Obtaining the Dewey Decimal Classification Number from other databases: a catalog clean-up project

Stefano Bargioni, Michele Caputo, Alberto Gambardella, Luigi Gentile

1 Introduction

The Library of the Pontifical University Santa Croce\(^1\) is a research library that belongs to the URBE (Roman Union of Ecclesiastical Libraries) Network.\(^2\) It possesses approximately 167,000 volumes corresponding to 145,000 bibliographical records cataloged in MARC21 format. In order to manage the library there have been three Integrated Library Systems (ILS): Aleph 300, Amicus 3.5.4, and the current Koha 3.2.7. With the implementation of the open source ILS, Koha,\(^3\) authority records were then introduced thanks to the software’s advanced productivity. Moreover, Koha’s flexibility has enabled the opening of new avenues for experimentation which is ordinarily impossible with a commercial ILS. With the intention of providing users with greater tools for catalog research from a semantic point of view, and bearing in mind that subject cataloging

\(^1\)http://www.pusc.it/bib.
\(^2\)http://www.urbe.it.
\(^3\)http://koha-community.org.
based on the Nuovo Soggettario of the Central National Library of Florence is fairly recent, the decision was made to develop the potential inherent in the Dewey decimal classification,\(^4\) which had already been partially implemented in the library for about ten years and was assigned to approximately 25\% of the documents in the library’s patrimony. Thus, the idea developed to increase the use of the Dewey decimal classification in the bibliographical records by importing the relevant information from other databases,\(^5\) using the International Standard Book Number (ISBN)\(^6\) as a key for the retrieval of missing numbers. We began by identifying the sources (databases) that would significantly meet our needs, both in terms of quality and quantity. The practice of copy cataloging — one of Koha’s strengths — was fundamental in this regard. Once both the national and international resources were determined, methods were identified through which it was possible to access the information therein by means of a program. As the various institutions use different modalities to publish their data, it became necessary to diversify the query methods in order to systematically gain access to the relevant information. These ranged from the more modern example of OCLC, which gave rise to Classify,\(^7\) a specific experimental web service for classification, to less simple cases of information retrieval from HTML pages. In order to ensure the quality of the

\(^4\)http://dewey.info.
\(^5\)The importation of data from other bibliographical sources is justified by the "principle of sharing", which one finds in public catalogs. This principle establishes the exchange of information through OPAC, Z39.50, web interfaces, etc. It also has as its objective the comparison and mutual control of registration and identification of the library information source, confirmed by field 035 in MARC21, for example. The importation of data occurred in accordance with the possible conditions or warnings expressed on the webpages of the queried sites. The case of a commercial use of the information retrieved could be different.
\(^6\)http://www.isbn.org/standards/home/index.asp.
\(^7\)http://classify.oclc.org.
Dewey classification numbers obtained, a special algorithm was created, which is described in the section “Quality Control”. The process of searching for and importing data was also analyzed under the stress it incurred both for the system that was the source of the data as well as for our Koha system. Queries of the servers cannot occur at an excessive pace, and that is why some of them expressly issue warnings to any possible software, such as crawlers or web robots, which access them.

2 Identifying the records to be modified

The records to be enhanced contained an ISBN (tag 020), but lacked a Dewey number (tag 082). They may be identified in Koha through an SQL query (listing 1) that is specific to the MySQL database and which is applied to the marcxml field of the biblioitems table.

Listing 1: Query for record identification in Koha.

```sql
SELECT biblionumber, ISBNlist
FROM biblioitems
WHERE isbn_present
AND dewey_absent
AND language_008='...' 
```

Since it was not an index driven search, the retrieval occurred via a record by record analysis of the database. This is an aspect of the project which depends on the computing power of the server hosting the ILS. Other ILSes allow to find the system number and the ISBN of a record without a Dewey classification number in a

---


9Main elements of the query are described in Table 9 on page 194.

---
manner quite different from that of Koha, due to the data structure used to store bibliographical data and the tools available to access it.

3 Sources

The ISBNs of each record, extracted from the query, were used to search seven different databases. The sources selected are listed in Table 1 in the temporal order of the query.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Classify</td>
<td>OCLC Classify</td>
</tr>
<tr>
<td>2</td>
<td>LC</td>
<td>Library of Congress</td>
</tr>
<tr>
<td>3</td>
<td>BNF</td>
<td>Bibliothèque nationale de France</td>
</tr>
<tr>
<td>4</td>
<td>DNB</td>
<td>Deutsche Nationalbibliothek</td>
</tr>
<tr>
<td>5</td>
<td>BNCF</td>
<td>Biblioteca Nazionale Centrale di Firenze</td>
</tr>
<tr>
<td>6</td>
<td>BNCR</td>
<td>Biblioteca Nazionale Centrale di Roma</td>
</tr>
<tr>
<td>7</td>
<td>BNB</td>
<td>British National Bibliography</td>
</tr>
</tbody>
</table>

**Table 1:** Dewey classification sources queried.

As the purpose of our work was essentially practical, no attempt was made to query each source with the same ISBN. In the event that a Dewey Decimal number should be retrieved and saved in a record, it was decided that each particular source would take priority over those following, so that the record would not be further processed. This way seemed more efficient to us than the other two possibilities, i.e. to query all sources with the same ISBN, either simultaneously or in succession. Moreover, in several cases the search was limited to the predominant language of the source queried, both to avoid an excessive number of searches and because it was deemed more reliable. Among the languages present in the catalog, Spanish language was not incorporated due to the absence of databases we consider adequate for this purpose. The method adopted does not allow for
comparisons among different sources on equal terms. However, a statistical analysis regarding the use of the Dewey number in different sources is still made possible, as will be seen later.

Figure 1 shows the address, the type of data returned, the type of service contacted for each source, and the language involved.

Sources other than the web provide connection data on their respective pages explaining the service. For web sources, however, connection and query are generally collected from the Advanced Search screen of the catalog. In order to identify the parameters to be sent, including the ISBN, one may proceed in any of the ways listed in the second paragraph of the Appendix. In the case of web pages, the technique adopted for the extraction of data is very specific. One

<table>
<thead>
<tr>
<th>Source</th>
<th>Type</th>
<th>Connection</th>
<th>Query</th>
<th>Data</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Classify</td>
<td>REST</td>
<td><a href="http://classify.oclc.org/classify2/Classify?summary=false&amp;isbn=75562">http://classify.oclc.org/classify2/Classify?summary=false&amp;isbn=75562</a></td>
<td>find $attr 1=7 75562</td>
<td>XML</td>
<td>All</td>
</tr>
<tr>
<td>2 LC</td>
<td>Z39.50</td>
<td>loc.law.gov:210/LCDB</td>
<td>find $attr 1=7 75562</td>
<td>MARC</td>
<td>All</td>
</tr>
<tr>
<td>3 BNF</td>
<td>Z39.50</td>
<td>z3950 bnf fr:2211/700T-UTF8 user id: Z3950password Z3950 BNF</td>
<td>find $attr 1=7 75562</td>
<td>MARC</td>
<td>All</td>
</tr>
<tr>
<td>4 DNB</td>
<td>web</td>
<td><a href="https://portal.dnb.de/opac_htx?query=isbn">https://portal.dnb.de/opac_htx?query=isbn</a> %20/55%method=simpleSearch</td>
<td></td>
<td>HTML</td>
<td>deu</td>
</tr>
<tr>
<td>5 BNF</td>
<td>web</td>
<td><a href="http://croc.bnf.fr/renzo_bnf_it/opac/controller.jsp?action=search">http://croc.bnf.fr/renzo_bnf_it/opac/controller.jsp?action=search</a> advanced&amp;searchquery fieldname key=keywords&amp;fieldname=identite&amp;query querstrina 1=identite%40%3A%2F%2F</td>
<td></td>
<td>HTML</td>
<td>ita</td>
</tr>
<tr>
<td>6 BNF</td>
<td>web</td>
<td><a href="http://193.296.215.172/BNE/result.php?text=expert=1%3Disbn%3Ccallback%3E1%3Dcode=">http://193.296.215.172/BNE/result.php?text=expert=1%3Disbn%3Ccallback%3E1%3Dcode=</a> completo%3Equencode=10%3Cscriptlanguage=it%3ElastRefinedQueryRPC=1%3Disbn%3Cercato%3Er %3Disbn%3Cnqueresults%3E1%3Estart%3Ereport&amp;query%4user=1%3Is%3Cscriptlanguage=it%3EformatShr%3E&amp;sortFormat=desc</td>
<td></td>
<td>HTML</td>
<td>ita</td>
</tr>
<tr>
<td>7 BNF</td>
<td>Z39.50</td>
<td>z3950cat.bl.uk:9909/BNF 0</td>
<td>find $attr 1=7 75562</td>
<td>MARC</td>
<td>eng</td>
</tr>
</tbody>
</table>

Figure 1: Characteristics of the queried sources containing a Dewey decimal classification number.
must apply what is normally referred to as web scraping,\textsuperscript{10} screen scraping, or more commonly data scraping. Essentially, it is necessary to understand whether one has a method for locating and extracting the data of interest from within the HTML code obtained. This operation is easier and more standardized when responses provided are structured data. Web 2.0 and even more, linked data, lead to the hope that the data sources might offer not only web interfaces, essentially intended for human use, but above all, interfaces with standard structured responses that are serviceable by other machines and stable in time. The logic used in the programs to query the data sources can be explained in the algorithm represented in figure 2.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{algorithm.png}
\caption{Representation of the logic used in the programs to query the data sources.}
\end{figure}

There is an exception in the case of Classify, which has already been mentioned above. The process of “querying the data source by the current ISBN” must be followed by:

The third paragraph of the Appendix provides examples for each of the three types of data obtained as a response: XML, MARC, and HTML.

\textsuperscript{10}http://en.wikipedia.org/wiki/Web_scraping.
The response in Classify$^{11}$ typically falls into four categories, as per the table 2.

<table>
<thead>
<tr>
<th>Response code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ISBN corresponds to a single work</td>
</tr>
<tr>
<td>4</td>
<td>ISBN corresponds to several works</td>
</tr>
<tr>
<td>101</td>
<td>ISBN incorrect</td>
</tr>
<tr>
<td>102</td>
<td>ISBN not found</td>
</tr>
</tbody>
</table>

Table 2: Categories of responses in Classify.

In the event of the response “ISBN corresponds to several works,” Classify$^{12}$ provides a list of OCLC# identifiers for related works. The first of these was preferred while locating the detailed record through its OCLC# with another query such as: http://classify.oclc.org/classify2/Classify?summary=false&swid=OCLC#.

This generates a response in the form of code 2, “ISBN corresponds to a single work”.

Classify’s response in the case of a single work (an example of which can be seen in paragraph 3 of the Appendix) reports both...

---

$^{11}$Classify APIs are described at http://classify.oclc.org/classify2/api_docs/index.html and may be tested through the Classify API Explorer at http://classify.oclc.org/classify2/api_docs/classify.html.

$^{12}$The aggregations in Classify occur through the application of FRBR. At http://www.oclc.org/research/activities/classify.html it states: “Bibliographic records are grouped using the OCLC FRBR Work-Set algorithm http://www.oclc.org/research/activities/frbralgorithm.html to form a work-level summary of the class numbers and subject headings assigned to a work. You can retrieve a summary by ISBN, ISSN, UPC, OCLC number, author/title, or subject heading.”
the combination of the Dewey number and the LCC classification assigned to a particular work by the many catalogs which contribute to OCLC, as well as a list of editions containing the classification number. It seemed preferable to import the number from the first edition in the list as it was often more complete in comparison with the others.

Z39.50 sources essentially require extracting the tag value of the Dewey number, according to the rules of the corresponding MARC format, as shown in Table 3.

<table>
<thead>
<tr>
<th>MARC format</th>
<th>tag</th>
<th>code subfield</th>
<th>edition subfield</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARC21</td>
<td>082</td>
<td>a</td>
<td>2</td>
</tr>
<tr>
<td>InterMARC or</td>
<td></td>
<td></td>
<td>v</td>
</tr>
<tr>
<td>UNIMARC</td>
<td>676</td>
<td>a</td>
<td>v</td>
</tr>
</tbody>
</table>

Table 3: Tags for the Dewey classification number in some MARC formats.

4 Quality Control

Before the project, Dewey decimal numbers referencing editions 19 through 23 populated the catalog. The decision not to introduce classification numbers from the abridged version or classification numbers of Dewey editions below 19 meant having to give up several classification numbers found, as reported in the statistics in Table 7 on page 188. Priority was assigned to quality rather than quantity in order to effect an enhancement that is more suitably aligned with the cataloging approach. In reality, beyond limiting the edition to 19 or higher, classification numbers with indicators 1 and
2 — different from “0 0” and “0 4”\textsuperscript{13} — were discarded. Classification codes containing non-numerical characters or lacking an edition were also discarded. Finally, classification numbers were standardized before being saved into the record.

5 Tag 035

While updating the record, it seemed appropriate to keep track of the details from which the imported Dewey classification number was obtained with the help of tag 035 in MARC21, as in the following example:

Listing 2: Example of the use of MARC21 tag 035.

\begin{verbatim}
00872nam a2200265 i 4500
001 00000035650
003 IT-RoPUS
005 20121112122621.0
008 041027r19851982xxk u000 u eng c
020 $a 0198247761
035 $a (OCoLC)007946090
040 $a IT-RoPUS $b ita
082 04 $a 111.85 $2 19
100 1 $a Savile, Anthony. $9 70779
245 14 $a The test of time : $b an essay in philosophical aesthetics / $c Anthony Savile.
\end{verbatim}

\textsuperscript{13}According to MARC21, the first indicator of field 082 with a “0” value signals use of the complete Dewey edition; the second indicator with a “0” value points to a Dewey number assigned by the Library of Congress, while the value “4” corresponds to a notation assigned by an agency other than the Library of Congress.
In the case of a non-MARC21 source, or one without a MARC Organization Code, it was decided to assign the most logical code possible, as shown in Table 4.

**Table 4:** Institution codes in 035.

<table>
<thead>
<tr>
<th></th>
<th>Institution Code</th>
<th>Organization Code</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Classify di OCLC</td>
<td>OCoLC</td>
<td>Official</td>
</tr>
<tr>
<td>2</td>
<td>Library of Congress</td>
<td>DLC</td>
<td>Official</td>
</tr>
<tr>
<td>3</td>
<td>Bibliothèque nationale de France</td>
<td>FR-PaBFM</td>
<td>Official</td>
</tr>
<tr>
<td>4</td>
<td>Deutsche Nationalbibliothek</td>
<td>DE-101</td>
<td>Official</td>
</tr>
<tr>
<td>5</td>
<td>Biblioteca Nazionale Centrale di Firenze</td>
<td>BNCF</td>
<td>Unofficial</td>
</tr>
<tr>
<td>6</td>
<td>Biblioteca Nazionale Centrale di Roma</td>
<td>BNCR</td>
<td>Unofficial</td>
</tr>
<tr>
<td>7</td>
<td>British National Bibliography</td>
<td>BNB</td>
<td>Unofficial</td>
</tr>
</tbody>
</table>

*a [http://dispatch.opac.d-nb.de/DB=1.2/LNG=EN](http://dispatch.opac.d-nb.de/DB=1.2/LNG=EN).*

The ID was derived from the record, which in each case appeared in different locations. For Z39.50 sources it is located in tag 001, while the Library of Congress makes use of tag 010. Classify also specifically reports this ID in the XML record, while ID retrieval from records in HTML format is particularly complex. This decision enables the linking of the bibliographical record to that of an external catalog. This is useful for creating a link of interest whether at the level of OPAC (figure 4 on page 186) or linked data.

A link in the OPAC is created — for every occurrence of tag 035 — on the basis of the links from Table 5 on the facing page. The permanence of some is ensured (permalink). In other cases, the link — which is essentially unstable — can be created from the simple view of each individual record offered in the catalog.

---

Table 5: Creation of a link in OPAC from an occurrence of tag 035.

<table>
<thead>
<tr>
<th>Library</th>
<th>Link</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classify di OCLC - WorldCat</td>
<td><a href="http://www.worldcat.org/search?q=no%3AID">http://www.worldcat.org/search?q=no%3AID</a></td>
<td>permalink&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Library of Congress</td>
<td><a href="http://lccn.loc.gov/ID">http://lccn.loc.gov/ID</a></td>
<td>permalink&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Bibliothèque nationale de France</td>
<td><a href="http://catalogue.bnf.fr/servlet/biblio?idNoeud=1&amp;SN1=0&amp;SN2=0&amp;host=catalogue&amp;ID=ID">http://catalogue.bnf.fr/servlet/biblio?idNoeud=1&amp;SN1=0&amp;SN2=0&amp;host=catalogue&amp;ID=ID</a></td>
<td></td>
</tr>
<tr>
<td>Deutsche Nationalbibliothek</td>
<td><a href="http://d-nb.info/ID">http://d-nb.info/ID</a></td>
<td>permalink&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Biblioteca Nazionale Centrale di Firenze</td>
<td><a href="http://opac.bncf.firenze.sbn.it/opac/controller.jsp?action=notizia_view&amp;notizia_idn=ID">http://opac.bncf.firenze.sbn.it/opac/controller.jsp?action=notizia_view&amp;notizia_idn=ID</a></td>
<td></td>
</tr>
<tr>
<td>British National Bibliography</td>
<td><a href="http://search.bl.uk/primo_library/libweb/action/search.do?vid=BLBNB&amp;fn=search&amp;vl%28freeText/zero.noslash%29=ID">http://search.bl.uk/primo_library/libweb/action/search.do?vid=BLBNB&amp;fn=search&amp;vl%28freeText/zero.noslash%29=ID</a></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> [http://www.oclc.org/worldcatorg/linking/how.htm#oclc-number.](http://www.oclc.org/worldcatorg/linking/how.htm#oclc-number.)

<sup>b</sup> [http://lccn.loc.gov/lccnperm-faq.html.](http://lccn.loc.gov/lccnperm-faq.html.)

<sup>c</sup> Concluded from the simple view of a single record at the end of any type of search.
6 Delay while searching sources

As mentioned in the Introduction, continual use may burden the queried server. This happens quite easily in the case of automated searches. Web pages such as “Terms and Conditions” allow the sources’ terms of use to be regulated. For example, the Library of Congress explicitly\(^{15}\) requires that crawlers use the Z39.50 server at a rate below 10 queries per minute. The Z39.50 server of the Bibliothèque nationale de France shuts down the connection after the tenth query. The program must then reopen the connection with the same frequency. It is not possible for the website of the Biblioteca nazionale centrale di Firenze to be accessed uninterruptedly, since it seems to be overloaded almost immediately. It is also opportune to verify, through the sources queried via http protocol, whether or not there are indications for crawlers in the file /robots.txt. At times restrictions on the frequency of access\(^{16}\) can also be found.

Therefore, a wait time of 4 to 6 seconds between queries was es-

\(^{15}\)http://lccn.loc.gov/lccnperm-faq.html#n12.

established for all sources. These intervals prevented our catalog from being overloaded as well. In fact, after every record modification, the Zebra\textsuperscript{17} search engine used by Koha and the independent search engine\textsuperscript{18} for ordered lists update their indexes and may slow down both the consultation of OPAC and regular usage. This is one aspect which must be assessed on the basis of the available processing power.

The pace imposed by these intervals actually prolongs the import process by hours if not days, depending upon the quantity of ISBNs to be processed. Such a pace may require adjustments to the program, e.g., by setting up parameters so that it operates only on a certain timetable.

7 Log

The import process was monitored in order to collect statistics on the work carried out. The types of log records listed in table 6 on the next page were recorded.

\textsuperscript{17}http://www.indexdata.dk/zebra.

\textsuperscript{18}At present, scrollable index searches – also known as browse searches - are not available in Koha. It was possible to add this feature to our installation of Koha through an application based on Solr [http://www.lucene.apache.org/solr/] and developed by our library. This browse feature was presented at the international meeting of Koha users held in Edinburgh in June 2012 [http://www.wiki.koha-community.org/wiki/KohaCon12_Schedule#Adding_browse_to_Koha_using_Solr_2815-20_min.29], and will be integrated into incoming versions of Koha, particularly when Solr will be an alternative to, or substitute for, Zebra.
The generated logs facilitated the creation of the following tables and certain comparisons between the different sources.

Table 6: Types of log records; types 2 and 3 are only related to Classify.

8 Statistics

The generated logs facilitated the creation of the following tables and certain comparisons between the different sources.

Table 7: Calculations.

Table 8: Distribution of editions related to the classification numbers found.

Table 8 has been reproduced in the graphs compiled in Figure 5,
one for each source.

Figure 5: Distribution of editions.

We see here some definite choices (e.g. BNF, DNB, and BNCR), which give priority to a single edition. On the other hand, with the amount of different editions reported by Classify, those who have used the Dewey number for some time do not seem to have allowed for an update of Dewey notations in the catalog. Certainly, this can be ascribed to the complexity of the operation. Finally, we note the (still) low usage of edition 23. As previously stated, the catalog has been increased by 19,710 new Dewey classification numbers in just as many bibliographical records. The increase amounted to 47.8%, given that earlier records with tag 082 totaled 41,255. The current distribution of Dewey numbers, shown in figure 7 on page 191, outline a profile of the library’s holdings, reflecting the areas of interest in the University’s different schools as well as the library’s growth.
Figure 6: Distribution of library works as a function of Dewey classification subdivisions.

Figure 7 on the facing page represents the distribution of Dewey editions in the catalog. The absence of editions for a significant number of bibliographical records is a case of inhomogeneous cataloging. To address this situation, a method of retrieval very similar to the one outlined in the present work could be used.

9 The Dewey Index in the OPAC

Through scrolling indexes, shown in Figure 8 on page 192 and mentioned in note 21, it is possible in the OPAC to offer a path of semantic search based upon the Dewey classification number. The search counts performed by users show that the index of greatest use is precisely the Dewey decimal classification, even higher than the name index, which is of particular importance for references of ancient authors as well as popes.
10 Software used

The seven query programs were written in Perl language, making use of the Koha API and the following libraries: LWP for HTTP connections, ZOOM for Z39.50 connections, DBI for connections to the MySQL database, XML::XPath for XML data processing, WWW::Scraper for HTML data processing, and MARC::Record for the processing of MARC records.

11 Conclusions

The present work helped us to understand the value and problems of retrieving information online, which can contribute to the improvement of bibliographical catalogs. Generally, copy cataloging is considered important in obtaining the entire record, but — through

---

19 Each library is documented and available at [http://search.cpan.org](http://search.cpan.org).
Figure 8: Browsing Dewey in Koha.
unique identifiers such as the ISBN etc. — it is possible to find partial or “atomic” information by means of which several objectives may be achieved:

- improving the catalog from a static perspective, as in the case presented
- dynamically enhancing the OPAC by the retrieval of data while viewing a record
- increasing navigability for better utilization of the OPAC
- helping to clean up the catalog
- performing quality checks
- providing support tools for cataloging
- increasing the number of unique identifiers in the catalog
- comparing databases.

12 Appendix

12.1 Query elements for the selection of records without a Dewey classification number

The function ExtractValue,\(^{20}\) which is present in MySQL 5.1.5 or higher, allows querying of XML data, specifying the field to be examined and an XPath expression as parameters.\(^{21}\)

Table 9: Principle elements of the query for the selection of bibliographical records to be reviewed.

<table>
<thead>
<tr>
<th>biblionumber</th>
<th>The system number of the bibliographical record</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISBNlist</td>
<td>ExtractValue(marcxml,'//datafield[@tag=&quot;020&quot;]/subfield[@code=&quot;a&quot;]/a') This is a list of occurrences of subfield $a in tag 020, which are separated by a space; normally, an occurrence is unique.</td>
</tr>
<tr>
<td>isbn_present</td>
<td>ExtractValue(marcxml,'count(//datafield[@tag=&quot;020&quot;]/subfield[@code=&quot;a&quot;])/a') At least one occurrence of 020$a</td>
</tr>
<tr>
<td>dewey_absent</td>
<td>ExtractValue(marcxml,'count(//datafield[@tag=&quot;082&quot;]/subfield[@code=&quot;a&quot;])/a') No occurrence of 082$a</td>
</tr>
<tr>
<td>language_008</td>
<td>substr(ExtractValue(marcxml,'//controlfield[@tag=&quot;008&quot;]/a'),36,3) = 'language_code'</td>
</tr>
</tbody>
</table>

12.2 Parameters for web searches

To identify the parameters that compose the search URL, including the ISBN, it is possible to proceed in one of the following ways:

- run the query and note the response URL. If it does not contain parameters, i.e. in the event of a form where method="post", change its method to the value get through “Inspect Element” (contained in various browsers), by pressing right-click on the form, and then run the query;

- analyze the http request sent from the query by means of a plugin for traffic analysis, or a special feature in the browser.
12.3 Examples of some responses

An example of XML response obtained from Classify\textsuperscript{22} is the following:

\begin{verbatim}
Listing 3: XML

<?xml version="1.0" encoding="UTF-8"?>
<classify xmlns="http://classify.oclc.org">
  <response code="2"/>
  <!-- Classify is a product of OCLC Online Computer Library Center: http://classify.oclc.org -->
  <work author="Beaucamp, Evode" editions="5" format="Book"
    holdings="69" itemtype="itemtype-book" title="Israel en prière : des Psaumes au Notre Père">014271167</work>
  <orderBy>hold desc</orderBy>
  <input type="isbn">22/04022659</input>
  <start>0</start>
  <maxRecs>25</maxRecs>
  <editions>
    <edition author="Beaucamp, Evode" format="Book" holdings="40"
      itemtype="itemtype-book" language="fre"
      oclc="014271167" title="Israel en prière : des Psaumes au Notre Père">
      <classifications>
        <class edition="19" ind1="0" ind2="4" sf2="19"
          sfa="220.6" tag="082"/>
        <class ind1="0" ind2="4" sfa="BS680.P64" tag="050"/>
      </classifications>
    </edition>
    <edition author="Beaucamp, Evode" format="Book" holdings="21"
      itemtype="itemtype-book" language="fre"
      oclc="299394640" title="Israel en prière : des psaumes au Notre Père">
      ...
    </edition>
  </editions>
</classify>
\end{verbatim}

\textsuperscript{22}http://classify.oclc.org/classify2/Classify?summary=false&isbn=2204022659.
<classifications>
  <class ind1="1" ind2="4" sfa="200" tag="082"/>
  <class ind1="" ind2="4" sfa="BX2033B42 1985" tag="050"/>
</classifications>
<edition>
  <edition author="Beaucamp, Evode" format="Book" holdings="5"
    itemtype="itemtype-book" language="fre" oclc="246374613"
    title="Israel en prière : des psaumes au Notre Père"/>
</edition>
<recommendations>
  [ ... ]
</recommendations>
</classify>

An example of a Z39.50\textsuperscript{23} (MARC21) response in readable format:

**Listing 4: MARC21**

```
00932cam 2200253 a 4500
001 500315
005 20050929180451.0
008 851021s1986 nyua 000 0 eng
035 $9 (DLC) 850733338
010 $a 850733338
```

\textsuperscript{23}From Library of Congress, lx2.loc.gov:210/LCDB, find @attr 1=7 0874472466.

JLIS.it. Vol. 4, n. 2 (Luglio/July 2013). Art. #8766 p. 196
020 $a 0874472466 (pbk.) : $c $8.95
040 $a DLC $c DLC $d DLC
050 00 $a LB2353.57 $b A16 1986
082 00 $a 371.2/6 $2 19
245 00 $a 10 SATs : $b the actual and [...] prepare for it.
250 $a 2nd ed.
260 $a New York : $b College Entrance Examination Board : $b ...
300 $a 304 p. : $b ill. ; $c 28 cm.
[ ... ]

An example in HTML code:\(^{24}\)

Listing 5: HTML

```html
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN"
"http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html xmlns="http://www.w3.org/1999/xhtml" xml:lang="de" lang="de" dir="ltr">
<head>
<title>DNB, Katalog der Deutschen Nationalbibliothek</title>
[ ... ]
</head>

<body onload="doLoad()">
[ ... ]
<tr>
<td width="25%">
   <strong>Link zu diesem Datensatz</strong>
</td>
<td>http://d-nb.info/977758214</td>
</tr>

```
The browser version is shown in Figure 9.

<table>
<thead>
<tr>
<th>Link zu diesem Datensatz</th>
<th><a href="http://d-nb.info/977758214">http://d-nb.info/977758214</a></th>
<th>ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Titel</td>
<td>Fides caritate formata : das Verhältnis</td>
<td></td>
</tr>
<tr>
<td>Person(en)</td>
<td>Rose, Miriam</td>
<td></td>
</tr>
<tr>
<td>Verleger</td>
<td>Göttingen : Vandenhoeck &amp; Ruprecht</td>
<td></td>
</tr>
<tr>
<td>Erscheinungsjahr</td>
<td>2007</td>
<td></td>
</tr>
<tr>
<td>Umfang/Format</td>
<td>303 S. ; 24 cm</td>
<td></td>
</tr>
<tr>
<td>Gesamttitel</td>
<td>Forschungen zur systematischen und ö</td>
<td></td>
</tr>
<tr>
<td>Hochschulschrift</td>
<td>Zugl.: München, Univ., Diss., 2004/200</td>
<td></td>
</tr>
<tr>
<td>EAN</td>
<td>9783525563427</td>
<td></td>
</tr>
<tr>
<td>Sprache(n)</td>
<td>Deutsch (ger)</td>
<td></td>
</tr>
<tr>
<td>Schlagwörter</td>
<td>Thomas &lt;de Aquino&gt;: Summa theologi</td>
<td></td>
</tr>
<tr>
<td>DDC-Notation</td>
<td>231.6 [DDC22ger]</td>
<td></td>
</tr>
<tr>
<td>Sachgruppe(n)</td>
<td>230 Theologie, Christentum</td>
<td></td>
</tr>
<tr>
<td>Links</td>
<td>Inhaltsverzeichnis</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 9:** Result of a search by ISBN in the Deutsche Nationalbibliothek catalog.

ABSTRACT: The increasing availability of online catalogs and bibliographical databases allows not only for copy cataloging, but also for the retrieval of atomic information useful within the catalog. To this end, Dewey decimal numbers were imported from national and international sources by means of the unique identifier ISBN. Technical specifications have been developed to locate the records to be enhanced, to query external databases, to extract the Dewey decimal classification numbers and add them to the catalog. The exceptionally large amount of Dewey numbers added to the catalog has improved the semantic usability of the OPAC. The procedure established has also facilitated the collection of information on the use of the Dewey Decimal System in the various databases used and allowed to make certain comparisons between them. The tools employed can be used analogously for data-retrieval operations in the catalog, as an aid in the cataloging process, or to improve the OPAC in either a static or dynamic manner. Taking into account its virtually exclusive practical purpose, this work is characterized by practical rather than theoretical choices. However, the experience acquired opens up areas even in the field of academic research.