

МЕТОДИКА АВТОМАТИЗИРОВАННОГО ФОРМИРОВАНИЯ СВЕДЕНИЙ О НАУЧНЫХ ПУБЛИКАЦИЯХ УНИВЕРСИТЕТА ДЛЯ ОТЧЕТА В СИСТЕМЕ УПРАВЛЕНИЯ НИР МИНОБРНАУКИ РОССИИ

О. С. Резниченко, С. И. Сиваков, Т. А. Резниченко

*Белгородский государственный национальный исследовательский университет
Россия, 308015, Белгород, ул. Победы, 85; oreznichenko@bsu.edu.ru*

Аннотация. Ручная подготовка и размещение сведений о публикациях университета в Системе управления НИР занимают значительное время, зачастую несовместимое с установленными сроками, что поставило перед авторами исследования задачу автоматизации указанных процессов. Разработанная методика предполагает использование исключительно инструментария стандартных офисных приложений (Microsoft Word и Excel) и может быть освоена сотрудниками со средним уровнем владения персональным компьютером, не имеющими специализированных знаний в сфере информационных технологий. Результаты исследования продемонстрировали сокращение временных затрат в 3,5 раза при автоматизированном формировании сведений о публикациях по сравнению с ручным вводом данных в профильную информационную систему.

Ключевые слова: Scopus, Web of Science, eLIBRARY.RU, научные публикации, Система управления НИР, Microsoft Excel, Microsoft Word, макросы, транслитерация.

Для цитирования: Резниченко О. С., Сиваков С. И., Резниченко Т. А. Методика автоматизированного формирования сведений о научных публикациях университета для отчета в Системе управления НИР Минобрнауки России // Университетское управление: практика и анализ. 2020. Т. 24, № 2. С. 44–58. DOI: 10.15826/umpa.2020.02.013.

METHOD OF AUTOMATED GENERATION OF INFORMATION ABOUT UNIVERSITY'S SCIENTIFIC PUBLICATIONS FOR REPORTING IN THE RESEARCH MANAGEMENT SYSTEM OF THE RUSSIAN MINISTRY OF SCIENCE AND HIGHER EDUCATION

O. S. Reznichenko, S. I. Sivakov, T. A. Reznichenko

*Belgorod State University 85, Pobedy St., Belgorod, 308015, Russia;
oreznichenko@bsu.edu.ru*

Abstract. Manual preparation and uploading information on university's scientific publications to the Research Management System takes considerable time, and often it is not able to meet the deadlines. This challenge posed the task of automating these processes to the authors of the study. The method involves the use of standard office tools (Microsoft Word and Excel), so that employees who do not have specialized information technology knowledge and have an average level of computer skills could use it. The study demonstrated a 3.5-fold time costs reduction when generating information about university's scientific publications automatically as compared to manual data entry in the Research Management System. *Keywords:* Scopus, Web of Science, eLIBRARY.RU, scientific publications, Research Management System, Microsoft Excel, Microsoft Word, macros, transliteration.

For citation: Reznichenko O. S., Sivakov S. I., Reznichenko T. A. Method of Automated Generation of Information about University's Scientific Publications for Reporting in the Research Management System of the Russian Ministry of Science and Higher Education. University Management: Practice and Analysis. 2020; 24 (2): 44–58. (In Russ.). DOI: 10.15826/umpa.2020.02.013.

Введение

Предоставление различного рода отчетности давно является неотъемлемым компонентом

деятельности образовательных учреждений. Применительно к государственным высшим учебным заведениям одной из платформ для отчетности выступает Единый портал информационного

взаимодействия с учреждениями, подведомственными Минобрнауки России (www.sbias.ru). С 2019 года в рамках исполнения поручений Президента РФ № Пр-2558 от 29 декабря 2018 года государственным высшим учебным заведениям необходимо представлять отчеты о публикационной активности за прошедший календарный год. От содержания и качества этих отчетов зависят будущие объемы госфинансирования вузов, а также возможность их участия в конкурсах на предоставление грантов. Формирование отчетов осуществляется в информационной системе управления НИР, функционирующей на базе штатной подсистемы «Нормирование НИР для государственного задания» (модуль ScientificRationing) программного продукта «ПАРУС-Бюджет 8», разработанного при помощи модуля «Конструктор отраслевых расширений» (КОР).

В данном исследовании рассматриваются вопросы подготовки данных для их импорта в раздел «Сведения о публикационной активности» (далее – информационная система, ИС) подсистемы «Нормирование НИР для государственного задания». Руководство пользователя ИС с описанием вариантов добавления сводных данных, списком исходных информационных баз данных для формирования сводных данных, а также с описанием свойств полей данных и рекомендаций по заполнению каждого поля представлены в Технологической инструкции по внесению данных в раздел «Сведения о публикациях» информационной системы Минобрнауки России по взаимодействию с подведомственными учреждениями (далее – Технологическая инструкция) [1]. Инструкция размещена на информационном портале «Система управления НИР», а также рассылается сотрудниками Минобрнауки подотчетным научным организациям.

В 2018 году Белгородский государственный национальный исследовательский университет (НИУ «БелГУ») подготовил более пяти тысяч научных публикаций, поэтому ручной ввод данных об этих публикациях занял бы значительное время, несопоставимое с требованиями министерства к срокам размещения информации в ИС. Соответственно применение представленной в статье методики нецелесообразно, если требуется внести сведения о не более чем полутора сотнях публикаций. Необходимо также отметить, что в настоящее время на рынке программного обеспечения существуют решения для управления научными исследованиями, включающие в том числе и возможности агрегирования информации из наукометрических баз, а также

формирования и экспорта отчетов в различных разрезах. В качестве примера успешного применения подобной системы можно привести комплексную систему учета и управления научной информацией Elsevier Pure [2, 3]. Эта система кроме агрегации данных из реферативных баз порталов Scopus и Web of Science Core Collection может экспортировать ранее выгруженные данные из базы eLIBRARY.RU, а также систему управления научно-исследовательской деятельностью VP GROUP Academic Line [4], позволяющую накапливать сведения о публикациях, выпускаемых организацией, в рамках встроенной или внешней электронной библиотеки. Условия Национальной подписки НИУ «БелГУ» не позволяют безвозмездно использовать возможности коммерческих решений от Elsevier и VP GROUP, а также интерфейс прикладного программирования (API) реферативной базы eLIBRARY.RU, что приводит к необходимости искать альтернативные способы подготовки и предоставления отчета о публикационной активности. Проблемой исследования является тот факт, что информацию для итогового файла импорта необходимо собирать из различных источников, тогда как средств обработки, форматирования, преобразования этой информации к пригодному для импорта виду имеется ограниченное количество. Кроме того, многие исполнители работ зачастую являются обычными пользователями, не владеющими навыками программирования.

Цель данного исследования состояла в сокращении временных затрат на предоставление сведений о публикациях научной организации в информационную систему управления НИР за прошедший календарный год посредством разработки алгоритмов формирования готового для загрузки в ИС файла, содержащего все необходимые записи.

Описание методов и средств реализации цели исследования

Подробное описание свойств файла для импорта и рекомендаций по заполнению сведений о научных публикациях сотрудников университета приведены в Технологической инструкции [1]. В качестве источников информации о публикациях можно выделить следующие:

- 1) данные, получаемые сотрудниками профильных административных подразделений от научных и научно-образовательных подразделений университета в течение отчетного года или по его окончании [при этом следует предварительно

сформировать пустую электронную форму (например, в виде документа Microsoft Excel) с полями, описанными в Технологической инструкции, разработать упрощенные рекомендации к заполнению этой электронной формы или предусмотреть возможность использования при заполнении формы полей с предопределенными списками];

2) данные, представленные в реферативных базах порталов Scopus, eLIBRARY.RU и Web of Science Core Collection (далее – WoS) [5–7], а также данные, представленные кадровыми службами и сотрудниками подразделений, курирующих грантовую активность научной организации.

Очевидными недостатками первого варианта являются вероятная неполнота предоставляемых сведений, ошибки, неточности в заполнении и подлоги со стороны недобросовестных сотрудников. Большой интерес представляют методика автоматизированного извлечения и обработки информации из реферативных баз данных, а также правила сборки итогового файла для импорта

в ИС. Общая схема, описывающая последовательность формирования итогового файла, представлена на рис. 1.

Способы извлечения данных из реферативных баз таковы:

1) последовательное открытие веб-страницы каждой публикации и ручной перенос требуемых сведений в файл для импорта в ИС (или непосредственно в веб-форму ИС);

2) автоматизированный экспорт требуемых сведений при помощи встроенного инструментария веб-порталов соответствующих реферативных баз.

В проведенном нами исследовании применялись методы системного анализа, методы анализа данных и библиографический анализ. В качестве инструментов исследования использовались приложения Word и Excel из состава пакета Microsoft Office, интернет-браузер с поддержкой возможности сохранения веб-страниц, макросы Microsoft Office Excel.

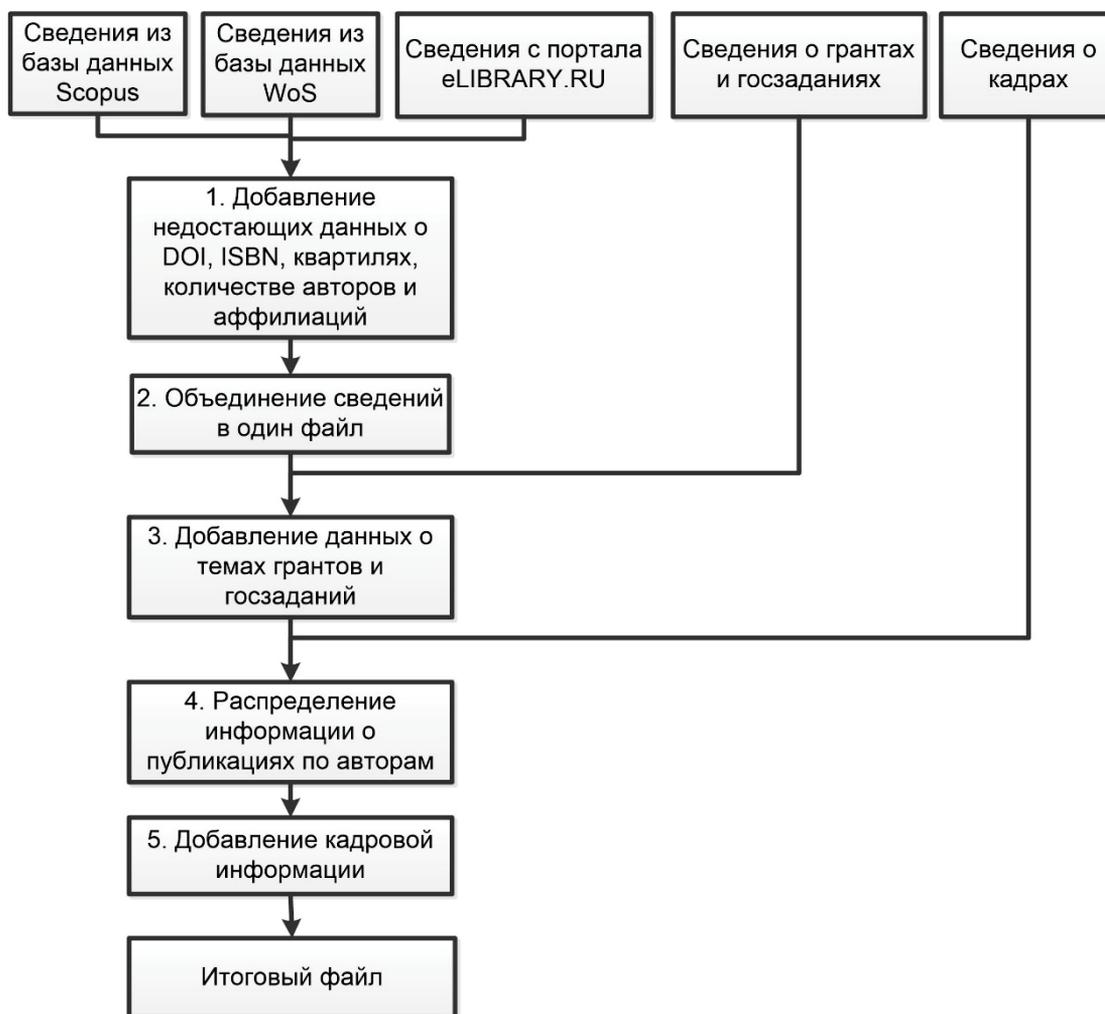


Рис. 1. Общая схема процесса получения итогового файла для импорта в ИС

Fig. 1. General scheme of the process of obtaining the final import file

Описание структур данных о публикациях, извлекаемых из различных источников

Итоговые сводные данные о публикациях организаций должны представлять собой текстовый файл (с расширением *.txt) в кодировке ANSI, поля в котором разделяются символом «;», а перечисления значений внутри полей – символом «,». Пример содержимого такого файла приведен в табл. 1.

Scopus [5] – это библиографическая и реферативная база данных и инструмент для отслеживания цитируемости статей, опубликованных в научных изданиях. Доступна эта база данных через веб-интерфейс на условиях Национальной подписки и только с определенных подпиской IP-адресов. Поисковый аппарат интегрирован с поисковой системой Scirus для поиска веб-страниц и позволяет экспортировать до 20 000 записей, в том числе в формате «Простой текст ASCII в HTML» (рис. 2). Структура записей, экспортируемых из базы данных Scopus, представлена в табл. 2.

Web of Science Core Collection (далее – WoS) [6] – поисковая интернет-платформа,

объединяющая реферативные базы данных публикаций в научных журналах и патентов, в том числе базы, учитывающие взаимное цитирование публикаций. В этой платформе предусмотрены возможности поиска и анализа библиографической информации и управления ею, а также возможность экспортирования до 5 000 записей в текстовом формате с разделителями табуляцией (рис. 3).

База данных Web of Science Core Collection также доступна через веб-интерфейс на условиях Национальной подписки. Структура записей, экспортируемых из базы данных WoS, представлена в табл. 3.

Весь комплекс выходных данных статьи в этом варианте экспорта удобнее объединить в одну запись, так как для основного контента конечного импортного файла это неважно, а важно только в процессе обработки данных экспорта этого типа.

eLIBRARY.ru [7] – это российская научная электронная библиотека, интегрированная с Российским индексом научного цитирования (RSCI). Национальная подписка для университета позволяет пользоваться встроенными

Таблица 1

Структура записи в сводном файле для импорта в ИС

Table 1

Structure of one entry in the summary import for Research Management System

Название поля	Пример записи
DOI	10.1002/a.012678
ISBN	987-3-11-141411-0
Номер темы	FSAD-2019–001
Квартиль	Q1
Количество авторов	12
Автор	Ivanov I. I.
ФИО автора на русском языке полностью	Иванов Иван Иванович
Должность автора в организации	ВНС
Ученая степень	ДОКТ (доктор наук)
Тип трудовых отношений	О
Год рождения автора	1976
Идентификатор	3100
Количество аффилиаций автора	1
Перечень тем автора	AAAA-A11-11223335555-1
Примечания	Адсорбенты из отходов сахарного тростника, их применение для извлечения фенола и 2,4-дихлорфенола из водной среды

Состав полей для экспорта из библиографической и реферативной базы данных Scopus с примерами записей

Table 2

Contents of the fields of the export file from the Scopus database with data sample

Название поля	Пример записи
Автор(ы)	Nelasov, I. V., Lipnitskii, A. G., Kartamyshev, A. I., Maksimenko, V. N., Kolobov, Y. R.
Название документа	Molecular-dynamics simulation of the O±-Ti plastic deformation under conditions of high-energy effects
Название источника, том, выпуск, страницы	AIP Conference Proceedings, 2053, article # 030047
DOI	10.1063/1.5084408
Ссылка на документ в базе	https://www.scopus.com/inward/record.uri?eid=2-s2.0-85059430379&doi=10.1063%2f1.5084408&partnerID=40&md5=3c92a8962985995ea5a236d1ca2e915d

поисковыми возможностями и выводить для просмотра в виде веб-страниц списки научных публикаций университета с применением различных фильтров. Невозможность задействовать в рамках подписки данного типа встроенный API причина того, что извлечь сведения о публикациях удается только одним способом: в виде постраничного сохранения содержимого веб-страниц в формате html средствами браузера по 100 записей (табл. 4)

Кадровая служба организации в нашем случае предоставляет информацию в виде экспортного файла из информационной системы управления кадрами в формате Microsoft Office Excel (Excel) (*.xlsx) и содержит следующие поля данных: полное имя, должность, ученая степень, дата рождения, СНИЛС, вид занятости.

Внутренняя служба, курирующая грантовую активность научной организации, предоставляет

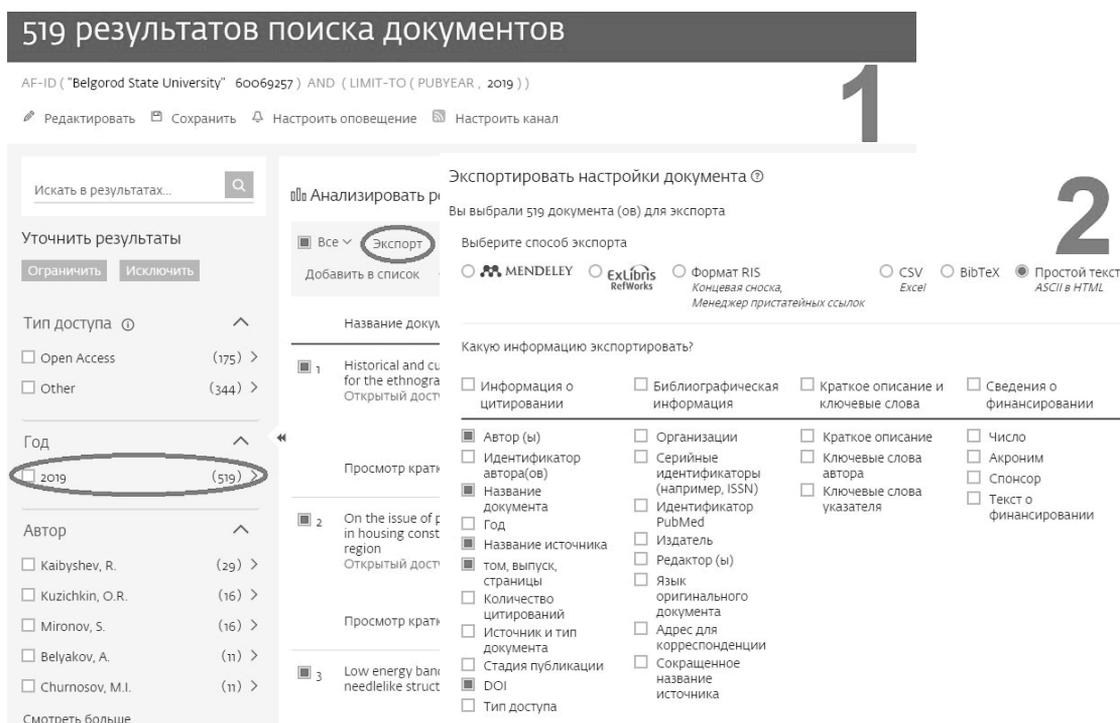


Рис. 2. Последовательность, приводящая к генерации файла-экспорта из библиографической и реферативной базы данных Scopus

Fig. 2. Screenshots describing the process of generating an export file from the Scopus database

Таблица 3

Состав полей для экспорта из поисковой интернет-платформы Web of Science Core Collection с примерами записей

Table 3

Contents of the fields of the export file from the Web of Science database with data sample

Название поля	Пример записи
Автор(ы)	Ivanov, Oleg; Danshina, Elena
Название документа	Peculiarities of the dielectric properties of ternary 0.5(Y0.1Zr0.9O2)-0.5(0.6SrTiO(3)-0.4BiScO(3)) ceramic system
Название источника	CERAMICS INTERNATIONAL
Том	44
Выпуск	18
Начальная страница	22856
Конечная страница	22864
Номер статьи	
DOI	10.1016/j.ceramint.2018.09.078
Индекс документа	WOS:000452345500104

Web of Science

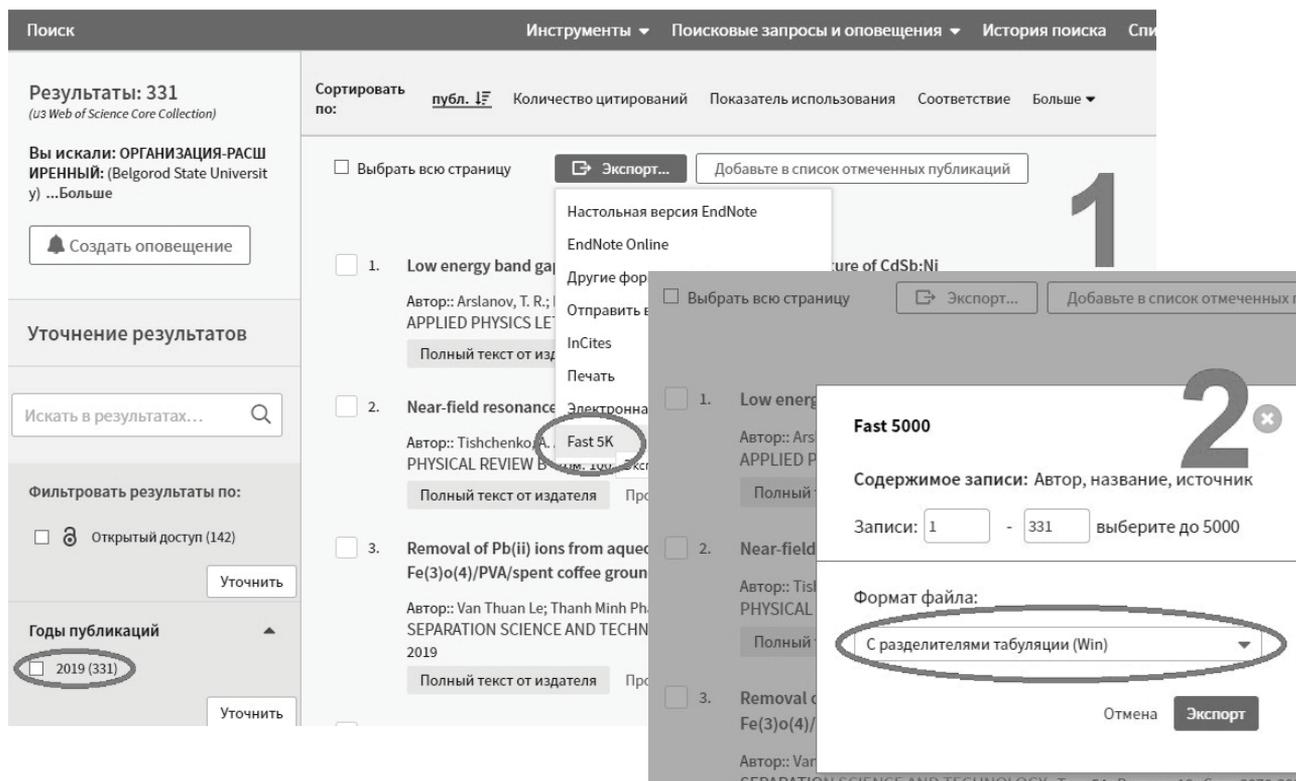


Рис. 3. Последовательность, приводящая к генерации файла-экспорта из базы поисковой интернет-платформы Web of Science Core Collection

Fig. 3. Screenshots describing the process of generating an export file from the Web of Science database

Таблица 4

Состав полей для экспорта из базы российской научной электронной библиотеки eLIBRARY.RU с примерами записей

Table 4

Contents of the fields of the export files from eLIBRARY.RU with data sample

Название поля	Пример записи
Название документа	MICROSTRUCTURE EVOLUTION OF A TI/TIB METAL–MATRIX COMPOSITE DURING HIGH-TEMPERATURE DEFORMATION
Автор(ы)	Ozerov M. S., Klimova M. V., Stepanov N. D., Zherebtsov S. V.
Название источника, год, том, выпуск, страницы	Materials Physics and Mechanics. 2018. Т. 38. № 1. С. 54–63

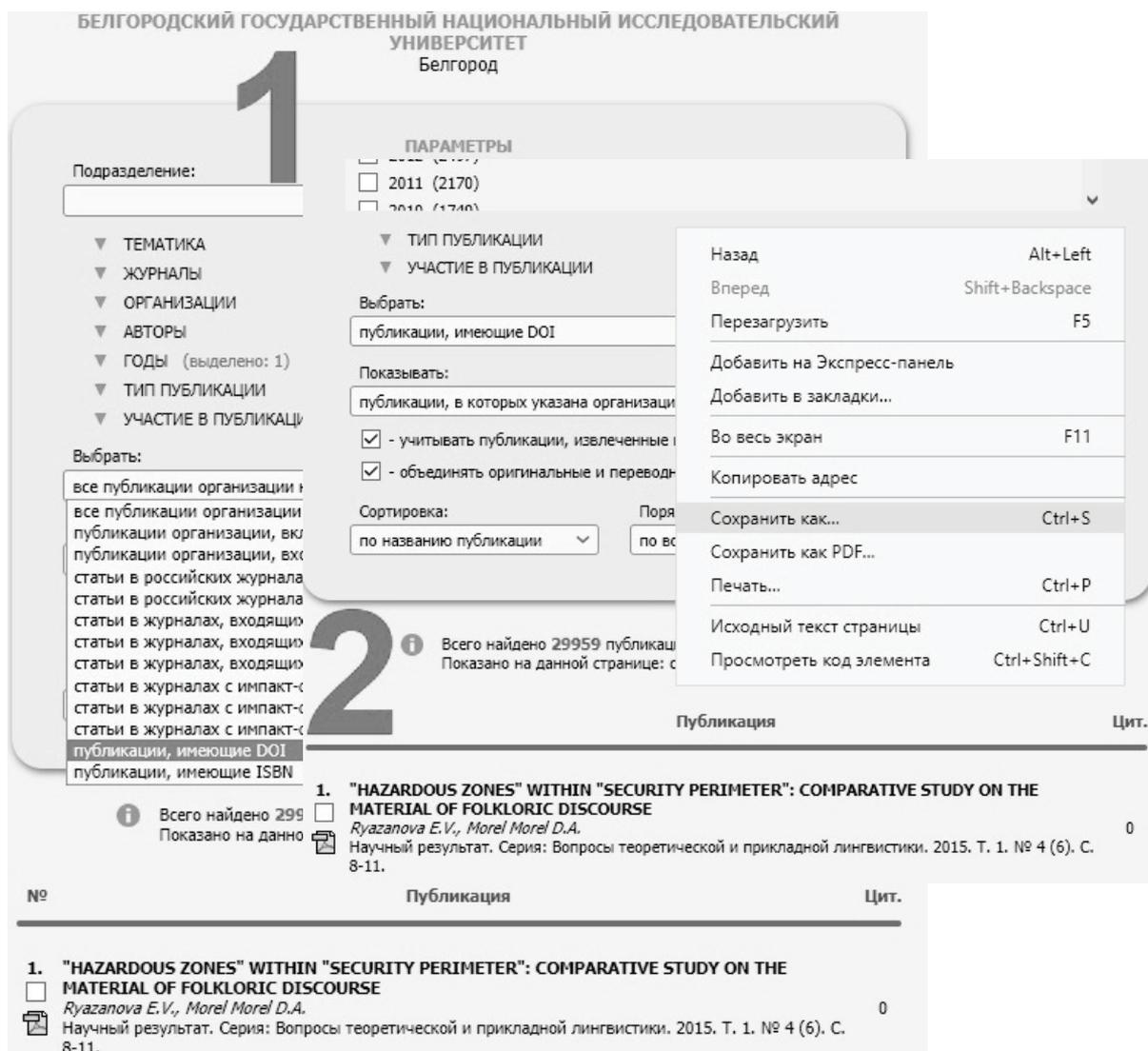


Рис. 4. Последовательность, приводящая к генерации серии экспортных файлов из российской научной электронной библиотеки eLIBRARY.ru

Fig. 4. Screenshots describing the process of generating an export file from eLIBRARY.ru

информацию о перечне показателей, которые характеризуют выполнение работ и достигнутые результаты по каждому научному проекту, имеющему номер ЕГИСУ НИОКР, в том числе списки значимых публикаций в международных базах цитирования. Сведения о научных проектах предоставляются в нашем случае в табличном виде в pdf-файлах. Номер темы, соответствующий каждому из проектов, выбирается, соответственно, согласно реестру тем из ИС.

Создание алгоритмов обработки данных при формировании файла импорта

Ниже описывается процесс обработки сведений о публикациях, полученных из соответствующих реферативных баз, а также процесс добавления недостающих сведений (см. шаг 1 на рис. 1).

На рис. 5 приведен алгоритм обработки данных файла экспорта при извлечении сведений о научных публикациях из библиографической и реферативной базы данных Scopus.

Описание алгоритма обработки файла экспорта при извлечении сведений о научных публикациях из базы данных Scopus

Шаг 1: процесс экспорта файла с данными из Scopus (отображен на рис. 2).

Шаг 2: процесс преобразования экспортного файла в формат Excel с полями, соответствующими полям в табл. 2 (отображен на рис. 6).

Шаги 3–5: согласно требованиям Технологической инструкции при отсутствии в статье DOI соответствующее поле необходимо заполнять значениями идентификатора библиографической и реферативной базы данных Scopus (EID). EID содержится в структуре URL-адреса и может быть извлечен в отдельное поле функцией Excel MID (ПСТР (например, вида =ПСТР(12;46;18).

Шаг 6: согласно требованиям Технологической инструкции для всех статей базы данных Scopus в поле «Квартиль» должно проставляться значение «S».

Шаг 7: для подсчета количества авторов публикации последовательно выполняются указанные ниже действия:

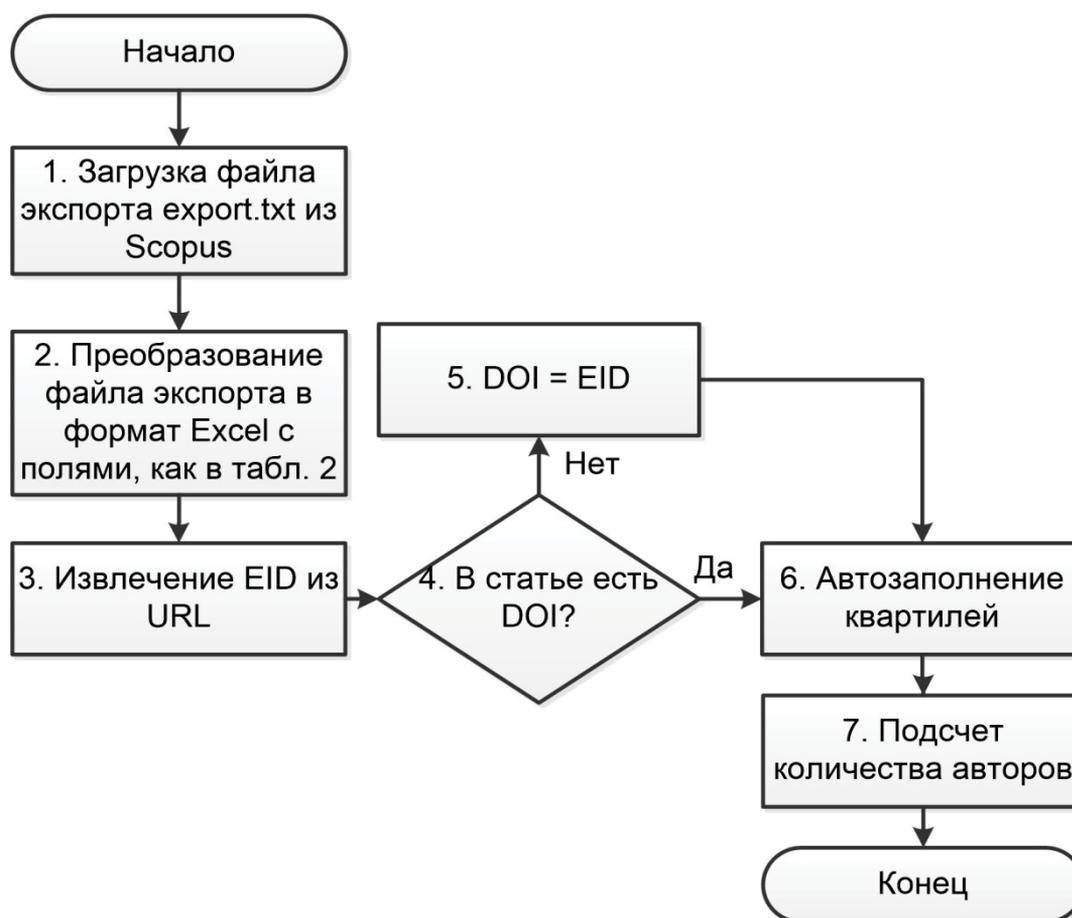


Рис. 5. Алгоритм обработки файла экспорта при извлечении сведений о публикациях из библиографической и реферативной базы данных Scopus

Fig. 5. Scopus publications data export processing algorithm

	A	B	C	D	E	F
1	Ladik, E.I., Elservi, O.E.M.M., Konev					
2	Historical and cultural landscape as					
3	(2019) IOP Conference Series: Mate					
4	https://www.scopus.com/inward/r=A1	=A2	=A3	=A4	=A6	
5						
6	DOI: 10.1088/1757-899X/698/3/03					
7						
8	Parfenova, E.N., Avilova, Gh.N., Poly					
9	On the issue of public-private partn					
10	(2019) IOP Conference Series: Mate					
11	https://www.scopus.com/inward/r=Parfenova, E.N., Avilova, On the issue of pul (2019) IOP Conferer https://www.scopu DOI: 10.1088/1757-89					
12						
13	DOI: 10.1088/1757-899X/698/7/07					
14						

Рис. 6. Преобразование экспортного файла export.txt в файл, пригодный для обработки в Excel

Fig. 6. Converting the export file «export.txt» to usable in Excel

- 1) поле Authors копируется на отдельный лист;
- 2) в значениях поля Authors последовательность символов «.,» заменяется на символ «.»;
- 3) применяется функция «Текст по столбцам» во вкладке «Данные» ленты с разделителями «.» [8];
- 4) с использованием функции COUNTA (СЧЁТЗ) в отдельном столбце вычисляется количество заполненных в строке ячеек (рис. 7).

На рис. 8 приведен алгоритм обработки данных файла экспорта при извлечении сведений о научных публикациях из базы поисковой интернет-платформы WoS.

Описание алгоритма обработки файла экспорта при извлечении сведений о научных публикациях из базы данных WoS

Шаги 1–3: в экспортном файле выходные данные каждой статьи распределены по нескольким столбцам, поэтому для того чтобы сформировать в отдельном столбце единую строку с выходными данными, необходимо осуществить сцепление

значений нескольких столбцов, используя операцию Excel «&» (рис. 9) [9].

Шаги 4 и 5: согласно требованиям Технологической инструкции при отсутствии DOI соответствующее поле необходимо заполнять значениями индекса документа (WoS ID, см. табл. 3).

Шаг 6: для формирования из значений поля UT файла savedrecs.txt строки с поисковым запросом для расширенного поиска статьи необходимо использовать формулу Excel вида «=»UT«&I2», где ссылка на ячейку I2 содержит значение поля UT.

Шаг 7: для подсчета количества авторов применяется алгоритм, использующийся для обработки файла экспорта из базы Scopus (пункт 2 исключается).

Шаг 8: в файле экспорта не содержится данных о квантиле журнала, поэтому квантиль необходимо искать на портале WoS [6] с применением средств расширенного поиска посредством поискового запроса, полученного на шаге 6.

При выгрузке данных о публикациях из портала eLIBRARY.RU в соответствии с рекомендациями Технологической инструкции на странице

	A	B	C	D	E	F	G
1	=СЧЁТЗ(B1:JX1)	Rumyantsev, M.B	Turanin, V.Y.	Akopyan, A.V.	Alontseva, D.V.	Batova, O.V.	
2	10	Polonikov, A.V.	Ponomarenko, I.	Bykanova, M.	Sirotnina, S.S.	Bocharov; Vagayt	
3	2	Buryak, Z.	Grigoreva, O.				
4	5	Grigorenko, N.V.	Tsurikova, L.V.	Kaliuzhnaya, E	Bubyreva, Z.A.	Lukyanova, E.V.	
5	4	Babintsev, V.P.	Goncharuk, Y.A.	Goncharuk, S.	Komarova, I.G.		
6	4	Voronkov, A.V.	Nikulin, I.N.	Kolesnikov, A.	Nikulina, D.E.		
7	5	Voloshina, L.N.	Buslovskaya, L.K.	Kovtunencko, I	Ryzhkova, Y.P.	Prokopenko, Y.A.	
8	2	Dabagov, S.	Gladkikh, Y.P.				

Рис. 7. Пример вычисления количества авторов публикации с использованием функции COUNTA (СЧЁТЗ)

Fig. 7. Example of calculating the number of publication authors using the Excel COUNTA function

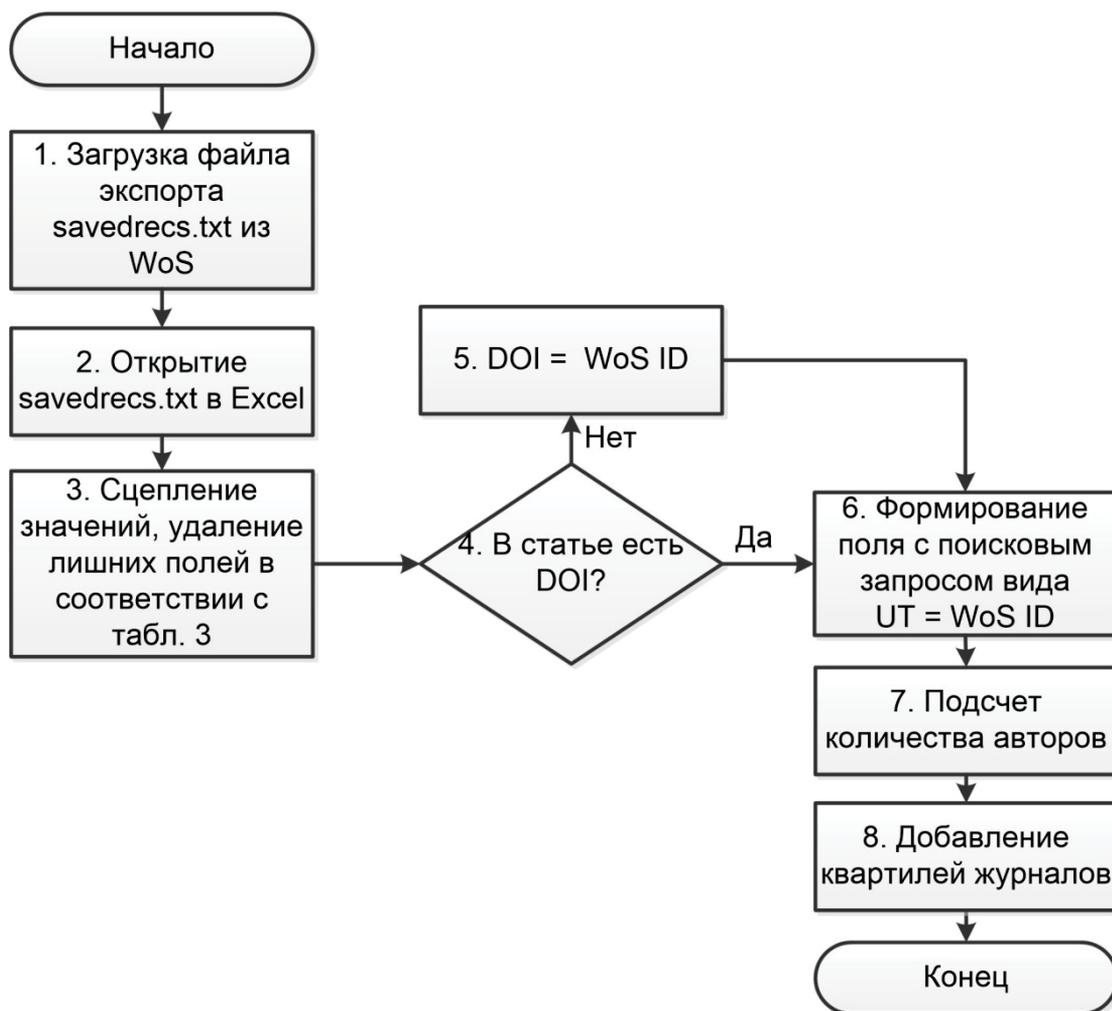


Рис. 8. Алгоритм обработки данных файла экспорта при извлечении сведений о публикациях из базы поисковой интернет-платформы WoS

Fig. 8. Web of Science publications data export processing algorithm

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P		
1	TI	AU	SO	VL	IS	BP	EP	AR	PY									
2	Superior cri Fedoseev; MATERIAL			262				127183	2020	-	Vol.262				-	at.#127, 2020	=C2&O2&J2&K2&L2&M2&N2	
3	Thermoelel Yaprntsev; JOURNAL	40	3	742	750				2020	-	Vol.40	-	Is.3	-	pp.742 -750		JOURNAL OF THE EUROPEAN CERAMIC SOCIETY, 2020. - Vol.40. -	
4	Crystal stru Vershinihine; VACUUM			172				UNSP 109	2020	-	Vol.172					-	at.#UN, 2020	VACUUM, 2020. - Vol.172. - at.#UNSP 109034
5	Creep strer Tkachev, i; MATERIAL			772				138821	2020	-	Vol.772					-	at.#138, 2020	MATERIALS SCIENCE AND ENGINEERING A-STRUCTURAL MATERIA/
6	Microstruc' Vysotskiy; MATERIAL			770				138540	2020	-	Vol.770					-	at.#138, 2020	MATERIALS SCIENCE AND ENGINEERING A-STRUCTURAL MATERIA/
7	Novel ceriui Golovin, S; CHEMICAL			74	1	367	370		2020	-	Vol.74	-	Is.1	-	pp.367 -370			CHEMICAL PAPERS, 2020. - Vol.74. - Is.1. - pp.367-370
8	The enhanc Goldberg; JOURNAL	9	1	76	88				2020	-	Vol.9	-	Is.1	-	pp.76 -88			JOURNAL OF MATERIALS RESEARCH AND TECHNOLOGY-JMR&T, 2
9	Specific ver Ukhnalevz; AMAZONI			9	25	126	134		2020	-	Vol.9	-	Is.25	-	pp.126 -134			AMAZONIA INVESTIGA, 2020. - Vol.9. - Is.25. - pp.126-134
10	Gum-like m Zhrebetso; INTERME			116				106652	2020	-	Vol.116					-	at.#106, 2020	INTERMETALLICS, 2020. - Vol.116. - at.#106652

Рис. 9. Пример сцепления значений столбцов из экспортного файла WoS для получения строки с выходными данными публикации

Fig. 9. Example of concatenating column values from a Web of Science export file to produce a single row with publication imprint

списка публикаций университета необходимо последовательно воспользоваться указанными ниже фильтрами:

- I. статьи в журналах, включенных в перечень ВАК;
- II. статьи в журналах, входящих в RSCI;

- III. публикации, имеющие ISBN;
- IV. публикации, имеющие DOI.

Алгоритм обработки файла экспорта при извлечении сведений о научных публикациях из библиографической и реферативной базы данных eLIBRARY.RU приведен на рис. 10.

Описание алгоритма обработки файла экспорта при извлечении сведений о научных публикациях из базы данных eLIBRARY.RU

Шаги 1–3: экспортированные веб-страницы со списком публикаций (см. рис. 4) сцепляются друг с другом в один файл и обрабатываются средствами Excel.

Шаг 4: при обработке экспортного файла базы данных eLIBRARY.RU для извлечения из названия статьи URL-адреса можно воспользоваться приведенным ниже адаптированным макросом Excel [10].

```
Sub GetURL()
With ActiveSheet
For I = 1 To .Hyperlinks.Count
Hyperlinks(I).Range.Offset(0, 3).
Value = .Hyperlinks(I).Address
Next I
End With
End Sub
```

При выполнении макроса URL-адрес на соответствующей статье пропишется в третьем столбце, следующим за столбцом с ее заголовком.

Шаги 5–14: предназначены для выполнения требований Технологической инструкции в контексте заполнения DOI и квартиля журнала публикации для списков публикаций, полученных при использовании различных фильтров.

Шаг 15: для подсчета количества авторов применяется алгоритм, использующийся для обработки файла экспорта из базы данных Scopus (пункт 2 исключается).

Наиболее трудозатратным и тяжело поддающимся автоматизации является процесс добавления в итоговый файл импорта информации о суммарном количестве аффилиаций авторов, которые публикуются от университета. Чтобы получить эти сведения, необходимо иметь прямые ссылки на страницы публикаций в соответствующих реферативных базах. В базах Scopus и eLIBRARY.RU такие ссылки можно получить непосредственно при экспорте данных или после

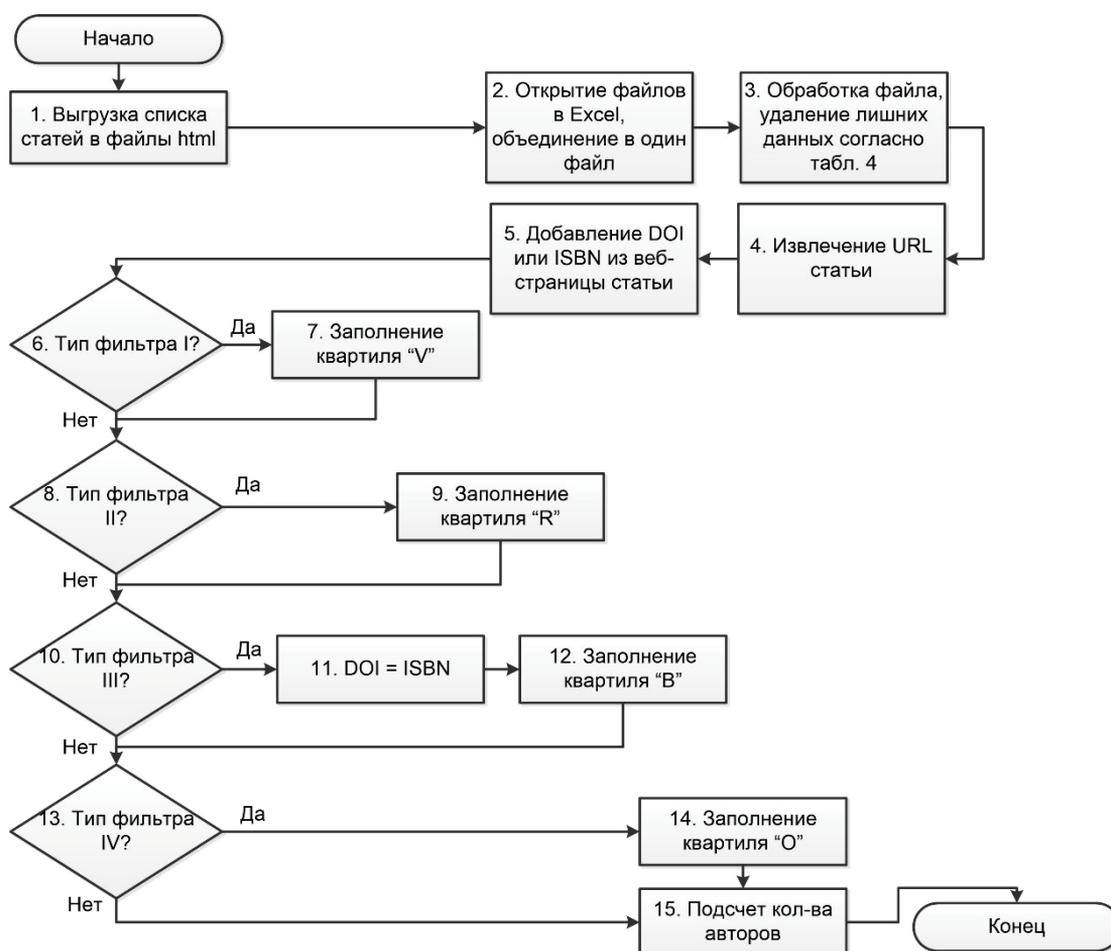


Рис. 10. Алгоритм обработки данных файла экспорта при извлечении сведений о научных публикациях из базы eLIBRARY.RU

Fig. 10. eLIBRARY.RU publications data export processing algorithm

соответствующей их обработки, а в базе WoS – только осуществив расширенный поиск по WOS ID статьи, используя поисковый запрос, полученный на шаге 6 (см. рис. 8). При отсутствии возможности организовать парсинг веб-страниц информацию о количестве аффилиаций авторов, а в случае с WoS – еще и данных о квинтиле журнала (по умолчанию посредством автозаполнения проставляется квинтиль Q4) можно получить только ручным переносом значения квинтиля со страницы статьи.

В результате обработки данных из трех источников (см. шаг 2 на рис. 1) получается электронная таблица, содержащая поля, представленные в табл. 5.

Необходимость экспорта из трех реферативных баз приведет в том числе к появлению одинаковых записей. Подобные дубликаты можно удалить, используя функцию Excel IF (ЕСЛИ). Применяя формулу =IF(LEFT(A2;35)=LEFT(A3;35);1;0), можно сформировать новый служебный столбец, по значениям которого будут отфильтровываться дубликаты. В вопросе о том, какую именно запись необходимо оставить, приоритет отдается публикациям WoS [1].

На следующем этапе добавляется информация о грантовых тематиках и номере грантовой темы непосредственно для каждой публикации (см. шаг 3 на рис. 1), которая осуществлялась в рамках соответствующего грантового исследования.

При обработке файла экспорта, предоставленного кадровой службой, необходимо придерживаться следующего алгоритма:

1) замена значений в столбцах, содержащих информацию о должности, ученой степени, виду занятости, на соответствующие

сокращения согласно требованиям Технологической инструкции [1];

2) извлечение года рождения автора публикации и последних 4 цифр СНИЛС;

3) разбиение полного имени сотрудника по столбцам «Фамилия», «Имя», «Отчество»;

4) транслитерация записей в столбцах «Полное имя», «Фамилия», «Имя», «Отчество» согласно алгоритму [11] в отдельные дополнительные столбцы;

5) формирование двух столбцов со значениями вида «Фамилия И. О.» и их транслитерированными вариантами.

Следующий этап состоит в распределении информации о статье по всем авторам этой статьи (см. шаг 4 на рис. 1), то есть таким образом, чтобы количество записей, соответствующих одной статье, было равно числу соавторов в ней. Подобное распределение можно осуществить посредством такого алгоритма:

1) копирование содержимого электронной таблицы в документ Microsoft Office Word (Word);

2) замена символов разделителей авторов в поле «Авторы» на символ абзаца;

3) копирование содержимого документа на новый лист документа Excel;

4) автозаполнение пустых значений полей DOI, ISBN, «Квинтиль», «Количество авторов», «Примечание» с использованием формулы =IF(A2="";B1;A2).

Следующий этап заключается в добавлении в результирующий файл импорта недостающих полей из адаптированного файла (см. шаг 5 на рис. 1), содержащего сведения о работниках. Для этого будет использоваться функция VLOOKUP (ВПР) [12], в зависимости от языка

Таблица 5

Состав полей итогового файла импорта, содержащего сведения о научных публикациях, с примерами записей

Table 5

Contents of the fields of the final import file with information about scientific publications with sample data

Название поля	Пример записи
DOI	10.1088/1748-0221/13/02/C02048
ISBN	–
Квинтиль	Q3
Количество авторов	5
Авторы	Vokhmyanina K. A.; Sotnikova V. S.; Kaplii A. A.; Sotnikov A. V.; Kubankin A. S.
Название статьи	About a contactless transmission of 10 keV electrons through tapering microchannels
Выходные данные статьи	JOURNAL OF INSTRUMENTATION, Vol. 13, art.# C02048

публикации добавляющая из таблицы с кадровой информацией поля «Полное имя», «Должность», «Ученая степень», «Год рождения», «СНИЛС», ориентируясь на равенство значений поля формата «Фамилия И. О.» (или его транслитерированного варианта) и значений поля «Авторы» итогового файла. В нашем случае используется Excel-функция вида =ВПР(A2; Employees!\$J:\$W; i; 0), где i – номер столбца из диапазона J:W, содержащий недостающее для итогового файла поле электронной таблицы с информацией о кадрах. В результате использования функции ВПР произойдет автоматическое заполнение персональной информации об авторе, если она имеется в таблице с информацией о кадрах. Нужно проанализировать, для каждой ли статьи нашлись соответствующие персональные данные. Наиболее вероятной причиной отсутствия в записи кадровой информации является различие между написанием фамилии автора в файлах экспорта из реферативных баз и транслитерированной записью в таблице с информацией о кадрах. Найденные несоответствия исправляются вручную. Записи, оставшиеся после проведенного анализа пустыми, отфильтровываются и из итоговой таблицы удаляются. На этом этапе формирование файла импорта завершено.

Заключение

Итогом данного прикладного исследования является разработка методики, позволяющей сформировать готовый файл электронной таблицы Excel, содержащий несколько типов записей о научных публикациях. Данный файл сохраняется в текстовом формате с требуемыми Технологической инструкцией разделителями.

В табл. 6 представлены временные затраты на внесение в ИС 2473 записей вручную и посредством формирования файла импорта. Временные затраты на перенос сведений о научных проектах организации в файл для импорта составили 87 минут. Замеры времени для переноса этих же сведений вручную непосредственно в ИС не производились, но можно с уверенностью утверждать, что этот процесс занял бы значительно больше времени, так как данные, относящиеся к одному научному проекту, пришлось бы вносить в записи по каждому из соавторов одной публикации. Эксперимент по ручному и автоматизированному заполнению системы совершал один и тот же сотрудник центра развития публикационной активности НИУ «БелГУ».

Приведенные в табл. 6 данные свидетельствуют, что увеличение количества сотрудников, осуществляющих подготовку и ручной перенос в ИС сведений о публикациях, кадровых данных и сведений о научных проектах, приведет к уменьшению суммарных временных затрат на выполнение работы. В свою очередь, увеличение количества публикаций, количества научных проектов организации, а также доли авторства в публикациях за отчетный период будет способствовать значительному увеличению разницы между временными затратами при ручном вводе сведений и при их автоматизированной обработке.

Рекомендации по использованию предлагаемой методики:

1) в связи с особенностями работы ИС управления НИР содержимое файла импорта необходимо разбить на несколько отдельных файлов объемом не более чем по 100 записей в каждом, а затем уже импортировать итоговые текстовые файлы в кодировке ASCII встроенными средствами импорта;

Таблица 6

Временные затраты на добавление сведений о публикациях в ИС вручную и посредством формирования файла импорта

Table 6

Time spent on adding information about publications to the Research Management System

База данных	Среднее время, затраченное на добавление одной записи, мин		Количество обработанных записей, ед.
	При ручном переносе данных	При автоматизации формирования и обработки данных	
Scopus	1,30	0,50	928
WoS	1,57	0,17	687
eLIBRARY.RU	1,52	0,48	858
Экспортный файл кадровой службы	32	32	–

Примечание. При ручном переносе данных в ИС затраченное время составило 3621 мин, при автоматизированном – 1025 мин.

2) в качестве поля с примечанием можно вставить данные о названиях или выходных данных научной публикации или ссепку «название статьи»&»выходные данные», чтобы облегчить в будущем навигацию по загруженным записям, а также быстро исправлять ошибки, возникшие при формировании итогового файла импорта;

3) информацию о публикациях в рамках грантов (проектов) следует получать в более формализованном табличном виде, а не в виде паспорта проекта.

«Узкими местами» предложенного метода являются неверное сопоставление авторов, имеющих одинаковые фамилии и инициалы, а также вариативность транслитерации фамилий авторов. В нашем случае корректировка результатов транслитерации фамилий производилась непосредственно перед формированием итогового файла путем поиска и исправления записей, для которых функция ВПР не добавила сведений из таблицы с данными о кадрах. Указанные проблемы могут быть решены путем ввода в анкеты сотрудников информационной системы управления кадрами организации дополнительных полей, содержащих их уникальные идентификаторы в крупных наукометрических базах (Scopus ID, Research ID, ORCID и другие).

В перспективе при разработке программных средств автоматизации формирования файла импорта для хранения экспортной информации можно использовать базу данных, структура которой представлена на рис. 11.

В рамках дополнительного модуля программных средств автоматизации с целью повышения качества программного обеспечения можно разработать и реализовать функционал онлайн-ового или офлайн-ого парсинга веб-страниц публикаций для автоматизации извлечения информации о количестве аффилиаций авторов применительно ко всем исходным реферативным базам; для автоматизации извлечения информации о квартиле журналов в базе WoS; а также для извлечения значений DOI и ISBN на страницах публикаций в базе eLIBRARY.RU, что приведет к значительному сокращению временных затрат на выполнение автоматизированного формирования итогового файла для импорта.

Список литературы

1. Инструкция по действиям в разделе «Сведения о публикациях» (обновленная) // Министерство науки и высшего образования Российской Федерации : официальный сайт. URL: <http://wnir.minobrnauki.gov.ru/index.php?newsid=67> (дата обращения: 04.04.2020).
2. Pure // Elsevierscience. URL: <http://elsevierscience.ru/products/pure/> (дата обращения: 15.05.2020).
3. Илюхин Д. Продвигаем ученых в мировом пространстве // Уральский Федеральний : [периодическое издание Уральского федерального университета им. Б. Н. Ельцина]. 2020. № 7. С. 2. URL: https://urfu.ru/fileadmin/user_upload/common_files/news/2020/02/20200210_gazeta_UF_7.pdf (дата обращения: 15.05.2020).
4. Система управления научно-исследовательской деятельностью вуза // VP Group : [сайт]. URL: <http://vpgroup.ru/Default.aspx?tabid=436> (дата обращения: 15.05.2020).

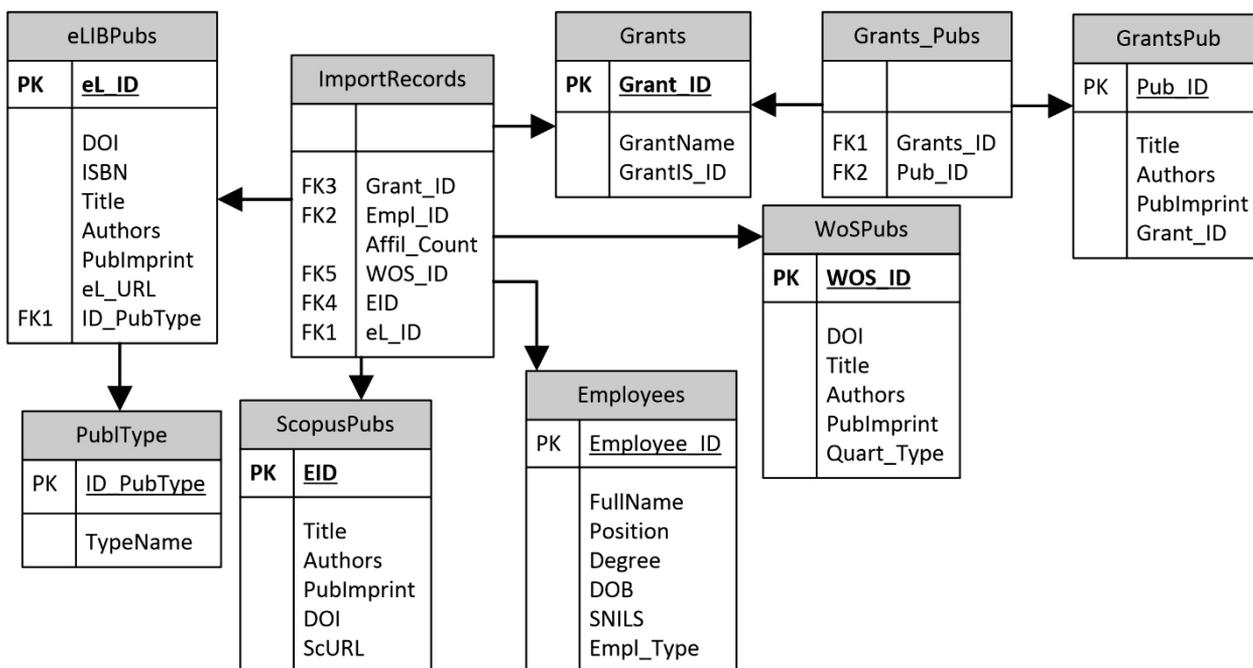


Рис. 11. Структура базы данных для хранения экспортной информации из различных реферативных баз
 Fig. 11. Data structure for storing exported information from Scopus, Web of Science, eLIBRARY.RU

5. Elsevier Scopus // Scopus : библиографическая и реферативная база данных. URL: <https://www.scopus.com/search/form.uri?display=basic=> (дата обращения: 04.04.2020).
6. Clarivate Analytics Web of Science Core Collection // Web of Science : поисковая интернет-платформа. URL: https://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=C3Qtws6Zp9bRCWtj7S7&preferencesSaved= (дата обращения: 04.04.2020).
7. eLIBRARY.RU : Российская научная электронная библиотека. URL: <https://elibrary.ru/defaultx.asp> (дата обращения: 04.04.2020).
8. Text to Columns // Excel Easy – #1 Excel tutorial on the net. URL: <https://www.excel-easy.com/examples/text-to-columns.html> (дата обращения: 15.05.2020).
9. Текстовые функции (справка) // Поддержка Office. URL: <https://support.office.com/ru-ru/article/Текстовые-функции-справка-cccc86ad-547d-4ea9-a065-7bb697c2a56e> (дата обращения: 15.05.2020).
10. How to extract a URL from a hyperlink on Excel // How to Use Excel. Tips on how to use Microsoft Excel. URL: <https://howtouseexcel.net/how-to-extract-a-url-from-a-hyperlink-on-excel> (дата обращения: 04.04.2020).
11. Транслитерация текста в Excel // Tutorexcel.ru. Excel простыми словами. URL: <https://tutorexcel.ru/makrosy-vba/transliteraciya-teksta-v-excel/> (дата обращения: 04.04.2020).
12. Функция ВПР в программе Microsoft Excel // Lumpics.ru. : [сайт]. URL: <https://lumpics.ru/examples-the-lookup-function-in-excel/> (дата обращения: 04.04.2020).
3. Piyukhin D. Prodviagam uchenykh v mirovom prostranstve [Promoting Our Scientists in World Science]. *Uralskii Federalnii*, 2020, no. 7, p. 2, available at: https://urfu.ru/fileadmin/user_upload/common_files/news/2020/02/20200210_gazeta_UF_7.pdf (accessed 15.05.2020). (In Russ.).
4. Sistema upravleniya nauchno-issledovatel'skoi deyatel'nosti vuza [University Research Management System], available at: <http://vpgroup.ru/Default.aspx?tabid=436> (accessed 15.05.2020). (In Russ.).
5. Elsevier Scopus, available at: <https://www.scopus.com/search/form.uri?display=basic=> (accessed 04.04.2020). (In Eng.).
6. Clarivate Analytics Web of Science Core Collection (2020), available at: https://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=C3Qtws6Zp9bRCWtj7S7&preferencesSaved= (accessed 04.04.2020). (In Eng.).
7. eLIBRARY.RU: Nauchnaya elektronnyaya biblioteka [Scientific Electronic Library eLIBRARY.RU], available at: <https://elibrary.ru/defaultx.asp> (accessed 04.04.2020). (In Russ.).
8. Text to Columns – Easy Excel Tutorial, available at: <https://www.excel-easy.com/examples/text-to-columns.html> (accessed 15.05.2020). (In Eng.).
9. Tekstovye funktsii (spravka) [Text functions (reference)], available at: <https://support.office.com/ru-ru/article/Текстовые-функции-справка-cccc86ad-547d-4ea9-a065-7bb697c2a56e> (accessed 15.05.2020). (In Russ.).
10. How to use Microsoft Excel. How to extract a URL from a hyperlink on Excel, available at: <https://howtouseexcel.net/how-to-extract-a-url-from-a-hyperlink-on-excel> (accessed 04.04.2020). (In Eng.).
11. Transliteratsiya teksta v Excel [Transliteration of text in Excel], available at: <https://tutorexcel.ru/makrosy-vba/transliteraciya-teksta-v-excel/> (accessed 04.04.2020). (In Russ.).
12. Funktsiya VPR v programme Microsoft Excel [VLOOKUP function in Microsoft Excel], available at: <https://lumpics.ru/examples-the-lookup-function-in-excel/> (accessed 04.04.2020). (In Russ.).

References

1. Instruksiya po deistviyam v razdele «Svedeniya o publikatsiyakh» (obnovlennaya)» [Manual for Using Section «Information on Publications» (Updated)], available at: <http://wnir.minobrnauki.gov.ru/index.php?newsid=67> (accessed 04.04.2020). (In Russ.).
2. Pure. The World's Leading Research Information Management System, available at: <https://www.elsevier.com/solutions/pure> (accessed 15.05.2020). (In Eng.).

Рукопись поступила в редакцию 05.04.2020

Submitted on 05.04.2020

Принята к публикации 26.05.2020

Accepted on 26.05.2020

Информация об авторах / Information about the authors

Резниченко Олег Сергеевич – старший преподаватель кафедры прикладной информатики и информационных технологий, Белгородский государственный национальный исследовательский университет; +7 4722 30-13-00 (*2166); oreznichenko@bsu.edu.ru; ORCID 0000-0003-4657-9428.

Сиваков Станислав Иванович – директор Департамента научной коммуникации и издательской деятельности, Белгородский государственный национальный исследовательский университет; +7 4722 30-14-83; sivakov@bsu.edu.ru; ORCID 0000-0003-4029-865X.

Резниченко Татьяна Алексеевна – кандидат физико-математических наук, доцент кафедры прикладной информатики и информационных технологий, Белгородский государственный национальный исследовательский университет; reznichenko_t@bsu.edu.ru.

Oleg S. Reznichenko – Senior Lecturer, Department of Applied Informatics and Information Technology, Belgorod State University; +7 4722 30-13-00 (*2166); oreznichenko@bsu.edu.ru; ORCID 0000-0003-4657-9428.

Stanislav I. Sivakov – Director of the Department of Scientific Communication and Publishing, Belgorod State University; +7 4722 30-14-83; sivakov@bsu.edu.ru; ORCID 0000-0003-4029-865X.

Tatyana A. Reznichenko – PhD (Physics and Mathematics), Associate Professor, Department of Applied Informatics and Information Technology, Belgorod State University; reznichenko_t@bsu.edu.ru.