BIBLIOGRAPHY OF THE BULGARIAN REVIVAL (1801-1878) AND INFORMATION RETRIEVAL SYSTEM TECHNOLOGIES

V. Zagorov
University of Library Studies and Information Technologies (BULGARIA)

Abstract

The study presents the second stage of the project related to the development of the Data Base and Information Retrieval System (IRS) for the bibliography of the Bulgarian books from the National Revival period (1801–1878). After the observation of the existing printed bibliographies of this period, which differ in structure, quantity and content, the project team is embarking on the final steps to finishing the editing of the bibliography entries. The paper will focus on exact steps for constructing the interface and structure of the IRS and importing the Data Base information into the system. Some technical and historical aspects and details of this procedure will be presented. The purpose of the final product is to make a radical transformation in the methods whereby the Bulgarian national bibliography is presented and taught. The project team work is related to different topics such as Academic Research Studies and Library and Resource Management in the field of Bibliography and Book History. The project Digitalization of the Repertoire of Books from the Bulgarian Revival (1801–1878) is funded by The Bulgarian National Science Fund (BNSF).

Keywords: Bulgaria, Revival Books, Retrospective Bibliography, Information Retrieval System.

1 INTRODUCTION

The Enumerative Bibliography is one of the most important humankind activities. It is mapping the achievements of the human civilisation during the centuries helping for sustainability and continual growth in the scientific field. This is why in the past few decades in relations with the dramatic information landscape changes the need for the transformation in the main bibliography tools is needed. Today, bulky bibliographic reference books merge into large Databases (DB), and slow page scrolling is replaced by fast Information Retrieval Systems (IRS). This transformation is already done in many countries who provide their national bibliographies online [1]. In Bulgaria, despite the development of several huge library e-catalogues, there is no DB and IRS build on the principals of the Enumerative Bibliography [2]. In December 2018 University of Library Studies and Information Technologies (ULSIT) started second stage of a project with financial help of the Bulgarian National Science Fund in attempts to create a bibliographic analytical repertoire based on cloud technologies. The project under the title Digitalization of the repertoire of the Bulgarian Revival books (1801-1878) focuses on a small part of the Bulgarian book production, published from the beginning of the 19th century until the Liberation of Bulgaria in 1878. As the main objectives of the project are already presented in the scientific space, this paper will focus on the development of the Database and IRS itself [3].

2 METHODOLOGY

The information about the Bulgarian Revival books (1801-1878) is collected and organized on the principles of the enumerative and descriptive bibliography synchronized with the principles and architecture of computer databases. Till now the DB is containing 2016 book entries but the final revision will be accomplished in the end of July 2019. The bibliographical information is collected from the previous Bulgarian printed bibliographies [4] and is compared with several huge e-catalogues such as WordCat, BnF Catalogue général, Bulgarian National Academic Library and Information System (NALIS), BG.Cobiss and etc. At the same time, in order to avoid mistakes, the project team also worked directly with copies of the books in the Bulgarian National Library of St. St. Cyril and Methodius, the Central Library of the Bulgarian Academy of Sciences, the Book collection of the Municipality Historical Museum in Samokov. Depositors of Bulgarian printed books in different libraries as Library of Congress (USA), Russian State Library in Moscow and Library of Russian Academy of Science in St. Petersburg (Russia), National Széchényi Library and Eötvös Loránd University Library (Hungary), Matica srpska (Serbia) and others have been viewed directly or indirectly. The bibliographical entries were stored in a

1 They will be an option to add new entries in the DB if an unknown title or edition appears.
Database in Excel format in Google Drive virtual space. In this conditions different participants of the team can work simultaneously under the control of the supervisor. Every single entry contains two types of elements – first group is related with the Bulgarian and International bibliographical standards [5] and the second group is related with the improvement of the searchability and the informational value of the records according to specific scientific or every day user’s needs and interests. Each bibliographic item in the standard bibliographic description is arranged in a stand-alone column that is linked to a numeric code to serve for filtering in the search. The main columns contain information about: number of the entry, main title, parallel title, authorship, information about edition, publisher, printer, place of publication, year of publication, physical measurements of the book, price of the book, presence or absence of digital copy. For more sustainability of the DB links between the reference numbers of the older bibliographies and nowadays e-catalogues are build [4].

Standard bibliographical information is served by columns with machine readable value. For example, the geographical dispersion of Bulgarian print production in the years of the Ottoman rule around the Europe and the multilingualism of the Ottoman Empire presupposes the presence in the Bulgarian books of many foreign language elements. They are parallel titles and notes related to censorship and administrative systems mainly in German, Russian and Ottoman languages. That’s why next to the main column “Parallel Title” in foreign language there is a column containing digit codes: 0 (Bulgarian); 1 (German); 2 (Greek); 3 (Russian); 4 (Ottoman); 5 (Latin). Codes are given according to the information in the books themselves when filling in the Database. Such arrangement of the information will help the user, especially non Bulgarian speaking one, to search in the database.

Table 1. Encoding the bibliographic entry with parallel title [6].

<table>
<thead>
<tr>
<th>Main title in Bulgarian</th>
<th>Code</th>
<th>Parallel title in German</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Свещеная история церковна от Ветхиет и Новиет завет</td>
<td>0</td>
<td>Kurze heiligen Kirchengeschichte des alten und neuen Testaments</td>
<td>1</td>
</tr>
</tbody>
</table>

The same is the situation with the alphabets used in the Bulgarian books in this period. They are different variation of the Cyrillic alphabet, also Greek, Latin and Arabic alphabet, used separately or simultaneously in the texts, sometimes without any logic. In this case separate columns with code 0/1 for the presence or absence of a specific alphabet in the book is created. For example, if the book is Greek-Bulgarian Phrasebook with permission from the Ottoman authorities it will contain Cyrillic, Greek and Arabic letters. The construction of the Database is making all these elements visible.

Because of the language issues related to the lack of literary Bulgarian language another element to every single entry is added – the picture of the title page. In this case every user will have opportunity to check and to verify the authenticity of the bibliographic record. In case of an error in the DB the user will have opportunity to report information about the incorrect information. In this way an absolute clearing of the bibliographic errors will be achieved. In addition to the language issue there are a lot of peculiarities in the book publishing practices which requires the user to have an access to at least 3 digital copies of every book described in the DB. For example, during the work on the DB entries the team understood that sometimes same editions have different color and different illustrations of the cover, and the bibliographical information is scattered along the front and back pages and covers. To solve the problem with possible misleading of the user along with digital copies the field Note is added. The user will be able to open it with more/less option under the ISBD. The lack of digital copies for certain books can help Bulgarian and foreign libraries to set their digitalization plan and to avoid over-digitalization.

Before the transfer of the bibliographic entry to the IRS container each record receives a number ensuring its authenticity as well as information on the physical specimens and repositories on which it is screened. It plays the role of auto-control and aims to avoid gags and gaps in the final product.

The Information Retrieval System for the project is built on the base of the Elasticsearch 7.0. Elasticsearch is a highly scalable open-source full-text search and analytics engine. It allows you to store, search, and analyze big volumes of data quickly and in near real time. It is generally used as the underlying engine/technology that powers applications that have complex search features and requirements” [7]. This system is of the NoSql type that stores the entries in the JavaScript Object Notation (JSON) and it is based on Apache Lucene. Currently, Elasticsearch is one of the most popular full-text search systems that employs complex and diverse filters. All this makes it extremely useful for the project. The choice of such a system will allow a quick and easy search in the linguistically unstable
textual environment created by the Bulgarian authors, printers and publishers in the 19th century and especially till the Liberation of Bulgaria from the Ottoman Empire (1878).

For the construction of the user interface and administrative panel of the IRS HTML 5, CSS, JavaScript, React, PHP and Go will be used. The searching engine will be integrated to the existing project webpage [8]. The primary goal of the search engine architecture is to have fewer clicks from the home screen to the relevant information for the user and to be user friendly for everybody (a library specialist or a regular user). The users will have an option for Simple and Advanced search. During the Simple Search the user will make his information request by using only one field for typing. The IRS will respond to him with results from certain elements of every single bibliographic entry as: main title and subtitle, parallel title in foreign language (searchable in the language of the parallel title), authorship (searchable also in Latin transcription of the name, place of publication, publishing and printing house, year of publication, subject). After displaying the results from the simple search, it will be possible for the user to make them more detailed by using additional filters and sorting the results already obtained. For example, if the user is searching by topic Geography he will have 113 answers collected from three main columns in the DB: Main Title and Subtitle and Subject. After choosing to filter results only form the column Subject the user will obtain the exact number of the geographical books printed during the period – 25 books, even if the title did not contain word Geography but the old Bulgarian equivalent – Zemleopisanie. The results can be sorted also by chronological principals – ascending, descending and within a certain range. The set of filters is consistent with the characteristics of the Bulgarian printed book from the period and with the long experience and good examples gained in the library e-cataloging. By using the Advance search option, the user will be able to specifically define search policies form that encompasses each detail from every single bibliographic entry. For every field the matching methods: "exact phrase", "to contain the phrase/word", "to begin with" can be chosen. After the final choice of criteria for searching the user can choose the type of logical relation between elements of the search: "and", "or", "nor". There will be Geographical Information System advanced option for searching which will present bibliographical resources to the user via map.

Options for users who do not speak Bulgarian will not be so diverse but they will still have the opportunity to use the system. The site itself and the search engine will have an English-language version. However, the user will be able to search only in a limited number of fields and through a limited number of filters. This will be possible for the foreign language elements contained in the books themselves and the used alphabetical systems. The only added option that is quite operational for the foreign users is the transliterated Latin name of the author. The names are transliterated according to the international standards [9].

<table>
<thead>
<tr>
<th>Name in BG</th>
<th>Year of birth</th>
<th>Equal</th>
<th>Year of birth</th>
<th>Transliteration Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Йоанович, хаджи Найден</td>
<td>1805?-1859</td>
<td>=</td>
<td>1805?-1859</td>
<td>Joanovič, hadži Najden</td>
</tr>
</tbody>
</table>

3 RESULTS

In my paper The Bulgarian 19th-Century Book as a Crossroad of Professional Literary Practices it was mentioned how unsystematic were the publishers’ practices and that the Bulgarian book at the Revival period (1801–1878) was a difficult object for bibliographic description. is [10]. Despite the characteristics of the book production the final revision of bibliographic entries will finish by schedule in the end of July. The web page of the project is fully operational and there is a lot of interest from the audience in convertor for recalculation of old Cyrillic letters dates in Arabic numerals [11]. In the end of September 2019 the Elasticsearch system have to be fully operational. In period September 2019 – January 2020 test groups, including students, for functionality of the IRS are planned2. It was already found out during the implementation of other projects of the team as the Erasmus+ project NAVIGATE – Information Literacy: A Game-based Learning Approach for Avoiding Fake Content that in terms of using information for learning the previous experience of the students provides the scaffolding allowing them to increase their existing knowledge [13]. The possible use of the Elasticsearch system in the teaching process of students from the Library Studies department corresponds to the increasing interest in ULSIT in the

2 USLIT have a good experience in organising groups for observing and testing the opinion of his students and collegial networks [12].
recent years towards the application of new methods for student preparation, methods that are designed
to make learners active participants in the educational process [14].

Till the end of May 2020, after corrections mentioned from the test group, the final product should be
accomplished. After the finalization of the project libraries and museums can compare their catalogues
with the DB and download for free bibliographic entries according to their needs.

Fig. 1: Old Church Slavonic Letter Date Convertor.

4 CONCLUSIONS

The successful finalization of the project will be a new stage in the development of the Bulgarian National
bibliography. The new IRS will make usage of bibliographic data easier for scientists in the field of
Bulgarian Literature, Librarianship and History. It will not only replace the work of voluminous and
difficult-to-use printed bibliographies, but will also allow the revision of old scientific notions and will be
an easy tool for selecting important topics for future generations of Bulgarian scientists. The final product
will also help foreign scholars to satisfy their interest in topics related to the Bulgarian printed book until
the Liberation of Bulgaria (1878). The product will have a practical value for all libraries in Bulgaria and
abroad, which have collections of Bulgarian books from the period 1801–1878.

ACKNOWLEDGEMENTS

The project 05/2 is being implemented with the financial support of The Bulgarian National Science Fund
at the Ministry of Education and Science, Republic of Bulgaria.

REFERENCES

(Състав. Васил Загоров, Алберт Бенбасат), с. 207–274, София : За буквите – О писменехь,
2016.
978-954-9837-22-3.


