of Milan and its Musical Collections Research Office. After a period at the Department of Animal Science library, University of Milan, she now works as a cataloguer in the libraries of Law and Economics at the University of Pavia, where she participates in the SciGator project.

Marcin Trzmielewski is a Polish philologist, translator, and librarian. He has previously obtained two master's degrees in French literature in France (University of Montpellier) and in romance philology in Poland (University of Opole). During the last academic year, he specialized in management and dissemination of digital information at the University of Montpellier. His research and main interests include information and communication technology applied to the library system and twentieth-century French literature.

Abstract: Exploring collections by their subject matter is an important functionality for library users. We developed an online tool called SciGator in order to allow users to browse the *Dewey Decimal Classification* (*DDC*) classes used in different libraries at the University of Pavia and to perform different types of search in the OPAC. Besides navigation of *DDC* hierarchies, SciGator suggests "see-also" relationships with related classes and maps equivalent classes in local shelving schemes, thus allowing the expansion of search queries to include subjects contiguous to the initial one. We are developing new features, including the possibility to expand searches even more to national and international catalogues.

† Paper presented at ISKO-Italy: 8' Incontro ISKO Italia, Università di Bologna, 22 maggio 2017, Bologna, Italia.

1.0 Introduction

While many library catalogues (OPACs) include various kinds of information on the subject contents of documents, especially applying classification schemes or verbal subject heading systems, surveys have shown that these are often poorly presented to catalogue users, as interfaces lack explanations of how they work, verbal equivalents to classification numbers that are otherwise obscure to most users, and proper display of relationships be-

Knowl. Org. 44(2017)No.8
M. Lardera, C. Gnoli, C. Rolandi, M. Trzmielewski. Developing SciGator, a *DDC*-based Library Browsing Tool

639

tween classes, both hierarchical and associative (Long 2000; Casson et al. 2011). Good examples exist for each of these functionalities that suggest how a general knowledge organization system, like the Library of Congress *Classification* (Bland and Stoffan 2008) or the Universal Decimal Classification (Pika and Pika-Biolzi 2015; Vukadin 2015), can be leveraged to enrich search and navigation of bibliographic data. However, full exploitation of these systems is often limited by the standard functionalities of the OPAC web interfaces and by problems in the dialogue between the multiple information layers that are implied in a library catalogue, including bibliographic description itself, shelfmark data, management of individual item records, the ways data from several libraries and institutions are integrated into union catalogues, and the web interfaces of these catalogues. This means that, in most cases, end users cannot easily explore the subject contents available in library collections despite the fact that much information on them has been stored in the catalogue records by professional indexers.

The *Dewey Decimal Classification* (*DDC*) is a very popular classification scheme, used by many libraries all around the world for indexing and shelving their resources. In Italy, the *DDC* is the most widespread library classification system, as all its recent editions have been translated by a specialized committee; an Italian version of the WebDewey digital edition is made available for subscription by the Italian Library Association (AIB) in collaboration with the Central National Library of Florence (BNCF).

The *DDC* thus offers the advantage of being well-known among both librarians and users, allowing them to use the same scheme in many different libraries. On the other hand, it also presents some intrinsic disadvantages. Because of its disciplinary hierarchical structure, the *DDC* forces the classifier to index a document in a rigid way, using a single specific class and ignoring the interdisciplinary relationships that are implicit in the subject content of most documents available in a modern library (Gnoli et al. 2015). Also, when shelving a book, one is required to choose a specific physical location where to place it, thus disallowing the user to easily reach related books that fall under different *DDC* classes. Therefore, there is a need to

present the available subjects and the relationships between them in order to allow the user to navigate through similar subjects and make wider searches that take into account the complexity and stratification of human knowledge by following links between classes or expanding a search to include related subjects, a technique known as "query expansion" (Tudhope et al. 2006; Lüke et al. 2012).

2.0 *DDC* at the University of Pavia

The University of Pavia is a well-known institution of Medieval origins, covering most fields of knowledge and especially renowned for research and training in medicine, engineering, and literature. Books and journals are collected in ten different library service points that in recent years have been reorganized into nine main libraries. Until a few years ago, these libraries only used local schemes to organize their collections that were created to satisfy their specific needs without following any recognized standard. No policy of subject indexing was adopted in the university online catalogue (OPAC). That made it very hard to navigate through the entire book heritage of the university library system.

For this reason, on the initiative of librarian Anna Bendiscioli, in the three union libraries covering scientific subjects (the Sciences Library, the Science and Technology Library, and the Medical Library), it has been decided to standardize the shelfmarks, adopting a single scheme based on *DDC* for them. The shelfmark pattern has the form "DDC AUTtit year," where *DDC* is a Dewey number (possibly shortened for subject areas in which the library does not specialize), "AUT" are the first three letters of the first author's or editor's name, "tit" are the first three letters of the first meaningful title word, and "year" is the year of publication, possibly preceded by the volume number and followed by the item number except for the first one (Figure 1). More recently, the Political and Social Sciences Library, led by Luisella Malattia, has also adopted *DDC* for some sections of their collections, although using it as additional subject metadata rather than in shelfmarks.

After this reform, it became possible to develop an online tool, called SciGator (http://scigator.unipv.it), that

☐ 1/4

Lehninger, Albert L. . I principi di biochimica di Lehninger / David L. Nelson, Michael M. Cox. - 4. ed. - Bologna : Zanichelli, ©2006. - XXIII, 1256 p. : ill. ; 28 cm.
Classification: 572 [D] [BIOCHIMICA]

Ateneo. Scienza e Tecnica:	Tamburo 574.1 LEHpri 2006
Ateneo. Scienza e Tecnica:	Tamburo 572 LEHpri 2006-2
Ateneo. Scienza e Tecnica:	Tamburo 572 LEHpri 2006-3

Figure 1. Example of some shelfmarks.

allows the user to navigate through the various sciences by browsing all the Dewey classes used in the university libraries, selecting one of them, and launching a search by subject that extracts from the OPAC all resources having that specific class as the beginning of their shelfmark or as additional metadata. At the same time, we have included in SciGator the possibility of recording relationships with *DDC* classes in different disciplines, or with classes in different local schemes, and of expanding the searches on their basis (Gnoli et al. 2016; Trzmielewski et al. 2017). This is one of the few tools we are aware of that permits the complete presentation of all the subjects in a library system and also provides a system of links between them.

3.0 The SciGator database

The SciGator database, based on MySQL, is composed of a single table containing several fields (Figure 2).
These are quickly described below:

Library: It consists of a number from one to nine, corresponding to the nine main libraries in the University of Pavia. It is only used for non-Dewey shelfmark classes that have not yet been converted into the new shelfmark scheme but are included in SciGator; the number indicates in which library the class is used. In the case of Dewey classes, this field is empty.
Notation: The Dewey class number.
Caption: The verbal equivalent of the class, in Italian.
Captione: The verbal equivalent of the class, in English.
Scopenote: A field for some notes.
Seealso (4 fields): The Dewey numbers of related classes.
Equivalent (2 fields): The Dewey numbers of equivalent classes, used to create a link between old and new shelfmarks.

These fields are used by a PHP script to extract data on the classes used in Pavia libraries and to present them dynamically in hierarchical chains and see-also relationships, as illustrated in the following section.

4.0 The SciGator interface

SciGator adopts an interface structured in a hierarchical form, thus reproducing the structure of *DDC* and of local shelving schemes. The home page shows the top-level *DDC* classes belonging to its one hundred main subdivisions (Figure 3), and clicking on one opens a page with all its subclasses that are included in the database. Recently we also introduced a search box, allowing the user to search for a specific class number or for its verbal equivalent (caption).

SciGator does not include all *DDC* classes but only those that are used to index a significant number of documents (typically at least five) actually owned by the university libraries. Thus, it does not replace WebDewey itself as an indexing tool, rather it focuses the attention of users on a fraction of particular classes that are of interest in their libraries. While navigation of WebDewey step by step, including such features as node labels or scope notes, is a long process requiring a good professional knowledge of the classification structure, navigation in SciGator is designed to be quick and intuitive to the average users of our libraries. For the same reason, class captions are often reformulated or shortened as compared to the official ones, to take into account the context and background of library users. This makes SciGator by no means an official source of information on *DDC* classes.

For each class, on the right side of the page, the corresponding related classes ("seealso" fields) and equivalent classes ("equivalent" fields) are displayed, the first ones preceded by the symbol "→," the second ones by "≈" (Figure 4).

For each class, SciGator, by communicating with the university OPAC through dynamic URLs, allows the user to perform three types of searches:

– "Shelf" search: it finds all the resources with a shelfmark starting with the *DDC* class in exam, applying right truncation of its notation. Thanks to the decimal structure of *DDC* notation, this will retrieve documents indexed by the class itself plus documents indexed by any class subordinate to it, which is more specific than it;
– "Catalogue" search: in addition to the results of the previous search, this search also retrieves the resources that include a *DDC* number as metadata, mostly inherited with records imported from SBN, the national database used by cataloguers. Some *DDC* metadata are now also contributed by the Political and Social Science Library as mentioned above;
– "Expand" search: it even extends the search coverage to the related and equivalent classes as recorded in the database.

Each type of search, therefore, allows the user to progressively extend the search query, thus retrieving more results including some belonging to knowledge areas contiguous to the initial one. This principle is in some way similar to the APUPA model by Ranganathan (1951), who believed that the librarian has the task of expanding the horizons of the user, making him capable of exploring the "penumbral" areas near the "umbral cone" projected by the main subject of the search.

Figure 2. The structure of SciGator database.

Figure 3. The main page of SciGator.

Figure 4. Detail of the search interface.

642

Knowl. Org. 44(2017)No.8
M. Lardera, C. Gnoli, C. Rolandi, M. Trzmielewski. Developing SciGator, a *DDC*-based Library Browsing Tool

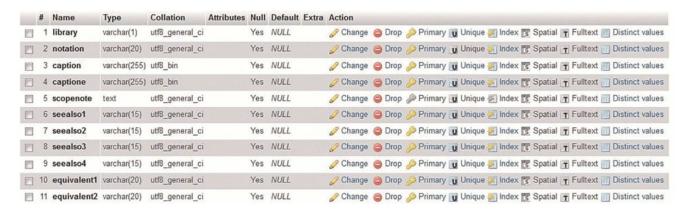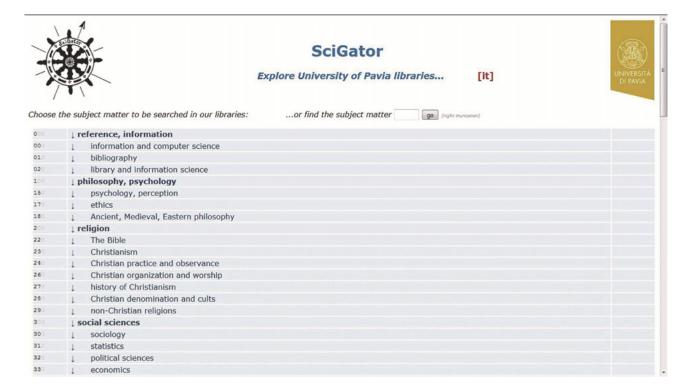The choice of developing a hierarchical interface, instead of a more original modern one, is derived from the nature of the *DDC*, which is suitable for being represented in this way (although interesting circular representations have also been proposed, see Green 2015), as well as the will of giving the user a simple and intuitive tool. The correctness of this choice seems to be confirmed by recent studies that show how this kind of interface brings more satisfaction to the user as compared to other fancy ways of visualizing Dewey classes, such as clouds or maps for example (Lin et al. 2017).

5.0 Limits and future developments

An intrinsic limitation of SciGator is its dependence on the OPAC interface in order to perform searches. Since the OPAC is developed by an external company, we have a very limited control on it and we cannot make direct modifications. This factor limits what SciGator can or cannot do. For example, our OPAC does not support the usage of a wildcard character on the left of a query, meaning we cannot launch searches of the kind "find everything that ends with *XX*," which, considering the rules of Dewey class construction, would be useful to retrieve such specific types of documents as dictionaries (-*03*), exercise books (-*076*) or biographies (-092), or documents treating any subject specifically in the geographical area of Pavia (-*094529*), which are owned in a significant number by our libraries.

An interesting development we are planning is the introduction of two further search types: one allowing to extend searches to SBN (the Italian national union catalogue) and another allowing to move them into WorldCat, the popular international catalogue developed by OCLC. The idea is to suggest a progressive extension of the search scope, from the local library shelves to libraries across the world. Clearly, these different search buttons should be used in their proper sequence, as recommended in the webpage instructions, so that only in case the first searches do not yield satisfying sets of results will the user move to the more expanded ones. On the other hand, if the number of documents assigned to the selected class on the local shelves is high already, expanding the search further would probably yield an exceedingly high recall, also introducing noise in the form of low precision.

Meanwhile, we are progressively introducing into SciGator local non-Dewey schemes used in various scientific and humanistic libraries of our university and mapping them to Dewey classes by creating the appropriate equivalences in the database. While the *DDC* remains the main common language by which the whole of knowledge fields can be navigated, for some *DDC* classes, suggestions of equivalent local classes are also displayed so that local shelves can also be explored by them. Although, ideally, all libraries in the university should adopt the *DDC* at some point in the future, in the meantime, mapping the resources organized by other systems to the common language of the *DDC* seems to work as a useful approximation (cfr. Mayr and Petras 2008).

Another important aspect that we hope to improve is gathering data about the users of SciGator; we can assume that, at the current state, the tool is mainly used by librarians, so a better promotion among other categories (students, researchers) who can benefit from it seems to be desirable.

In conclusion, our experience shows that devoting enough attention to the implementation of a classification scheme, by designing interfaces specifically conceived for its conceptual structure, can allow better leverage of existing subject data and improve the way librarians and users explore and navigate the collections of a system of university libraries.

References

Bland, Robert N. and Mark A. Stoffan. 2008. "Returning Classification to the Catalog." *Information technology and libraries* 27, no. 3: 55-60.

Casson, Emanuela, Andrea Fabbrizzi and Aida Slavic. 2011. "Subject Search in Italian OPACs: An Opportunity in Waiting?" In *Subject Access: Preparing for the Future*. Berlin: De Gruyter Saur, 37-50.

Gnoli, Claudio, Rodrigo de Santis and Laura Pusterla. 2015. "Commerce, See Also Rhetoric: Cross-Discipline Relationships as Authority Data for Enhanced Retrieval." In *Classification & Authority Control, Expanding Resource Discovery: Proceedings of the International UDC Seminar 2015, 29-30 October 2015, Lisbon, Portugal*, ed. Aida Slavic and Maria Inês Cordeiro. Würzburg: Ergon, 151-62.

Gnoli, Claudio, Laura Pusterla, Anna Bendiscioli, and Cristina Recinella. 2016. "Classification for Collections Mapping and Query Expansion." In *NKOS 2016: proceedings of the 15th European Networked Knowledge Organization Systems Workshop, Hannover, Germany, September 9, 2016*, ed. Philipp Mayr, Douglas Tudhope, Koraljka Golub, Christian Wartena and Ernesto William De Luca. CEUR Workshops Proceedings. http://ceur-ws.org/Vol-1676/paper3.pdf

Green, Rebecca. 2015. "Relational Aspects of Subject Authority Control: The Contributions of Classificatory Structure." In *Classification & Authority Control: Expanding Resource Discovery: Proceedings of the International UDC Seminar 2015, 29-30 October 2015, Lisbon, Portugal*, ed. Aida Slavic and Maria Inês Cordeiro. Würzburg: Ergon, 39-52.

Knowl. Org. 44(2017)No.8

M. Lardera, C. Gnoli, C. Rolandi, M. Trzmielewski. Developing SciGator, a *DDC*-based Library Browsing Tool

643

Lin, Xia, Michael Khoo, Jae-Wook Ahn, Doug Tudhope, Ceri Binding, Diana Massam and Hilary Jones. 2017. "Mapping Metadata to DDC Classification Structures for Searching and Browsing." *International Journal on Digital Libraries* 18: 25-39.

Long, Chris Evin. 2000. "Improving Subject Searching in Web-Based OPACs: Evaluation of the Problem and Guidelines for Design." In *Internet Searching and Indexing: The Subject Approach*, ed. Alan R. Thomas and James R. Shearer. New York: Haworth, 159-86. Also in *Journal of Internet Cataloguing* 2, nos. 3-4: 159-86.

Lüke, Thomas, Philipp Schaer and Philipp Mayr. 2012. "Improving Retrieval Results with Discipline-specific Query Expansion." In *Theory and Practice of Digital Libraries: Second International Conference, TPDL 2012, Paphos, Cyprus, September 23-27, 2012: Proceedings*, ed. Panayiotis Zaphiris, George Buchanan, Edie Rasmussen and Fernando Loizides. Lecture Notes in Computer Science 7489. Berlin: Springer, 408-13.

Mayr, Philipp and Vivien Petras. 2008. "Cross-concordances: Terminology Mapping and its Effectiveness for Information Retrieval." Paper presented at the World Library and Information Congress: 74th IFLA General Conference and Council, 10-14 August 2008, Québec, Canada. http://www.ifla.org/IV/ifla74/index.htm

Pika, Jiri and Milena Pika-Biolzi. 2015. "Multilingual Subject Access and Classification-Based Browsing through Authority Control: The Experience of ETH-Bibliothek, Zürich." In *Classification & Authority Control: Expanding Resource Discovery: Proceedings of the International UDC Seminar 2015, 29-30 October 2015, Lisbon, Portugal*, ed. Aida Slavic and Maria Inês Cordeiro. Würzburg: Ergon, 99-110.

Ranganathan, Shiyali Ramamrita. 1951. *Classification and Communication*. Delhi: University of Delhi.

Tudhope, Douglas, Ceri Binding, Dorothee Blocks, and Daniel Cunliffe. 2006. "Query Expansion via Conceptual Distance in Thesaurus Indexed Collections." *Journal of Documentation* 62: 509-33.

Trzmielewski, Marcin, Claudio Gnoli, Marco Lardera, Gaia Heidi Pallestrini and Matea Sipic. 2017. "Mapping Classifications and Linking Related Classes through SciGator, a DDC-Based Browsing Library Interface." *Catalogue and Index* 188: 30-33.

Vukadin, Ana. 2015. "The Development of a Classification Oriented Authority Control: The Experience of National and University Library in Zagreb." In *Classification & Authority Control: Expanding Resource Discovery: Proceedings of the International UDC Seminar 2015, 29-30 October 2015, Lisbon, Portugal*, ed. Aida Slavic and Maria Inês Cordeiro. Würzburg: Ergon, 111-21.