

ONLINE BIBLIOGRAPHY OF MACEDONIAN LANGUAGE¹

Katerina Zdravkova, Milica Mirkulovska

A b s t r a c t: Online bibliography of Macedonian language <http://bmj.manu.edu.mk> is a system consisting of two interconnected applications, a desktop application for automatic batch uploading of the volumes and their representation as bibliographic records, and an ASP.NET client-server application based on C# for the server code. It is based on recently published volumes covering the period between 1953 and 1985. This paper presents the implementation and the functioning of online bibliographic system. In the beginning, the steps leading to its creation are presented in details. Afterwards, all the functionalities of the system are presented, together with illustrations of their active operation. Then, the administrative part of the system is presented. The paper ends with the conclusions, and direction for further development of the system.

Key words: automatic extraction of bibliographical informations; searching and editing textual databases

1. INTRODUCTION

Creation of bibliography of Macedonian language initially started in 1953 [1]. The project was temporarily stopped in 1986, to be again re-established in 1993. New project is still in progress, and it is conducted by the Centre for Areal Linguistics within Macedonian Academy of Sciences and Arts. As part of this project, two new volumes have been prepared so far [2, 3].

¹ The present article is a long english version of the article „Bibliografia Języka Macedońskiego Online: Doświadczenia i osiągnięcia (Przyczynek do dyskusji nad systemem informacji slawistycznej Isybisław)“. In: *Zagadnienia informacji naukowej*, **2**, 90 (2007) 41–53.

As a result of their recent publishing, they are also available in electronic form. This was convenient for the creation of a Web-based online bibliographic system <http://bmj.manu.edu.mk>. This system was made at the Institute of Informatics from the Faculty of Natural Sciences and Mathematics in 2007.

Any bibliographic record in the recent two volumes conform the standards given in the International Standard Bibliographic Description (ISBD) with minor adaptations according to the standards from 1996 [4]. They include different metadata (title, classification, author, editor, publisher etc.) thus enabling access to bibliographic records with the same value of corresponding metadata.

The paper starts with a description of the bibliography of Macedonian language. It is followed by a review of related work, with particular attention to Linguistic Bibliography / Bibliographie Linguistique (BL), published by the Permanent International Committee of Linguists under the auspices of the International Council for Philosophy and Humanistic Studies² [5], which was used as a sample to create our own system. Next section presents the structure of bibliographic records in the published volumes, and their representation in the database. The paper continues with a detailed description of the development of online system. Section 5 presents the phases of the system. It also covers batch uploading of the volumes, and the extraction of bibliographic constituents out of textual bibliographic strings. Section 6 presents the implementation of client-server part of the system. Section 7 is dedicated to the administrative part of the system, which enables the editing within the database. Illustration of both parts of the system is presented in Section 8. The paper ends with the conclusions, and the direction for further development of the system.

2. CREATION OF THE BIBLIOGRAPHY OF MACEDONIAN LANGUAGE

In 1953 Bozhidar Vidoeski, assistant professor at the time at the Department of Macedonian Language within the Faculty of Philosophy (later on the department became part of the Faculty of Philology), published his interpretative bibliography of publications “A Contribution to the Bibliography of the Macedonian Language” (BML 1). This work included publications up till 1952 inclusive, which refer to the structure of the Macedonian lexical fund. He continued collecting the bibliographical data that refer to Macedonian Studies and

² Library of Congress Catalogue card Nr.A50-397 2rev. Subvention UNESCO since 1983.

arranged them in a card file, known as “Card file of the Macedonian Bibliography”. Since 1973 Milena Milev, manager of the Macedonian Academy of Science and Arts’ library at the time, assists Vidoeski. Since 1983 the academic Zuzanna Topoljinska participates in the ordering and simpleton of the “Card file”. In that period, thanks to Vidoeski’s initiative, the first version of a legitimate classification has been made. Unfortunately, due to personnel issues the project of the “Card file” was terminated in 1986.

The following period of reanimation of the meticulously collected “Card file of the Macedonian Bibliography” started with the formation of RCAL (Research Center of Areal Linguistics) within MASA (the Macedonian Academy of Science and Arts) and the idea of completing and publishing the materials from the “Card file”, formerly set with Vidoeski’s “A Contribution to the Bibliography of the Macedonian Language” in 1953, was accomplished. We should emphasize that since 1953 a complete bibliography of the Macedonian language has not been published. We have at our disposal partial bibliographies of several Macedonian scientific journals from specific periods, as well as more or less complete segments dedicated to the Macedonian language in the bibliographical extracts found in the “South-Slavonic philologist” journal; in the “Bibliography of Slavic Linguistics” (BSL) published in Poland (up till 1993 it covered the period from 1908 till 1981 in the “Rocznik Slawistyczny” journal, and in 1995 it covered the period from 1992-1997, as a special edition of the Institute for Slavistics at the Polish Academy of Science); in the Linguistic bibliography/Bibliographie Linguistique (BL) published in Brussels, Hag, Leaden (a project, financed by UNESCO since 1947, whose data base covering the period from 1993 till 2002 can be found on the internet since 2001 <http://www.blonline.nl>).

In 2002 a second volume of “A Contribution to the Bibliography of the Macedonian Language 2” (1954-1970) was published by RCAL within MASA. The edition provided simpleton to the following years: 1950, 1951, 1952, 1953) (BML 2), based on of the old “Card file”. The entire data are verified and the material is ordered according to the accepted principles in the contemporary bibliographical processing (a product of the team work by the academics: Topolnjska, PhD. Mirkulovska and Sonja Milenkovska, qualified Slavist, a volunteer in RCAL at the time). The volume contains 1521 bibliographical positions (the reviews are displayed under the registry position due to economical reasons) + an index of authors, editors, translators + an index of names included in the title of the bibliographical position. The bibliographical material

is presented with an original orthography (Cyrillic alphabet and Roman alphabet) of the language in which the bibliographical unit was written.

The material is ordered according to the specific year and a further classification³ is provided within each year. Later, in 2006 the third edition of the “Enclosure to the Bibliography of the Macedonian Language 3 (1971–1985)” (BML 3) was published (with suppletion for the previous year). It came along with certain amendments to the classification, that were required in the same material from the “Card file of the Macedonian Bibliography”. Right after the second volume was published, the writing of the third one followed. Due to the ample character of the material the team for processing the data included more members. The project was under the management of academic Zuzanna Topolinjska (editor-in-chief) and PhD. Milica Mirkulovska. Aleksandra Stojanovich, qualified Slavist and Anglicist, volunteer in RCAL during the period from June till September 2003 participated in the process of internal ordering and entering data from the “Card file of the Macedonian Bibliography” for the period from 1971 till 1985. Afterwards came the stage of checking and amending the data from the “Card file”. In this stage Aneta Buchevska took part as a qualified Slavist, volunteer in RCAL during the period from September 2003 till July 2004, inclusively with the completion of the volume. The material required elaboration of the classification⁴. In this stage of the process as a con-

³ CLASSIFICATION BML 2.

1. General, 1.1. Chronicle (jubilees, anniversaries, congress reports ect.), 1.2. Bibliographies, 1.3. Periodicals, 1.4. Collections (from conferences, symposiums, discussions etc.), 1.5. Collections of articles (in honor to...).

2. Contemporary standard language, 2.1. General (grammar, textbooks, teaching aids), 2.1.1. Mother tongue training, 2.2. Codification, standardization, language culture, 2.3. Phonology and phonetics, prosody, 2.4. Grammar, 2.5. Lexicology, lexicography (dictionaries, phraseology, etymology, terminology), 2.6. Stylistics, poetics, versification, 2.7. Authors' and literary texts' language, 2.8. Folk poetry's language, 2.9. Typological and confrontational research.

3. History, 3.1. General, 3.2. Phonology and phonetics, prosody, 3.3. Grammar, 3.4. Lexicology, lexicography (dictionaries, phraseology, etymology, terminology), 3.5. Authors' language in certain texts, 3.5.1. Texts (critical editions of texts).

4. Dialectology, 4.1. General (linguistic geography), 4.2. Phonology and phonetics, prosody, 4.3. Grammar, 4.4. Lexicology, lexicography (dictionaries, phraseology, etymology, terminology), 4.5. Studies about certain dialects, 4.6. Dialect texts, 4.7. Social dialects,

5. Onomastics, 5.1. General, 5.2. Toponomastics.

⁴ CLASSIFICATION BML 3.

1. General, 1.1. Chronicle (jubilees, anniversaries, congress reports ect.), 1.2. Bibliographies, 1.3. Periodicals, 1.4. Collections (from conferences, symposiums, discussions etc.), 1.5. Collections of articles (in honor to...),

sultant in the section **3. History**, Ph.D. Liljana Makarijoska, took part as a senior research fellow (at the time) in the History department within the Institute for Macedonian language “Krste Misirkov” in Skopje. At the end, the volume contained 3614 bibliographical positions (without the reviews given under the registry bibliographical unit). Same as in the second volume, the third volume included several slavistic and balcanistic articles that were part of the “Card file”. These articles are of great significance because they depict the Macedonian conditions. The two volumes BML 2 and BML 3 together contain approximately 8 000 bibliographical descriptions and over 600 pages written text. The documents are arranged chronologically according to their year of publishing and classified into 28 categories and 12 sub-categories.

We can conclude that we have successfully updated Vidoeski’s long-term work programme. With our efforts and the devotion of the manager of RCAL within MASA, academic Zuzanna Topolinjska, by the end of 2008 the forth volume of the “Enclosure (BML 4) for the period from 1986 till 2006” should be published. For this volume the Government of The Republic of Macedonia provided a grant within the “2008- Year of the Macedonian Language” manifestation.

3. RELATED WORK

There are many bibliographic systems available online. They cover many different areas, including linguistics. As previously mentioned, the majority of Macedonian linguistic publications have already been included into international linguistic bibliographic database in Hague, which is available online

2. Contemporary standard language, 2.1 General (grammar, textbooks, teaching aids), 2.1.1. Mother tongue training, 2.2. Codification, standardization, language culture, 2.3. Phonology and phonetics, prosody, 2.4. Grammar, 2.5. Lexicology, lexicography (dictionaries, phraseology, etymology, terminology), 2.6. Stylistics, poetics, versification, 2.7. Authors’ and literary texts’ language, 2.8. Folk poetry’s language, 2.9. Typological and confrontational research, 2.10. Sociolinguistics, 2.11. Translation,

3. History, 3.1. Macedonian language from the 19th century, 3.1.1. General, 3.1.2. Phonology and phonetics, prosody, 3.1.3. Grammar, 3.1.4. Lexicology, lexicography (dictionaries, phraseology, etymology, terminology), 3.5. Authors’ language in certain texts, 3.5.1. Texts (critical editions of texts),

4. Dialectology, 4.1. General (linguistic geography), 4.2. Phonology and phonetics, prosody, 4.3. Grammar, 4.4. Lexicology, lexicography (dictionaries, phraseology, etymology, terminology), 4.5. Studies about certain dialects, 4.6. Dialect texts, 4.7. Social dialects,

5. Onomastics, 5.1. General, 5.2. Toponomastics, 5.3. Anthroponomastics.

on www.blonline.nl [6]. Furthermore, the system has a very user-friendly interactive client-server application, with a search engine capable of effective simple (upper part) and advanced search (lower part).

Consequently, the basis for the creation of Macedonian online bibliographic system was namely this system. Our main intention was to create a system that resembles to it, and to enable all the functionalities it covers. Specific goal of our system was the capability to cope with different encodings in parallel. Namely, in BL Online, all the publications dealing with the Macedonian are written with the Latin script. In order to present the publications written in the Cyrillic script, this system uses the transliterated characters, which sometimes hinders the search, or gives ambiguous results (Fig 1.).

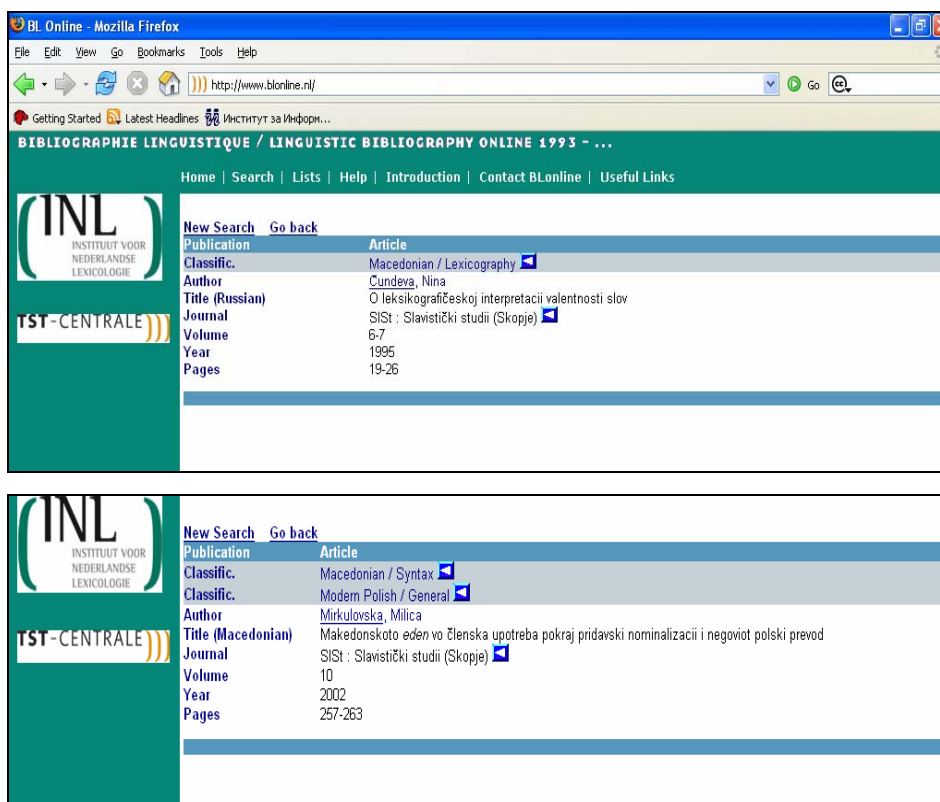


Fig. 1. Both, the Russian and the Macedonian Cyrillic are presented with the Latin script

Another problem of the Dutch system was that it sometimes incorrectly interprets specific characters. For example, the title of the publication by Spasov and Feleszko (Fig. 2) contains a text in Polish and in Macedonian. Macedonian part is transliterated into: “*se kačiv po skalite → gi kačiv skalite*”. However, the real transliteration of the verb “*качув*” is not “*kačiv*”, but “*kačiv*”. The problem of obstructed search in the Dutch system is partly resolved by the suggestion to avoid diacritics. In our system, we decided to enable the usage of several encodings in parallel.

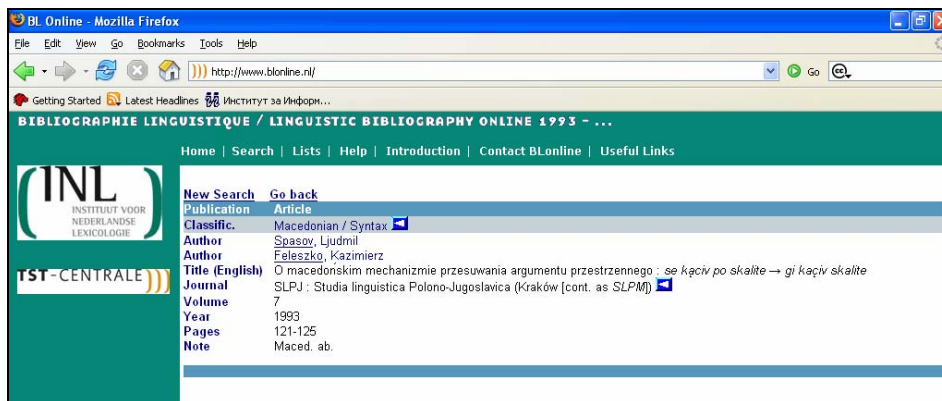


Fig. 2. Mistaken transliterations can seriously obstruct the search

In order to enable automatic extraction of the information out of bibliographic records, fruitful was the experience from Cornell University (Bergmark, 2000). The paper presents several algorithms for extracting metadata from online texts and linking full-text documents together. Although the system Bergmark presents is mainly related to online journals, the approach implemented was found useful for online linguistic bibliography. AutoBib system, created at Duke University (Geng, 2004) was an additional support for this ambitious task. This system defines algorithms for identifying segments within bibliographic Web sites, and their conversion into structured bibliographic records using a parser based on hidden Markov model.

Both approaches are hopefully compatible with NET Framework Software Development Kit (SDK), which contains the **System.Text.RegularExpressions** Namespace class (STRE). This regular expression namespace enabled smooth conversion of textual files into database of bibliographic records.

4. STRUCTURE OF BIBLIOGRAPHIC RECORDS

The data used in the system Online Bibliography is structured, that is it has defined structure according to the International standard for bibliographical description from the International Federation of Library Associations (IFLA) with certain modifications imposed by the material. The editors tried to coordinate mostly with the standards accepted by BL (Linguistic Bibliography / Bibliographie Linguistique).

Bibliographical data of one publication includes information about the authors, the title, the publisher, the year and place of publishing, the source, the number of pages, supplements etc. Each publication is connected with a specific field, characters, reviews/ discussions. The data about the reviews, discussions and answers to the discussions are signaled near the bibliographical unit to which they refer to, along the flat line (|). Further classifications in a certain bibliographical units are signaled with the appropriate number of angle brackets <....>.

This sort of data structure represents bibliographical unit (Fig. 3a).

2.5. Лексикологија, лексикографија (речници, фразеологија, етимологија, терминологија)	1104 Feleszko, Kazimierz: Zdanie bezosobowe w języku macedońskim. – <i>SFPS</i> 7, 1967, 205–217.
кирилица	1105 Panzer, Baldur: <i>Der slavische Konditional. Form. Gebrauch. Funktion.</i> – München : Wilhelm Fink Verlag, 1967. – 317с.
1106 Китановски, Дано: Француско-македонски речник (со краток преглед на француските гласови и на глаголските форми). – Скопје : Просветно дело, 1967. – 548с.	
1107 Миличкиќ, Владимир: <i>Обратен речник на македонскиот јазик</i> / ред.: Кирил Конески, Александар Цукески. – Скопје : Институт за македонски јазик, 1967. – 387с. <i>MJ</i> 18, 1967, 165–168 Оливера Јашар-Настева <i>НМ</i> , 2.05.1968 Оливера Јашар-Настева <i>СовСлав</i> 6, 1968, 108–109 Р. Цейтлин	
1108 Ройзензон, И. Л.: Странички из славјанской лингвостатистики. – <i>MJ</i> 18, 1967, 107–109.	
1109 Руси, Љутви; Корвезироски, Миле: <i>Македонско-албански речник.</i> – Скопје : Просветно дело, 1967. – 712с. <i>НМ</i> , 6.11.1968 Ристо Елимов	
1110 Сејфула, Кевсер: <i>Македонско-турски речник.</i> – Скопје : Просветно дело, 1967. – 657с. <i>НМ</i> , 6.11.1968 Ристо Елимов	

Fig. 3a. Extract from the “Bibliography of the Macedonian Language 2” where the bibliographical unit 1107 is found, colored in blue

The data from this bibliographical unit (1107) have the following meaning (Fig. 3b):

Тип на издание : книга

Број	1107
БМЈ/ том	2
Наслов	Обратен речник на македонскиот јазик
Година на издавање	1967
Место на издавање	Скопје
Издавач	Институт за македонски јазик
Страници	387
Класификација	2.5 Современ стандарден јазик/Лексикологија, лексикографија (речници, фразеологија, етимологија, терминологија)
Автори	Владимир Миличиќ
Уредници	Александар Цукески Кирил Конески
Прикажувачи/	Оливера Јашар-Настева
Автори на приказ	Р. Цейтлин
Дискусии, Прикази	автор: Оливера Јашар-Настева извор: МЈ 18 година: 1967 страници: 165–168
	автор: Оливера Јашар-Настева извор: НМ година: 1968 страници:
	автор: Р. Цейтлин извор: СовСлав 6 година: 1968 страници: 108–109

Fig. 3b. Data from the bibliographical report 1107

5. CREATION OF THE ONLINE BIBLIOGRAPHIC SYSTEM

The process of creating online bibliography went through these stages:

- a) Conversion of previously used encoding into Unicode UTF-8 encoding
- b) Batch insertion of printed volumes into a textual database
- c) Transformation of textual records into a database of bibliographical records
- d) Implementation of search algorithms for simple and advanced search of full and partial phrases
- e) Presentation of the entities related to a particular bibliographical entry
- f) Addition of editing options and to correct accidental mistakes inherited from published volumes, or mistakes originating from manual addition of new records.

Explanation of the stages is presented in details in the following six subchapters.

5.1. Unification of the encodings

Conversion of different fonts used to present bibliographical records with one encoding was manually done using the commercial system LC KonvertMak V3.0, Login Systems [6]. This system enables conversion of TrueType fonts into encoding with language support. This approach was found the most convenient, because it includes all the languages used in the printed bibliography of Macedonian language [6]. Conversion is bijective, and it exchanges one character at a time (Fig. 4a, and Fig. 4b).

As a result, all the characters were converted into Unicode UTF-8 encoding. Conversion produced few mistakes (Fig 5), which were noticed during the evaluation of the system, and manually corrected in the database.

5.2. Batch insertion of the volumes

Converted textual documents were automatically uploaded in the system using self-developed desktop application. The application consists of two smaller units. The first unit enables inserting of all the periodic editions, the list of areas, and the list of all persons appearing as authors, editors, and reviewer. All these lists exist in the volumes, and their amount is limited so their addition was straightforward.

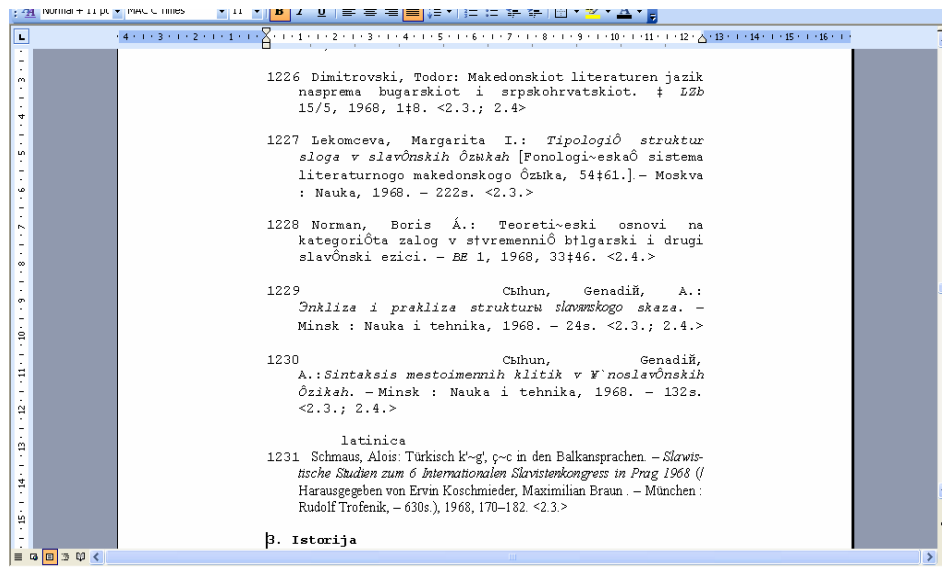


Fig. 4a. Part of the document before conversion, written with TrueType fonts

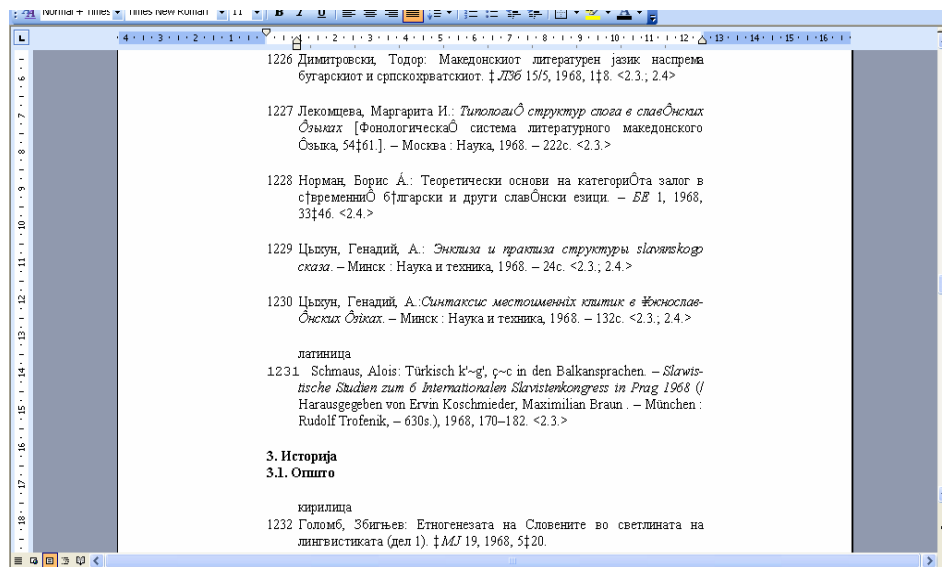


Fig. 4b. The same part of the document after conversion into Unicode UTF-8

broj	ime	BMJ
3.5.1.	Историја/Јазикот на писатели на одделни текстови/Текстови (крити...	2
4.	Дијалектологија	2
4.	Дијалектологија	3
4.1.	Дијалектологија/Општо (лингвистичка географија)	2
4.1.	Дијалектологија/Општо (лингвистичка географија)	3
4.2.	Дијалектологија/Фонологија и фонетика; прозодија	2
4.2.	Дијалектологија/Фонологија и фонетика; прозодија	3

Fig. 5. The converter added squares in item 4.1 from volume 2, and in 4.2 from volume 3

The second unit is divided into three classes:

- WordDocument – a class containing the attributes and the methods for handling documents written in Microsoft Word (Fig. 6.);
- BibRecord – a class containing the attributes and the methods for extraction of bibliographic record into separate paragraphs (Fig. 7.)
- Parser – a class containing the attributes and the methods responsible to compose and write database entities (Fig. 8.)

```

public class WordDocument
{
    private string name;
    private Microsoft.Office.Interop.Word.Document document;
    private Microsoft.Office.Interop.Word.Paragraphs paragraphs;
    Microsoft.Office.Interop.Word.ApplicationClass app;
    public WordDocument(string name);
    public void Close();
    public Microsoft.Office.Interop.Word.Document getDocument()
    {
        return document;
    }
    public Microsoft.Office.Interop.Word.Paragraphs getParagraphs();
    {
        return paragraphs;
    }
    public string getName();
    public void setName(string i);
}

```

Fig. 6. The code of the class WordDocument

```

public class BibRecord
{
    public string index = "";
    public string[] author;
    public string[] responsible;
    public string[] discussion;
    public string[] area;
    public string persons = "";
    public string title = "";
    public short tyoe = 0;
    public string source = "";
    public string place = "";
    public string publisher = "";
    public int year;
    public string year_number = "";
    public string pages = "";
    public string details = "";
    public string subject = "";
    public string text = "";
    public string journal;
    public short volume ;

    public BibRecord(string index, string[]author, string[] discussion,
string[] area, string persons, string title, short type, string
source, string place, string publisher, int year, string year_number,
string pages, string details, string subject, short volume);

    public BibRecord(Microsoft.Office.Interop.Word.Paragraph p, string g,
string o, short vol);

    public void generate(Microsoft.Office.Interop.Word.Paragraph p);

    public string toStr();
}

```

Fig. 7. The code of the class BibRecord

```

public class Parser
{
    private WordDocument doc;
    string year;
    string areas;
    bool valid;
    Microsoft.Office.Interop.Word.Paragraphs paragraphs;
    BiblioDataSet1 biblioBDataSet1;
    System.Windows.Forms.BindingSource bindingSource1, bindingSource2,
bindingSource3, bindingSource4, bindingSource5, bindingSource6, bindingSource7,
bindingSource8;
    DIPL1.BiblioDataSet1TableAdapters.RELATIONTableAdapter
relationTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.PERSONSTableAdapter personsTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.BibRecordTableAdapter
bibRecordTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.PERIODICTableAdapter
periodicTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.DISCUSSIONSTableAdapter
discussionsTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.AREASTableAdapter areasTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.CLASSIFICATIONTableAdapter
classificationTableAdapter;
    DIPL1.BiblioDataSet1TableAdapters.AliasTableAdapter aliasTableAdapter;

    public Parser(string p);
    public void enterAreas(short bmj);
    public void enterRelations(short role, short vol);
    public void enterPeriodic();
    private void enterBibRecord(BibRecord[] bib);
    public BibRecord[] parse(short vol);
}

```

```
}
```

Fig. 8. The code of the class Parser

In order to avoid the bottlenecks due to huge amount of text, each publishing year was separately uploaded. As a result, a textual database consisting of bibliographical records was obtained (Fig. 9).

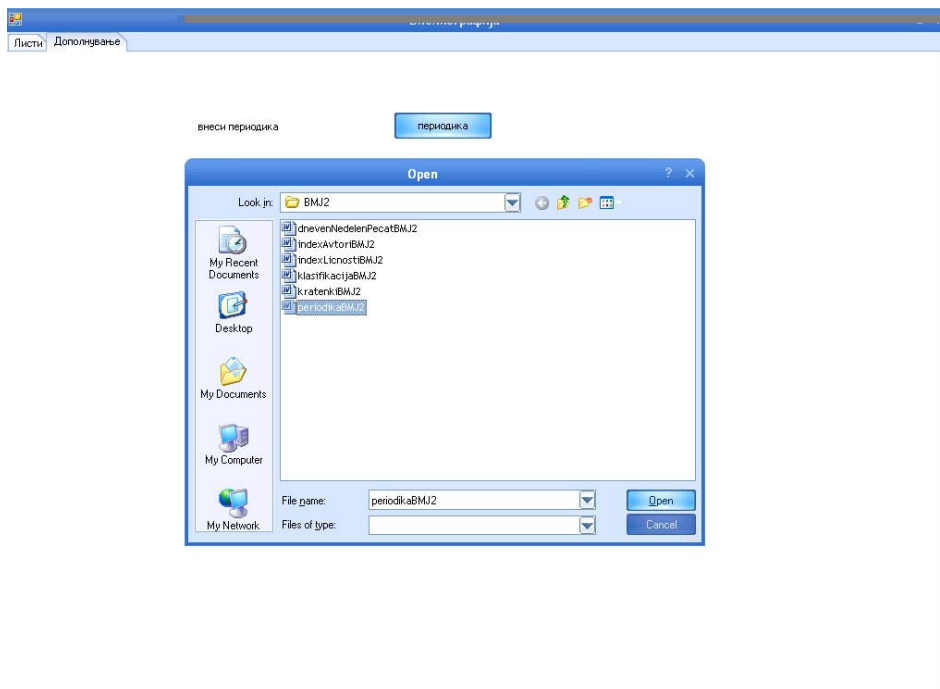


Fig. 9. Module for uploading of the database from textual documents

Transformation of flat textual records into bibliographical records was a process done by implementation of a parsing algorithm based on regular expressions. Namely, each bibliographic record can be represented with a finite-state automaton, consisting of bibliographic record, Unicode characters, initial state, rewriting rules, and the set of final states. Finite-state representation enabled the implementation of **System.Text.RegularExpressions** Namespace class existing in .NET Framework SDK.

As a result of this process, the table of all the publications is obtained (Fig. 10). Main table is the table of bibliographic records. It consists of records which include the bibliography index, the title as well as the type of the publication, the number of the volume, the place of publication, the publisher, the page(s), details, the source, the edition, and finally the short description. Bibliography index and the volume are primary keys, while the source and the edition are foreign keys from BibRecord. The database also contains records for the persons involved in the publication of the record, whenever necessary, their aliases, a record of periodic publications, discussions, areas, and finally a record for all the classifications. Such structure of the database is natural, and furthermore it is simple enough to enable high scalability.

индекс	наслов	тип	ВМЈ	место	izdavac	stranici	godBr
1119	Замечанья о не...	0	2			40-45	18
1119	Frekwencija i dist...	0	3			1976, 105-118	
112	За граматиката...	0	2			188-189	4/8
112	Билтен на одбо...	1	3	Скопје	МАНУ	27	1/1
1120	Заселване на б...	1	2	София	БАН	350	
1120	Општествено-по...	0	3			1976, 67-77	
1121	К. вопрос о мак...	0	2			5-36	12
1121	Од лексиката н...	0	3			53-56	23/2
1122	Историја на мак...	1	2	Скопје	Култсра	241	
1122	Поглед врз лек...	0	3			15-17	21/9-10
1123	Акцентските ан...	1	2	Скопје	(машинпис)		
1123	Дијалектна б...	0	3			7-16	27
1124	An sjet di sagac...	0	2			5-10	2/2
1124	Пресоодавање ...	0	3			30-32	21/1
1125	За един труд вь...	0	2			380-381	17/4
1125	Дијалектизмите...	0	3			1976, 119-123	
1126	Интернационал...	0	2			80-102	18
1126	Новооткриен ап...	0	3			207-214	27
1127	Кон посветата в...	0	2			16-18	14/6
1127	Григор С. Прли...	0	3			237-241	26/1
1128	За првиот прев...	0	2			47-68	18
1128	Граноти на мак...	0	3			1	
1129	Нови материјал...	0	2			353-391	19
1129	Обиди за вовед...	0	3			1976, 57-65	
113	За зборот осито...	0	2			65-70	4/3
113	Билтен на одбо...	1	3	Скопје	МАНУ	74	1/2
1130	О писменех на ...	0	2			1-6	14/5
1130	Српско-босански ја...	0	2			105-08	25-26

Fig. 10. Resulting table of publications

5.3. Creation of the bibliographic records

The converting process from flat textual records into bibliographic records was one of the most demanding tasks of the whole process of creation of online bibliography. Namely, all the information is presented with different styles, as explained in paragraph 4.

General structure of a bibliographic record is in fact a finite automaton $M = (K, \Sigma, s, \Delta, F)$. In this automaton K is the set of all the authors, titles, editors, publishers, publications, sources, years, pages, discussions, and classifications. Σ is the external alphabet consisting of Unicode symbols. The automaton starts with the initial state s , and finishes with the set of final states F . The relation Δ is the set of rules that transfer the states from the initial state to the final state. All the bibliographic records recognised by this finite automaton determine a regular grammar, i.e. all the records can be transformed into regular expressions.

The process of transformation started with a partitioning of the bibliographic strings. In the beginning, the number of the record is separated from the remaining contents, using the regular expression:

$$\text{^(?<number>\d+)\s*(?<tail>.\+)$}$$

Afterwards, the style of the tail is checked. If the first word from the tail is in italic, then the record starts with the title. Otherwise, the tail continues with the author list. In such case, all the authors are separated from the tail with the expression:

$$\text{^(?<authors>.\+?:)\s*(?<tail>.\+)$}$$

where new tail with extracted author replaces the initial tail. The title and the editors are extracted with the expression:

$$\text{^(?<title>[^|\+?])(?<redaction>/\s.\+?)\s*\.\s*[\-\+](?<tail>.\+)$}$$

All the code was implemented in **System.Text.RegularExpressions** platform. Using this approach, all the regular expressions were transformed into a relational database of bibliographical records. The database consists of eight natural entities, representing the structure of the record.

5.4. Search engine

Search engine used for the simple and advanced search was implemented in ASP.Net environment. It enables both possibilities: to search for a particular entry, or to make a list of all the records that fulfil selected criteria. The user can select according to these eight constraints:

- Word which exists in the record
- Word or character string existing in the title
- Area
- Author
- Editor
- Reviewer
- Person who is a subject of the bibliographic record
- Person with undefined role (Fig. 12)

Search was implemented with the existing ASP.Net modules:

`search.aspx`, `results.aspx`, `adsearch.aspx`, `adresults.aspx` and `details.aspx`.

Similarly to BL Online, search can be restricted to books, articles and discussions, as well as according to exact year, or according to a period (Fig 13). As a result, a list of bibliographic records fulfilling the selected criteria appears on the screen (Fig. 14). It is sorted according to the number of the publication, which is preceded by the volume number.

5.5. Access to information related to a bibliographic record

Each bibliographic record gives a possibility to access several related information. As first, the identification number is a hyperlink towards the details related to the record (Fig. 15). This function was implemented using ASP.Net program component `list.aspx`. Presentation of the bibliographical entry contains their own hyperlinks towards the classifications, which can be multiple (as shown at Fig. 15), towards the authors, and towards the editors. They were also implemented with the same program component. These three lists can also be accessed directly, independently from the bibliographic records they appear in.

5.6. Administrative part of the system

Five previously explained parts of online bibliography are client oriented, and they are actually very static. But, the system is ongoing, and addition of new volumes is expected. Furthermore, during testing phase, many mistakes have been noticed. Extension of the system, and possibility of editing lead to creation of an administrative part of the system. As far as automatic uploading of the new volumes can remain desktop based, the dynamic part comprises only the editing options and correction of accidental mistakes inherited from published volumes, or mistakes originating from manual addition of new records.

Administrative part of the system has been created over the same database, using SQLExpress as data server. In order to disable unintended editing, this database is password protected. This service was made using standard `log.aspx` form, presented at Fig. 11.

```
string Login(string strname, string strPassword)
{
    SqlConnection conn;
    SqlDataAdapter daAdmin;
    string ss =
System.Configuration.ConfigurationManager.AppSettings["cs1"];
string queryString = "Select * from Admin where
Admin.password='"+strPassword+"' and Admin.name='"+strname+"' ";

    conn = new SqlConnection(ss);
    daAdmin = new SqlDataAdapter(queryString, conn);
    DataSet ds = new DataSet();
    daAdmin.Fill(ds, "aaa");
    if (ds.Tables[0].Rows.Count != 0)
    { return strname; }
    else
    { return ""; }
}
```

Fig. 11. Source code used to enable authentication of the administrator

Editing options are applicable for all the information stored within bibliographic records. These records are located according to the publication index, authors, editors, classifications and finally, according to keywords at any place within the record.

Administrative part has been realised into five different layers, all of them realised in Asp.net environment. Their responsibilities are:

- To locate bibliographic records
- To edit basic information concerning selected publication
- To edit the discussions concerning selected publication
- To edit the classification a selected publication belongs to, including the possibility of a publication to have multiple classifications, and
- To edit names of persons involved in the creation of a publication

Source code of these five modules is a standard Asp.net code, and it exceeds the scope of this paper.

6. ILLUSTRATION OF THE BIBLIOGRAPHIC SYSTEM

Search algorithms implemented over this database enable search of books, articles and discussions according to publication type and the year or period of publishing. For selected publication type and period, search can be performed with the exact phrase, or with part of the phrase (Fig 12). Advanced search, which combines search criteria is also enabled (Fig 13).

Пребарувај по :		Внесете збор или фраза за пребарување:	
	<input type="text" value="Збор или фраза од наслов"/>	<input type="text" value="Граматика"/>	
Тип на издание:	<ul style="list-style-type: none"> <input type="checkbox"/> Било кој збор <input type="checkbox"/> Збор или фраза од наслов <input type="checkbox"/> Област <input type="checkbox"/> Автор <input type="checkbox"/> Уредник <input type="checkbox"/> Критичар <input type="checkbox"/> Лице-предмет на библ. единица <input type="checkbox"/> Име (во било која функција) <input type="checkbox"/> Издавач 	<input type="button" value="Барај"/>	
Година на издавање :	<input checked="" type="radio"/> Сите години <input type="radio"/> Пред <input type="radio"/> По <input type="radio"/> Во <input type="text" value="1950"/> <input type="radio"/> Меѓу <input type="text" value="1950"/> и <input type="text" value="1950"/>	За комбинирање на различни термини користете	
Избриши ги сите полиња	Напредно пребарување		

Fig. 12. Simple search

Пребарувај по :		Внесете збор или фраза за пребарување:	
	Автор		пеев
и	Збор или фраза од наслов		земјодел
и	Област		терминологија
	Било кој збор		
	Било кој збор		
Тип на издание:		<input type="button" value="Барај"/>	
<input checked="" type="checkbox"/> Книги <input checked="" type="checkbox"/> Статии <input type="checkbox"/> Дискусии (Критики, Прикази)			
Година на издавање :		Или користете	
<input checked="" type="radio"/> Сите години <input type="radio"/> Пред <input type="radio"/> По <input type="radio"/> Во 1969 <input type="radio"/> Меѓу 1950 и 1950		<input type="button" value="Едноставно пребарување"/>	
Избрани ги сите полиња			

Fig. 13. Advanced search according to selected type and period

Bibliographical records corresponding to defined search criteria are presented in a window. They are sorted according to their identification number, preceded by volume number (Fig. 14). Selected record offers links towards all records with same classification or sub-classification, and all persons involved with its creation (Fig. 15).

Editing options are part of administration part of the application, which is also Web-based. It enables manual addition of new bibliographical records, addition of new categories and subcategories, modification of selected parts within a record, temporary and permanent modification of the name of the person in the database, and deletion of obsolete records. For security purposes, all the previous versions of the database are locally stored, and exchange of accidentally wrong editing with previous stage is possible (Fig. 16).

ON-LINE БИБЛИОГРАФИЈА НА МАКЕДОНСКИОТ ЈАЗИК 1950-1980

[главна](#) [пребарување](#) [листи](#) [контакт](#)

[Назад](#)

[Ново пребарување](#)

Литературен збор 23/1 (Списание на сојузот на друштвата за македонски јазик и литература на СРМ) [томот е посветен на 30-годишнината од првата Граматика на македонскиот литературен јазик] / редакција: Благоја Корубин, Димитар Најчевски, Ружа Паноска, Александар Спасов, Трајко Стаматоски. – Скопје : Сојузот на друштвата за македонски јазик и литература на СРМ, 1976. – 106с. <1.5.>

3/1022

3/1031 Конески, Блаже: Граматика на македонскиот литературен јазик: дел 1 и 2. – Скопје : Култура, 1976. – 552с.

2/1079 Конески, Блаже: Граматика на македонскиот литературен јазик. Дел 1 и 2. – Скопје : Култура, 1967. – 552с. (Избрани дела во седум книги):

2/110 Конески, Блаже: Граматиката на Ѓорѓи Пулевски. – МЛ 4/2, 1953, 45–47. <1.1.; 3.5.>

3/110 Цукески, Александар: 10 Редовно годишно собрание на Сојузот на првата школска граматика на современиот македонски јазик. – ЛЗБ 18/1, 1971, 94–96.

2/112 Поленаковиќ, Харалампие: За граматиката на Пулевски. – МЛ 4/8, 1953, 188–189. <1.1.>

Литературен збор 24/2 (Списание на сојузот на друштвата за македонски јазик и литература на СРМ) [томот е по повод 25-годишнината од излегувањето на Граматиката на македонскиот литературен јазик од Блаже Конески] / редакција: Вера Антиќ, Атанас Николовски, Александар Цукески. – Скопје : Сојузот на друштвата за македонски јазик и литература на СРМ, 1977. – 138с. <1.5.>

3/1219

3/1228 Видоески, Божидар: По повод 25-годишнината од излегувањето на Граматиката на македонскиот литературен јазик од Блаже Конески. – ЛЗБ 24/2, 1977, 3–8.

3/1229 Видоески, Божидар: По повод 25-годишнината од излегувањето на Граматиката на македонскиот литературен јазик од Блаже Конески. – МЛ 28, 1977, 203–206.

3/1402 Стаматоски, Трајко: Делото на проф. Блаже Конески. По повод на 25-годишнината од излегувањето на неговата Граматика. – ЛЗБ 25/1, 1978, 3–7. <2.1.>

1 2 3 4 5 6

Fig. 14. List of all the publications containing the name “Конески” with different roles

Тип на издание книга	
Број	1219
БМЈ	3
Наслов	Литературен збор 24/2 (Списание на сојузот на друштвата за македонски јазик и литература на СРМ) [томот е по повод 25-годишнината од излегувањето на Граматиката на македонскиот литературен јазик од Влаже Конески]
Година на издавање	1977
Место на издавање	Скопје
Издавач	Сојузот на друштвата за македонски јазик и литература на СРМ
Годиште/Број	24/2
Страници	138
Класификација	1.3. Општо /Периодика
	1.5. Општо /Збирки на трудови (во чест ... /по повод ...)
Уредници	Александар Цукески
	Атанас Николовски

Fig 15. Presentation of a selected publication with links towards classifications it belongs to, and towards the editors

As mentioned in the previous chapter, administrative part comprises five different possibilities. As first, a publication can be located according to different criteria. Fig 3b. presents search according to publication number (биб. единица 1107), and volume number (том 2). This search is straight forward, because this information is publication index key. If such a record exists, it appears on the screen. As presented at the same figure, information can be changed at different levels. As first, the title (наслов) can be altered. Secondly, the type (тип), publication year (година год. бр.), place (место), source (извор), publisher (издавач), details (детали), and number of pages (страници) are altered. Periodical source is selected from a drop-down menu.

Modifications at this stage are local. They are mainly used to correct accidental spelling mistakes during the first insertion of the record. Similarly, the discussions (Прикази/дискусии) are altered.

More complex is the editing of classification (Fig. 17), because it affects referential integrity of the database. Namely, all the publications belonging to a particular classifications will also be moved to new classification. Therefore, the warning “Attention, modification affects all the bibliographic records appearing in that classification” (“Внимателно, промената се однесува на област и се однесува на сите биб. единици кои се во оваа област”) appears on the screen.

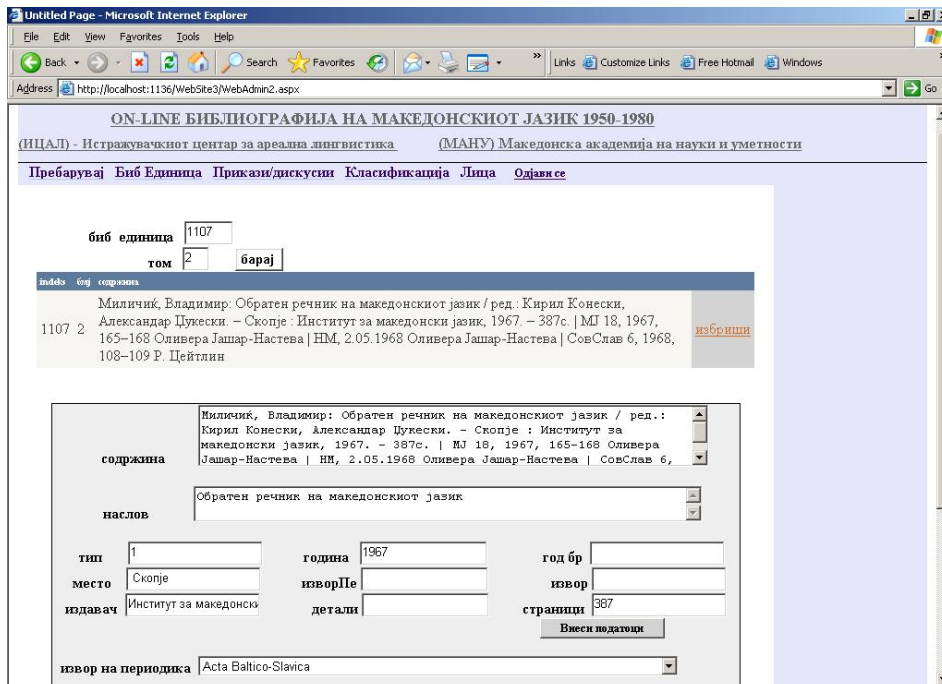


Fig 16. Modifications are made at several levels

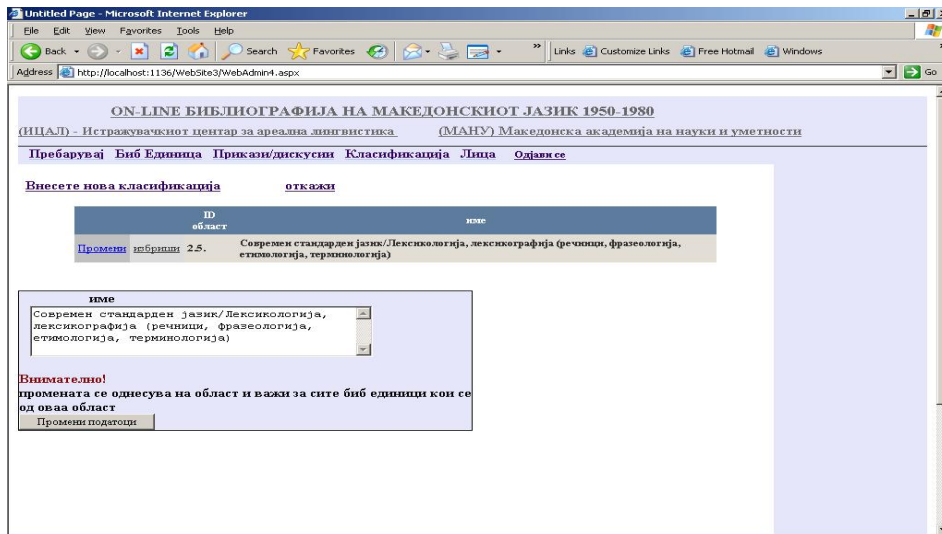


Fig 17. Modification of a classification

Particular attention has been paid to persons in different roles, because in many cases, the same person appears with different names. Such is the case with Božidar – Božo Vidoeski (Божидар – Божо Видоески). Furthermore, the same person can appear in publications written in different languages. Therefore, the entity alias has been established.

Whenever the name is correctly spelled, only the role can be changed (Fig. 18). This modification is local and it doesn't affect the role of the same person within other publications.

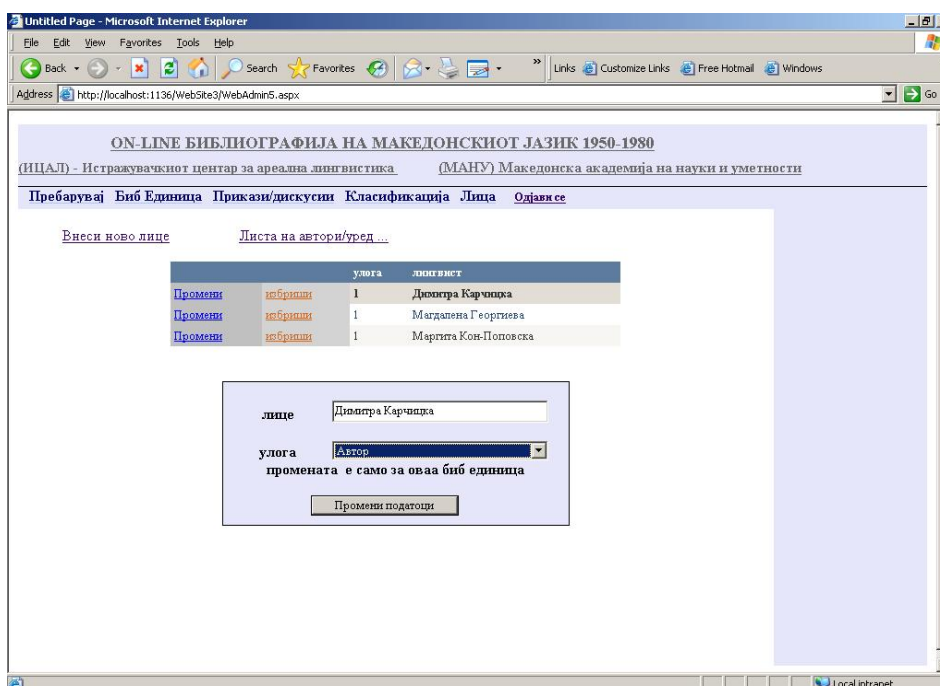


Fig. 18. Modification of the role is local

Finally, if the name of a person has been used as several different names, in such case, the names will remain unchanged in the bibliographic record, while the alias will unit all of them into one person. For example, Димитра Карчицка appears as Карчицка in two publications, and as Корчицка in another two. After the change, she appears as two different persons in the records, but as only one person in the list of persons (Fig. 19).

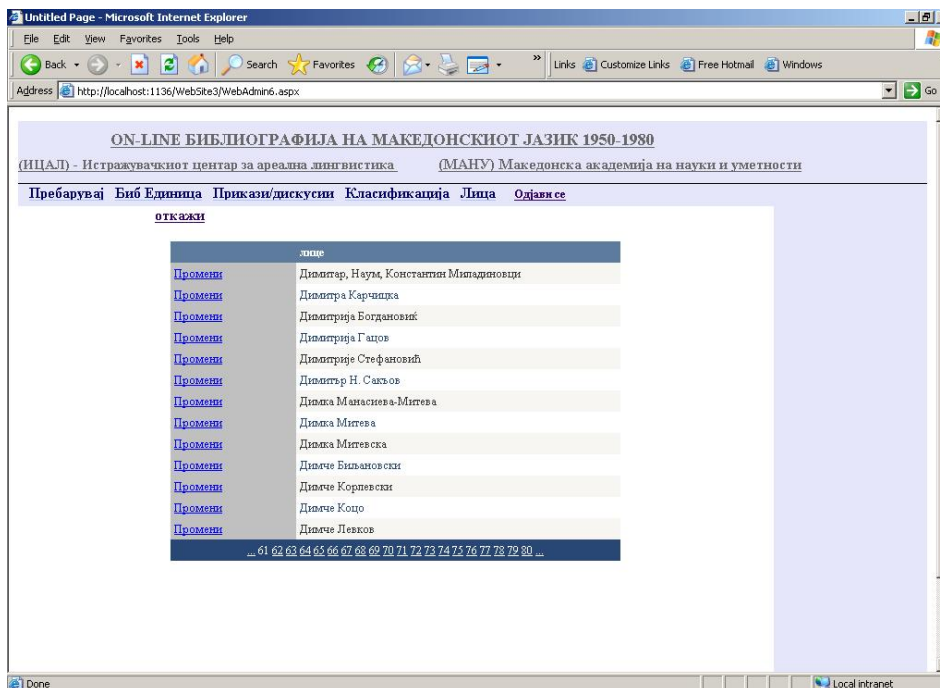


Fig. 19. Modification of person names is global and it doesn't affect the records

7. CONCLUSIONS AND FURTHER DEVELOPMENT OF THE SYSTEM

Creation of online bibliography of Macedonian language is an ongoing project between the Centre for Areal Linguistics within Macedonian Academy of Sciences and Arts, and the Institute of Informatics from the Faculty of Natural Sciences and Mathematics. The main goal of this common effort was to enable gathering of all the publications about Macedonian language from various sources, and to use online availability of these publications for further referencing and indexing.

Current online bibliography contains almost 3000 publication of Macedonian language published in the period between 1954 and 1985, appearing in two recently published volumes [2, 3]. The project presents these two volumes in a form that enables retrieval of all the collected publications, with possibility to correct accidental mistakes, to delete potential multiple records, and to add all the publications which will be collected in the future.

Both applications, client-server and administrative one, are hosted at the Macedonian Academy of Sciences and Arts site, <http://bmj.manu.edu.mk>. At the same time, first volume [1] will soon be available in electronic form, and the fourth volume is under construction. They will verify the scalability, and the correctness of the system. Long-term system evolution is a responsibility of the software manager, and initially of its developers.

Acknowledgment: The authors thank to Irena Anastasova and Goran Nikolov for the development of the code, and for their willingness to correct all the errors noticed by the authors of this paper.

REFERENCES

- [1] B. Vidoeski, *Supplement to Macedonian Language Bibliography, Vol 1*, Institute of Macedonian Language, Skopje (1953).
- [2] Z. Topolinska, M. Mirkulovska, S. Milenkovska (ed), *Supplement to Macedonian Language Bibliography, Vol. 2. (1954–1970)*, Macedonian Academy of Sciences and Arts, Centre of Areal Linguistics (2002) (in Macedonian).
- [3] Topolinska, Z., Mirkulovska, M., Bučevska, A. (ed): *Supplement to Macedonian Language Bibliography, Vol. 3. (1971–1985)* (with additions from previous years), Macedonian Academy of Sciences and Arts, Centre of Areal Linguistics – XXXIII, (2006) (in Macedonian).
- [4] M. Janse, S. Tol, (ed.), *Linguistic Bibliography for the year 1996*, Dodrecht / Boston / London : Kluwer (2000).
- [5] <http://www.blonline.nl/>
- [6] <http://www.loginsys.com.mk/DesktopDefault.aspx?tabindex=0&tabid=160>
- [7] <http://unicode.org/>
- [8] ST RE: NET Framework SDK System.Text.RegularExpressions Namespace, available from [http://msdn2.microsoft.com/en-us/library/system.text.regularexpressions\(VS.71\).aspx](http://msdn2.microsoft.com/en-us/library/system.text.regularexpressions(VS.71).aspx)
- [9] D. Bergmark, *Automatic Extraction of Reference Linking Information from Online Documents*, Technical Report: TR2000-1821 (2000), Cornell University Ithaca, NY, USA, available from www.cs.cornell.edu/cdlrg/Reference%20Linking/extraction.pdf
- [10] J. Geng, J. Yang, AUTOBIB: Automatic Extraction of Bibliographic Information on the Web (2004), in International Database Engineering and Application Symposium, available from <http://doi.ieeecomputersociety.org/10.1109/IDEAS.2004.14>

Резиме

ОНЛАЈН БИБЛИОГРАФИЈА НА МАКЕДОНСКИОТ ЈАЗИК

Онлајн Библиографија на македонскиот јазик (ОнлајнБМЈ) е <http://bmj.manu.edu.mk> е систем составен од две меѓусебно поврзани апликации. Првата претставува десктоп апликација за автоматско групно полнење на томовите и нивно претставување како библиографски записи, а втората е ASP.NET клиент – сервер апликација врз база на C# сервер код, наменета за пребарување и уредување на записите. ОнлајнБМЈ е изработена врз основа на печатените томови на *Прилоџ кон библиографијата на македонскиот јазик* (БМЈ) што го опфаќаат периодот меѓу 1953 и 1985 година. Во оваа статија е претставена примената и функционирањето на онлајн библиографскиот систем. На почетокот се претставени деталите поврзани со неговата изработка. Потоа се претставени сите функционални можности на системот и нивните илустрации. Натаму следи административниот дел на системот. Статијата завршува со заклучоци и насоки за натамошното развивање на системот.

Клучни зборови: автоматска екстракција на библиографските информации; пребарување и уредување текстуални бази податоци

Adress:

Katerina Zdravkova

*Faculty of Natural Sciences and Mathematics,
SS. Cyril and Methodius University
Arhimedova bb, MK-1000 Skopje, Republic of Macedonia
keti@ii.edu.mk*

Milica Mirkulovska

*Faculty of Philology "Blaže Koneski", SS.
Cyril and Methodius University
Krste Misirkov Bld bb, MK-1000 Skopje, Republic of Macedonia
mmirk@manu.edu.mk*

Received: 28. V 2008

Accepted: 11. IX 2008