

УДК 004.62,004.42

DOI 10.52575/2687-0932-2025-52-1-181-193

Алгоритмы и инструменты обработки данных о публикациях научных организаций, размещенных на портале eLIBRARY.RU

Резниченко О.С.

Белгородский государственный национальный исследовательский университет
Россия, 308015, г. Белгород, ул. Победы, 85
oreznichenko@bsuedu.ru

Аннотация. Научным организациям приходится ежегодно формировать разные виды внешних и внутренних отчетов о публикационной активности, значительная часть которых представляется в виде табличных данных. Большинство из отчетов создается на основе сведений с портала eLIBRARY.RU. В открытом доступе не представлены программные инструменты автоматизированного экспорта из eLIBRARY.RU списка публикаций научной организации в формат электронных таблиц, поэтому данное исследование ставит перед собой цель в разработке таких инструментов. В работе создан алгоритм экспорта и преобразования публикационных данных, использующий процессор электронных таблиц Microsoft Excel и скрипт на языке Python на базе функций класса Pandas.DataFrame. Разработанный алгоритм и программный инструментарий позволят сэкономить время специалистов, анализирующих публикационную активность научных организаций и их подразделений, за счет исключения ручной обработки данных.

Ключевые слова: наукометрия, xml в Excel, парсинг xml, Microsoft Excel, Pandas.DataFrame, Science Index, Python, список публикаций, отчет о публикационной активности

Для цитирования: Резниченко О.С. 2025. Алгоритмы и инструменты обработки данных о публикациях научных организаций, размещенных на портале eLIBRARY.RU. *Экономика. Информатика*, 52(1): 181–193. DOI 10.52575/2687-0932-2025-52-1-181-193

Algorithms and Tools for Processing Data on Scientific Organizations' Articles Uploaded to eLIBRARY.RU

Oleg S. Reznichenko

Belgorod State National Research University
85 Pobeda St, Belgorod 308015, Russia
oreznichenko@bsuedu.ru

Abstract. Scientific organizations frequently generate various types of external and internal publication activity reports. Most of the reports are presented in a tabular dataset format and are based on content from the eLIBRARY.RU portal. After reviewing and analyzing existing software tools for automated export of articles data from eLIBRARY.RU into a spreadsheet format, it was found that there are no ready-made software tools or open-source solutions publicly available that can be easily modified to solve this problem. This research aims to develop such software tools. The paper designs an algorithm for exporting and converting articles' XML metadata using Microsoft Excel and an original Python script. The Python script is created using the Pandas.DataFrame class. This algorithm and the software tools will reduce the time specialists spend analyzing the publication activity of scientific organizations and their departments by eliminating manual data processing.

© Резниченко О.С., 2025



Keywords: scientometrics, xml to Excel, xml parsing, Microsoft Excel, Pandas.DataFrame, Science Index, Python, list of articles, publication activity report

For citation: Reznichenko O.S. 2025. Algorithms and Tools for Processing Data on Scientific Organizations' Articles Uploaded to eLIBRARY.RU. *Economics. Information technologies*, 52(1): 181–193 (in Russian). DOI 10.52575/2687-0932-2025-52-1-181-193

Введение

Крупным государственным научным организациям в течение года приходится предоставлять большой объем отчетности о различных направлениях своей деятельности. Отчетность предоставляется как организацией в целом для внешних контролирующих органов (например, Минорбнауки), так и в форме внутренней отчетности между своими структурными подразделениями. Значительный объем этой отчетности касается публикационной активности организации, её сотрудников и структурных подразделений. Сейчас основной наукометрической базой данных, на которую ориентируется большинство государственных научных организаций, является российская научная электронная библиотека eLIBRARY.RU (далее – elibrary.ru). В Белгородском государственном национальном исследовательском университете (далее – НИУ «БелГУ», университет) сотрудникам отдела научных коммуникаций и публикационной активности приходится формировать тот или иной отчет на основе данных их портала elibrary.ru минимум раз в неделю. Одной из частых задач в рамках предоставления отчетности является задача получения списка публикаций университета в различных разрезах. В качестве исходных данных для анализа публикаций применяется встроенная в elibrary.ru система фильтрации публикаций организаций, включающей набор predefined признаков для фильтра. Среди примеров задач, включающих этапы экспорта и анализа списка публикаций, можно привести следующие задачи, которые регулярно возникают перед специалистами университета:

- 1) создание отчетов о деятельности Научно-образовательных центров мирового уровня;
- 2) формирование внутреннего рейтинга журналов, в которых публикуются сотрудники университета;
- 3) расчет фракционной доли представленности в публикациях структурных подразделений и университета в целом [Гуськов, Косяков, 2020];
- 4) мониторинг количества публикаций у сотрудников и подразделений.

При авторизованном доступе к elibrary.ru на странице со списком публикаций университета в результате применения фильтров представляется до 100 публикаций, включающих заглавие статьи, информацию об источнике и не всегда полный список авторов. Когда требуется обработать менее 100 публикаций, то имеющегося стандартного функционала достаточно для проведения анализа. Например, можно легко преобразовать веб-страницу с результатами поиска после применения фильтра в табличный документ, используя процессор электронных таблиц, например, Microsoft Office Excel (далее – Excel). Чаше возникают задачи, которые подразумевают обработку многостраничных результатов поиска. В таких случаях предлагается постранично сохранять и преобразовывать в табличный формат результаты фильтрации. Вопросы автоматизированной обработки сохраненных страниц со списком публикаций будут рассматриваться в другом исследовании, описывающем алгоритмы парсинга содержимого набора последовательно выгруженных веб-страниц с результатами поиска после применения фильтра.

В университете имеется подписка на использование функционала информационно-аналитической системы Science Index, в рамках которой, в том числе, доступна возможность экспортировать реферативную информацию о публикациях в рамках описанной ранее системы фильтров в формате xml [Научная электронная библиотека, 2024].

Вопросы экспорта и преобразования библиографических записей рассматриваются в работе [Желнов, 2022], но в представленном там программном обеспечении (далее – ПО)

имеется возможность обрабатывать только результаты использования расширенного поиска, в котором информация о публикациях представлена в урезанном виде. Кроме этого, описанные скрипты применимы только в связке с менеджером пользовательских скриптов Tampermonkey в виде расширения для браузера [Biniok, 2025], который надо устанавливать и настраивать дополнительно. Также среди существенных минусов работы можно указать наличие защиты от автоматического онлайн-парсинга, которое имеется на портале elibrary.ru [Попов и др., 2022], и которое, в свою очередь, потребует значительной адаптации предложенного скрипта для офлайн-парсинга содержимого веб-страниц. В нескольких источниках приводятся описания разработок и инструкции по экспорту публикационных данных с последующим их импортом в специализированные профильные библиографические информационные системы, в том числе:

- 1) «ИСТИНА» – Интеллектуальная Система Тематического Исследования НАукометрических данных [НИИ механики МГУ, 2018; ИФЗ РАН, 2018];
- 2) компьютерный обучающий комплекс для обучения, аттестации, переподготовки персонала на базе системы электронного обучения «Гефест» [Маврин, 2020];
- 3) 1С:Первый БИТ [ПервыйБИТ, 2017];
- 4) 1С: Университет ПРОФ [Сахалинский государственный университет, 2024];
- 5) Регистратор «CrossRef» и веб-сайты на базе системы управления содержимым сайта с открытым исходным кодом «WordPress» [Чуриков, 2018].

При этом особенности практической реализации приведенного инструментария в открытом доступе не представлено. Требуемый функционал, вероятно, реализован в исследованиях Кондратьева и Бондякова [2024], а также Шишкина и др. [2021], но их результатами также невозможно воспользоваться, так как они отсутствуют в открытом доступе.

Также в рамках договора Science Index разработчики портала elibrary.ru предоставляют доступ к API для получения только показателей публикационной активности авторов-сотрудников организации без возможности анализировать списки аффилированных публикаций [Научная электронная библиотека, 2024].

Используя имеющиеся в открытых источниках наработки, а также с учетом описанных выше ограничений, можно сформулировать задачу исследования, которая будет состоять в разработке метода и программного инструментария преобразования списка публикаций университета из экспортного формата xml в формат xlsx, в котором этот список был бы представлен в удобном для анализа виде.

В исследовании используется следующий программный инструментарий:

- 1) архиватор, поддерживающий распаковку zip-архивов;
- 2) Microsoft Office Excel версии не ниже версии 12 (2007) [Microsoft, 2025] для первичного преобразования экспортного xml-файла во входной файл для последующей программной обработки;
- 3) интерпретатор Python версии не ниже 3.12 [van Rossum, 2025];
- 4) PyCharm Community Edition версии не ниже 2024.3.1 в качестве интегрированной среды разработки для редактирования кода скриптов и компиляции конечной исполняемой программы [JetBrain, 2025].

Описание структуры входных и выходных данных

Экспортный xml-файл, полученный при использовании встроенного функционала elibrary.ru [Научная электронная библиотека, 2024], представляет собой многомерный массив записей (рис. 1).

В данном исследовании предлагается не производить синтаксический анализ xml-файла напрямую, а предварительно преобразовать его в таблицу, так как проще обрабатывать двумерные наборы данных, чем разбирать многомерную структуру xml-тегов и их значений.

Для преобразования xml-файла в двумерную таблицу можно воспользоваться стандартным функционалом Excel, используя инструкцию, например, приведенную в статье [Microsoft, 2025]. В процессе открытия файла xml Excel предложит создать xml-

схему автоматически. Необходимо отметить, что процесс преобразования в формат Excel может занять длительное время и зависит от размера исходного xml-файла, а также от характеристик гаджета, на котором производится преобразование.

```
1519-6984</issn><eissn>1678-4375</eissn><publisher>International Institute of Ecology
</publisher><country>BRA</country><vak>no</vak><rsci>no</rsci><wos>no</wos><scopus>yes
</scopus><white_list>no</white_list><doaj>yes</doaj></journal><issue><year>2024
</year><volume>84</volume></issue></source><pages>e257071</pages><language>EN
</language><cited>0</cited><titles><title lang="EN">Testing for the ability to modify
antibiotics of Panus tigrinus 8/18 Lentinus strigosus 1566 laccase
</title></titles><doi>10.1590/1519-6984.257071</doi><edn>CNFUJZ</edn><grnti>340000
</grnti><risc>yes</risc><corerisc>yes</corerisc><citation>Testing for the ability to
modify antibiotics of Panus tigrinus 8/18 Lentinus strigosus 1566 laccase / U. A.
Krut, G. M. Shaidorova, A. I. Radchenko [et al.] // Brazilian Journal of Biology. -
2024. - Vol. 84. - P. e257071. - DOI 10.1590/1519-6984.257071. - EDN CNFUJZ.
</citation><authors><author num="1" lang="EN"><lastname>Krut</lastname><initials>U.A.
</initials><authorid>758213</authorid><affiliations><affiliation num="1" lang="EN"
><orgname>Belgorod State University</orgname><orgid>340</orgid><country>rus
</country><town>Belgorod</town></affiliation></affiliations></author><author num="2"
lang="EN"><lastname>Shaidorova</lastname><initials>G.M.</initials><authorid>1005440
</authorid><affiliations><affiliation num="1" lang="EN"><orgname>Belgorod State
University</orgname><orgid>340</orgid><country>rus</country><town>Belgorod
</town></affiliation></affiliations></author><author num="3" lang="EN"><lastname>
Radchenko</lastname><initials>A.I.</initials><authorid>1144088
</authorid><affiliations><affiliation num="1" lang="EN"><orgname>Belgorod State
University</orgname><orgid>340</orgid><country>rus</country><town>Belgorod
</town></affiliation></affiliations></author><author num="4" lang="EN"><lastname>
Kuzubova</lastname><initials>E.V.</initials><authorid>1126845
</authorid><affiliations><affiliation num="1" lang="EN"><orgname>Belgorod State
University</orgname><orgid>340</orgid><country>rus</country><town>Belgorod
</town></affiliation></affiliations></author><author num="5" lang="EN"><lastname>
```

Рис. 1. Содержимое экспортного файла elibrary.ru
Fig. 1. Contents of the elibrary.ru export file

Результирующий файл представляет собой таблицу, в которой каждая запись – это статья, разбитая по количеству соавторов, их аффилиациям, используемым источникам, ключевым словам и некоторым другим полям (рис. 2).

Структура выходного табличного файла определяется потребностями решаемых задач. Предлагается создать избыточную структуру файла, чтобы он максимально перекрывал потенциальные потребности запрашиваемой формы отчетности. Набор полей выходных данных сформирован субъективно относительно круга решаемых сотрудниками отдела научных коммуникаций и публикационной активности задач на протяжении последних 8 лет аналитической деятельности. Этот набор полей может быть легко изменен под нужды конкретной задачи. Состав полей и их описание представлены в табл. 1.

Описание алгоритмов и особенностей реализации скриптов

На рис. 3 представлен алгоритм действий пользователя для формирования выходного табличного файла в формате Excel.

На рис. 4 представлен скриншот с примером настройки параметров фильтрации публикаций университета на портале elibrary.ru (шаги 1–2). Необходимо отметить, что выполнение поисковых запросов по фильтрам на этой странице может занимать значительное время и зависит как от общего количества публикаций научной организации на портале, так и загруженности портала в момент использования фильтров.

Процесс формирования экспортного файла также может занять длительное время. Письмо со ссылкой на сформированный файл придет на почту, привязанную к аккаунту, из-под которого выполняется выгрузка.

| | A | B | C | D | E | F | G | H | I | J | K | L |
|----|----------|---------------------------------------------|--------------|----------------|-----|-------------------------|-----------|-----------|--------------|---------|-----|------|
| 1 | id | linkurl | genre | type | id2 | title | issn | eissn | publisher | country | vak | rsci |
| 2 | 48188165 | https://elibr. | статья в жур | научная статья | | Brazilian Jou 1519-6984 | 1678-4375 | 1678-4375 | Internationa | BRA | no | no |
| 3 | 48188165 | https://elibr. | статья в жур | научная статья | | Brazilian Jou 1519-6984 | 1678-4375 | 1678-4375 | Internationa | BRA | no | no |
| 4 | 48188165 | https://elibr. | статья в жур | научная статья | | Brazilian Jou 1519-6984 | 1678-4375 | 1678-4375 | Internationa | BRA | no | no |
| 5 | 48188165 | https://elibr. | статья в жур | научная статья | | Brazilian Jou 1519-6984 | 1678-4375 | 1678-4375 | Internationa | BRA | no | no |
| 6 | 48188165 | https://elibr. | статья в жур | научная статья | | Brazilian Jou 1519-6984 | 1678-4375 | 1678-4375 | Internationa | BRA | no | no |
| 7 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 8 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 9 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 10 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 11 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 12 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 13 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 14 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 15 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 16 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 17 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 18 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 19 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 20 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 21 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 22 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 23 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 24 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 25 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |
| 26 | 56656587 | https://elibr. | статья в жур | научная статья | | Теория и пр 0040-3601 | | | Научно-изд | RUS | yes | yes |

Рис. 2. Содержимое файла Excel, полученного из экспортного файла xml
 Fig. 2. Contents of Excel file converted from xml-file export

Таблица 1
 Table 1

Структура табличного файла с выходными данными
 Structure of the output table file

| Название полей | Описание назначения полей и их типы |
|----------------|-------------------------------------------------------------------------------------------------------------------------------|
| id | идентификатор публикации на портале elirbary.ru, целое |
| URL | интернет-ссылка на публикацию в elirbary.ru, строка |
| authors | список авторов, разделенный запятыми, строка |
| title | название источника публикации, строка |
| title5 | название публикации, строка |
| year | год публикации, целое |
| imprints | выходные данные публикации, включающие название источника, год публикации, номер тома и выпуска, диапазон страниц в источнике |
| type | тип публикации, строка |
| risc | признак вхождения публикации в РИНЦ, строка |
| corerisc | признак вхождения публикации в Ядро РИНЦ, строка |
| doiб | DOI статьи, строка |
| pages | диапазон страниц публикации в источнике, строка |
| issn | идентификатор ISSN источника публикации, строка |
| eissn | идентификатор eISSN источника публикации, строка |
| vak | признак вхождения публикации в Перечень, рекомендованный ВАК, строка |
| rsci | признак вхождения публикации в RSCI, строка |
| wos | признак индексации публикации в БД Web of Science, строка |
| scopus | признак индексации публикации в БД Scopus, строка |
| white_list | признак вхождения публикации в журнал из «Белого списка», строка |
| number | номер выпуска публикации, строка |
| volume | номер тома выпуска публикации, строка |
| source | наименование источника публикации, строка |
| abstract | аннотация к публикации, строка |



Рис. 3. Общий алгоритм экспорта и преобразования данных из elibrary.ru
 Fig. 3. General algorithm for exporting and converting data from elibrary.ru

Рис. 4. Пример выполнения фильтрации в списке публикаций научной организации
 Fig. 4. An example of filtering a list of articles of a scientific organization

Как видно из инструкции пользователя [Научная электронная библиотека, 2024], экспортный файл приходит на почту в заархивированном виде. Архив нужно скачать, распаковать любым имеющимся архиватором, поддерживающим формат zip (шаги 4–5), а затем открыть вложенный xml-файл при помощи Excel согласно инструкции пользователя [Microsoft, 2025].

Анализ состава полей (шаг 6) промежуточного файла Excel, получившегося после выполнения предыдущего шага необходим, так как некоторые поля меняют своё наименование от выгрузки к выгрузке. На шаге 7 актуальные наименования переменных полей необходимо внести в скрипт конвертации.

Далее более подробно рассматривается структура и особенности создания скрипта конвертации, используемого на шаге 8.

Скрипт использует в качестве входных данных файл “_el_import.xlsx”, который был получен на шаге 5 общего алгоритма, и должен быть размещен в папке расположения скрипта. Структура исходного кода скрипта для конвертации формируется в соответствии с необходимостью последовательно генерировать содержимое столбцов электронной таблицы, согласно структуре файла итоговой электронной таблицы (см. табл. 1). Данные большинства столбцов формируются по алгоритму, представленному на рис. 5.



Рис. 5. Алгоритм генерации полей 'title', 'year', 'id', 'type', 'doiб', 'pages', 'issn', 'number', 'volume', 'source', 'abstract' и других в итоговом наборе данных

Fig. 5. Algorithm for creating columns 'title', 'year', 'id', 'type', 'doiб', 'pages', 'issn', 'number', 'volume', 'source', 'abstract' and others in the final data frame

В лист. 1 приведен пример скрипта реализации алгоритма формирования столбца 'abstract'. Данный скрипт использует встроенные функции класса Pandas.DataFrame (далее – dataframe) [Van den Bossche Joris at al, 2024], и может считаться типовым шаблоном, на основе которого формируются остальные скрипты для генерации содержимого полей 'title', 'year', 'id', 'type', 'risc', 'corerisc', 'doiб', 'pages', 'issn', 'eissn', 'vak', 'rsci', 'wos', 'scopus', 'number', 'volume', 'source', 'abstract'. В результате работы скрипта формируется фрейм данных, содержащий поля 'id' и 'abstract', как результат исключения всех дублей значений в поле 'id', кроме первого (или последнего, в зависимости от столбца)

Листинг 1. Пример генерации фрейма данных с полями 'id' и 'abstract'
Source Code 1. Example of creating a dataframe with columns 'id' and 'abstract'

```
dftmp = elinp.dropna(subset=['abstract'])
dfabs = dftmp[['id', 'abstract']]
dfabs = dfabs.drop_duplicates(subset = ['id'], keep = 'first')
```

Данные полей 'id' и 'linkurl' получаются в виде списков посредством применения функции извлечения уникальных значений столбца данных dataframe.unique()

```
ids = elinp['id'].unique()
urls = elinp['linkurl'].unique()
```

На рис. 6 представлена часть алгоритма генерации поля 'authors', содержащего данные о соавторах публикаций. Особенностью реализации данного алгоритма является необходимость предварительно преобразовать данные поля 'initials' из формата «Имя Отчество» или «И.О.» в формат «И.О.». Результатом выполнения алгоритма является новый столбец 'FIO', в котором имена соавторов записаны в формате «Фамилия И.О.». На основе значений столбца 'FIO' генерируется список 'au_lst', в котором каждый элемент представляет список соавторов каждой из публикаций в формате «Фамилия И.О.», разделенных запятыми (лист. 2).

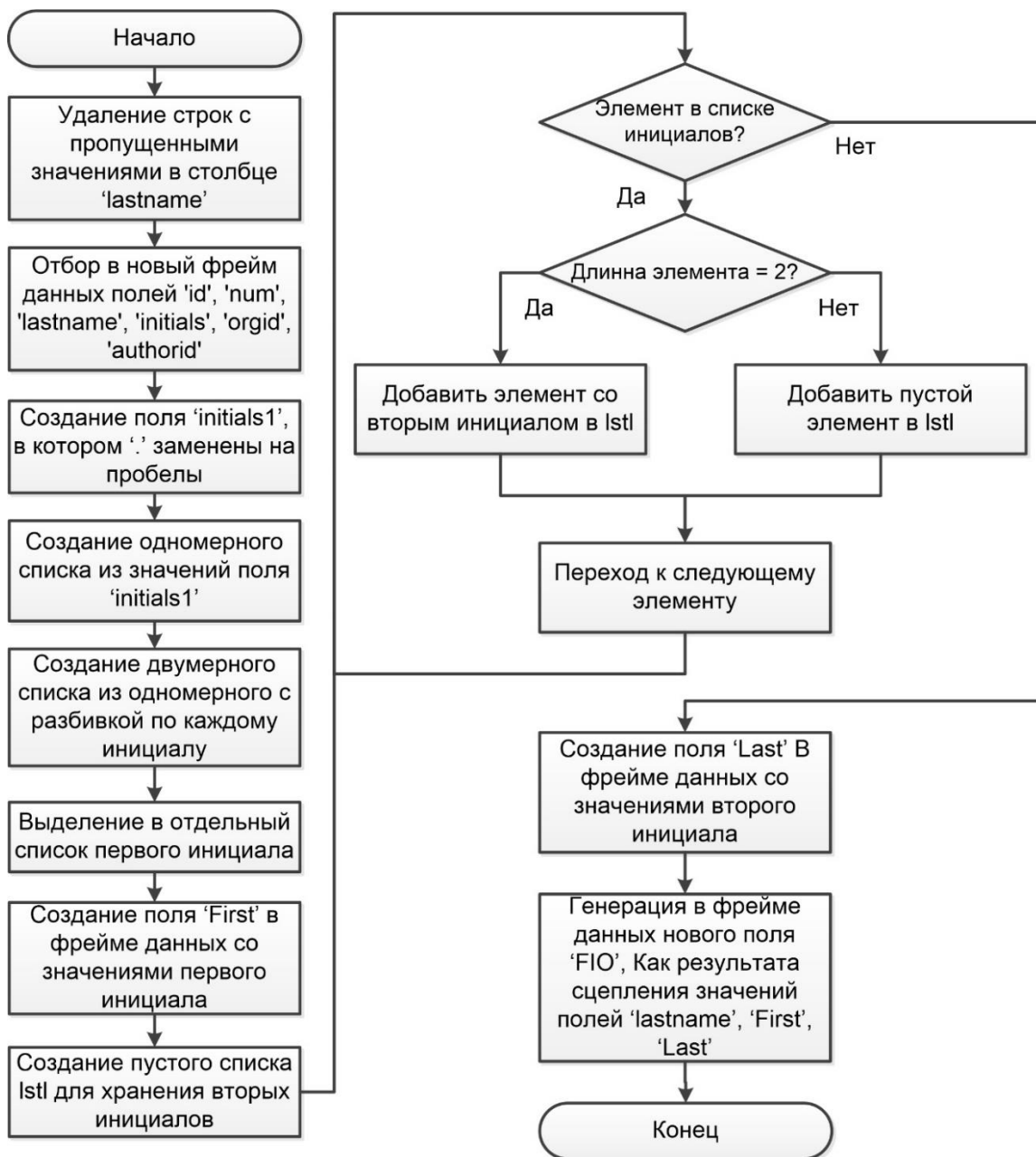


Рис. 6. Генерация значений полей с фамилией и инициалами в формате «Фамилия И.О.»
 Fig. 6. Creating column values with last name and initials in form "Last name Initial1.Initial2."

Листинг 2. Генерация значений поля 'authors' в виде списка 'au_lst'
Source Code 2. Creating 'authors' column values as a list 'au_lst'

```
au_lst = []  
for i in ids:  
    dfnew = elout1[elout1['id'] == i]  
    dfnew = dfnew.drop_duplicates(subset=['FIO'], keep='first')  
    tmp = dfnew['FIO'].values  
    au_lst.append(', '.join(tmp))
```

Зачастую в публикационной отчетности встречаются задачи, связанные с формированием списков публикаций с выходными данными. В выходном файле типа Excel имеется достаточно полей для формирования подобного списка путем сцепления значений нескольких уже имеющихся столбцов. В данной работе предлагается автоматизировать и этот функционал. На рис. 7 представлен алгоритм формирования списка 'vyhodnye', в котором каждый элемент – это строка с выходными данными публикаций, составленный по следующему шаблону:

[Название источника\журнала (source)], [Год (year)]. – [Том (volume)]. –
[Выпуск (number)]. – с. [Страницы (pages)]

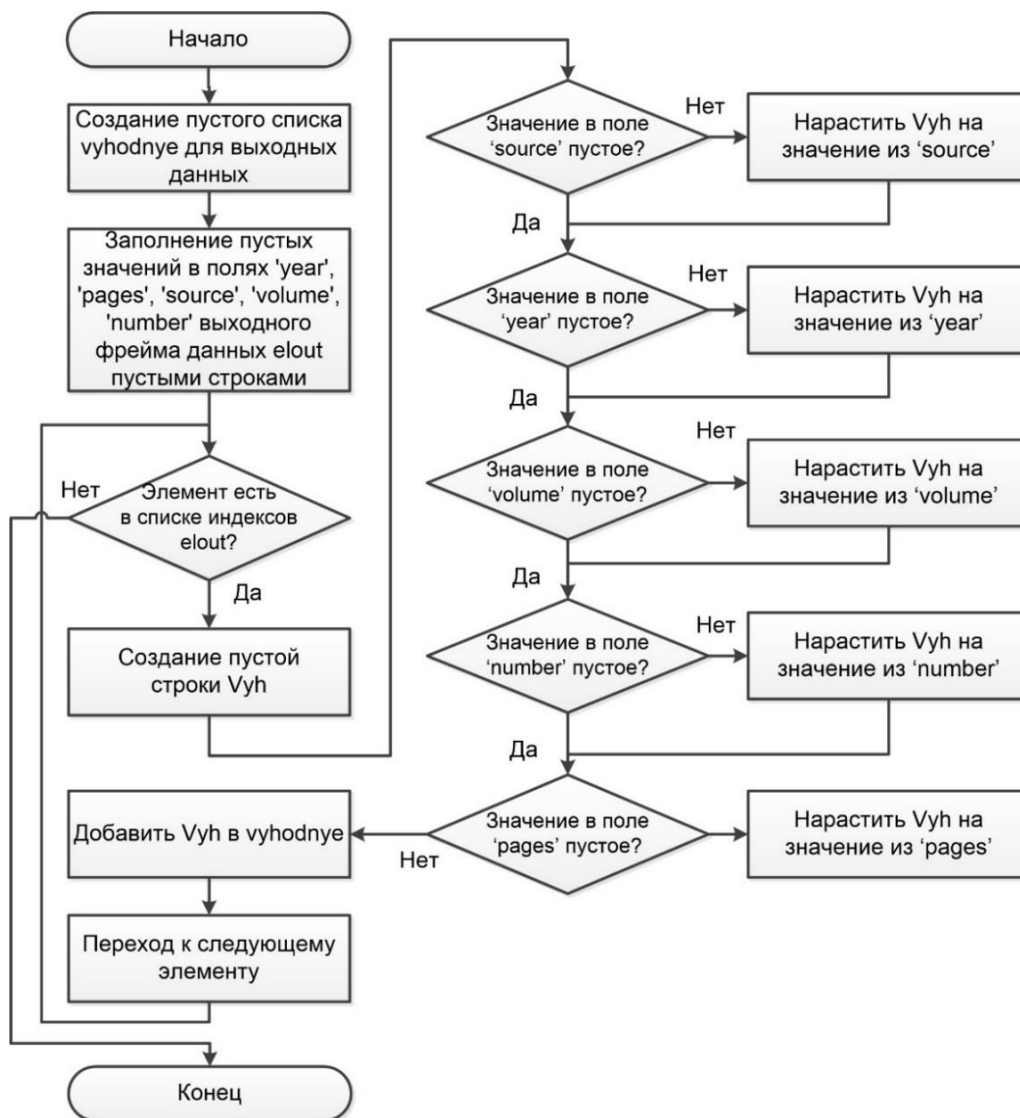


Рис. 7. Алгоритм генерации списка с выходными данными публикаций
Fig. 7. Algorithm for creating a list with articles output data

Последний блок кода посвящен объединению всех сгенерированных наборов данных, полученных на прошлых этапах в один фрейм данных. В список для объединения попадают фреймы данных `dftitles`, `dfyear`, `dfstype_`, `dfpages`, `dfissn_`, `dfsourсe`, `dfabs`. Эти фреймы объединяются в один общий фрейм `elout` при помощи функции `dataframe.merge`, а столбцы 'id', 'URL', 'authors', 'imprints' заполняются данными из списков `ids`, `urls`, `au_lst` и `vyhodnye` соответственно.

После применения описанной последовательности скриптов будет получен файл `_el_export.xlsx`, имеющий пригодный для дальнейшего использования формат и набор данных (рис. 8).

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | |
|----|----------|---------------------------------------------------------------|------------|------------|------|-------------|-----------|------|----------|-----------|---------|-----------|-----------|-----|-----|
| 1 | id | URL | author | title | year | imprint | type | risc | corerist | doi7 | pages | issn | eissn | vak | |
| 2 | 48188165 | https://eli.kit.edu.tr/ | Krut U.A., | Testing fo | 2024 | Brazilian J | научная с | yes | yes | 10.1590/1 | e257071 | 1519-6984 | 1678-4375 | no | no |
| 3 | 56656587 | https://eli.kit.edu.tr/ | Третьяков | Формиро | 2024 | Теория и | научная с | yes | yes | | 106-108 | 0040-3601 | | yes | yes |
| 4 | 56739445 | https://eli.kit.edu.tr/ | Михайлич | ОСОБЕНН | 2024 | Проблемн | научная с | yes | no | | 3-7 | | 1561-691€ | no | no |
| 5 | 56739446 | https://eli.kit.edu.tr/ | Синицына | ТРЕТЕЙСК | 2024 | Проблемн | научная с | yes | no | | 8-11 | | 1561-691€ | no | no |
| 6 | 56739447 | https://eli.kit.edu.tr/ | Сова П.Н. | ПРОБЛЕМ | 2024 | Проблемн | научная с | yes | no | | 12-17 | | 1561-691€ | no | no |
| 7 | 56739448 | https://eli.kit.edu.tr/ | Рассолов | АДМИНИ | 2024 | Проблемн | научная с | yes | no | | 18-23 | | 1561-691€ | no | no |
| 8 | 56739450 | https://eli.kit.edu.tr/ | Голубева | РАЗВИТИЕ | 2024 | Проблемн | научная с | yes | no | | 28-33 | | 1561-691€ | no | no |
| 9 | 56739451 | https://eli.kit.edu.tr/ | Дахова А. | ОБЪЕКТЫ | 2024 | Проблемн | научная с | yes | no | | 34-37 | | 1561-691€ | no | no |
| 10 | 56739452 | https://eli.kit.edu.tr/ | Банченко | ПРОБЛЕМ | 2024 | Проблемн | научная с | yes | no | | 38-40 | | 1561-691€ | no | no |
| 11 | 56739453 | https://eli.kit.edu.tr/ | Михайлич | САЙТ КАК | 2024 | Проблемн | научная с | yes | no | | 41-47 | | 1561-691€ | no | no |
| 12 | 56739454 | https://eli.kit.edu.tr/ | Петрова Е | АКТУАЛЬ | 2024 | Проблемн | научная с | yes | no | | 48-51 | | 1561-691€ | no | no |
| 13 | 56739455 | https://eli.kit.edu.tr/ | Лазовская | СУЩНОСТ | 2024 | Проблемн | научная с | yes | no | | 52-54 | | 1561-691€ | no | no |
| 14 | 56739456 | https://eli.kit.edu.tr/ | Тарасенко | ИСТОРИЯ | 2024 | Проблемн | научная с | yes | no | | 55-60 | | 1561-691€ | no | no |
| 15 | 56739457 | https://eli.kit.edu.tr/ | Димитров | ПРАВОВО | 2024 | Проблемн | научная с | yes | no | | 61-63 | | 1561-691€ | no | no |
| 16 | 56739458 | https://eli.kit.edu.tr/ | Рязанцев | АКТУАЛЬ | 2024 | Проблемн | научная с | yes | no | | 64-68 | | 1561-691€ | no | no |
| 17 | 56739461 | https://eli.kit.edu.tr/ | Тесленко | НАДРУГА | 2024 | Проблемн | научная с | yes | no | | 69-73 | | 1561-691€ | no | no |
| 18 | 56739462 | https://eli.kit.edu.tr/ | Шевкунов | Уголовно | 2024 | Проблемн | научная с | yes | no | | 74-80 | | 1561-691€ | no | no |
| 19 | 56739463 | https://eli.kit.edu.tr/ | Фендрикс | ПОНЯТИЕ | 2024 | Проблемн | научная с | yes | no | | 81-88 | | 1561-691€ | no | no |
| 20 | 56739465 | https://eli.kit.edu.tr/ | Амелина | ПРОБЛЕМ | 2024 | Проблемн | научная с | yes | no | | 89-94 | | 1561-691€ | no | no |
| 21 | 56739467 | https://eli.kit.edu.tr/ | Клавдиев | ПОЛНОМ | 2024 | Проблемн | научная с | yes | no | | 101-106 | | 1561-691€ | no | no |
| 22 | 56739468 | https://eli.kit.edu.tr/ | Чубукова | ПРЕДУПР | 2024 | Проблемн | научная с | yes | no | | 107-112 | | 1561-691€ | no | no |
| 23 | 56739469 | https://eli.kit.edu.tr/ | Турчин О. | ОСОБЕНН | 2024 | Проблемн | научная с | yes | no | | 113-117 | | 1561-691€ | no | no |
| 24 | 56739470 | https://eli.kit.edu.tr/ | Косова А. | ВОЗДЕЙС | 2024 | Проблемн | научная с | yes | no | | 118-121 | | 1561-691€ | no | no |
| 25 | 56739471 | https://eli.kit.edu.tr/ | Трушкова | УГЛУБЛ | 2024 | Проблемн | научная с | yes | no | | 122-126 | | 1561-691€ | no | no |

Рис. 8. Пример содержимого выходного файла `_el_export.xlsx`
 Fig. 8. Example of the contents of the output file `_el_export.xlsx`

Обсуждение результатов и заключение

В ходе проведенного исследования были разработаны алгоритмы и инструменты для преобразования экспортного xml-файла с информацией о публикациях в файл Excel, который может быть использован для формирования нескольких видов публикационной отчетности. Необходимо отметить, что преобразованный в табличный вид на этапе 5 общего алгоритма (см. рис. 3) xml-документ уже можно применять для анализа. Однако стандартными средствами Excel, не включающими создание макросов, в том числе на языке Visual Basic for Application, файл подобного формата пришлось бы обрабатывать длительное время, удаляя дубликаты записей в столбцах 'lastname', 'abstract', 'keyword', 'reference', 'code' и вручную формируя список авторов. Кроме этого, разработанные алгоритмы и сформированные на их основе скрипты можно легко доработать, сформировав отдельный табличный лист, где каждая публикация будет продублирована для каждого автора с его аффилиацией, что позволит быстро выполнить расчет фракционной доли аффилиации научной организации в статье, что в свою очередь может потребоваться при формировании некоторых видов внешних отчетов.

В данном исследовании не решена проблема вариации наименований некоторых столбцов при формировании очередной выгрузки из `elibrary.ru`. Эта проблема также оставляется на доработку потенциальным пользователям разработанного инструментария. В качестве варианта реализации универсального алгоритма, независимого от наименований столбцов в исходных наборах данных, предлагается использовать следующие подходы:

- 1) создание дополнительного конфигурационного файла, из которого скрипт будет считывать названия полей выгрузки с неоднозначными именами;
- 2) реализация и компиляция скрипта в исполняемый файл с поддержкой параметров запуска.

Следует также отметить, что разработанные инструменты не подойдут для обработки xml-выгрузок таких типов публикаций в eLibrary.ru, как книги, монографии и их главы, а также статей из сборников конференций из-за значительного отличия в таких выгрузках наборов и имен полей исходных данных. Однако представленный скрипт легко можно адаптировать для использования после применения других типов фильтров в eLibrary.ru.

В целом можно сказать, что разработанные алгоритмы и инструментарий в виде скрипта на языке Python [Резниченко, 2025] позволят значительно сэкономить время сотрудникам научных организаций, формирующим различного рода публикационную отчетность, а также в других задачах, где подразумевается использование списков публикаций в виде таблиц с различным набором выходных данных, включающих в том числе DOI статей, ISSN журналов, и других свойств научных публикаций.

Список литературы

- Гуськов А.Е., Косяков Д.В. 2020. Национальный фракционный счет и оценка научной результативности организаций. *Научные и технические библиотеки*, 1(9):15–42. DOI:10.33186/1027-3689-2020-9-15-42
- Желнов Павел. 2022. Средство экспорта библиографических записей из eLIBRARY.RU (СЕБЗЕР). GitHub. URL: <https://github.com/plm-ortho/xs-sebzer> (дата обращения: 23.01.2025)
- ИФЗ РАН. 2018. Инструкция по внесению списка публикаций из РИНЦ в систему ИСТИНА. Москва. URL: https://ifz.ru/documents/All/Instrukcija_po_dobavleniju_v_ISTINU_spiska_publicacii_iz_RINC_fi_n_03.pdf (дата обращения: 24.01.2025)
- Кондратьев Андрей, Бондяков Алексей. 2024. Автоматизированный сбор и систематизация публикаций. Открытые системы архитекторам информационных систем, 2024(01). Москва, ООО «Издательство «Открытые системы». URL: <https://www.osp.ru/os/2024/01/13058272> (дата обращения: 24.01.2025)
- Маврин Владислав. 2020. Экспорт из РИНЦ в Гефест. URL: <https://www.youtube.com/watch?v=rImH64dUrqo> (дата обращения: 24.01.2025)
- Научная электронная библиотека. 2024. Интерфейс программирования приложений API. 2024. URL: https://www.elibrary.ru/projects/api/api_info.asp (дата обращения: 24.01.2025)
- Научная электронная библиотека. 2024. Научная электронная библиотека eLIBRARY.RU. URL: <https://www.elibrary.ru/defaultx.asp> (дата обращения: 24.01.2025)
- Научная электронная библиотека. 2024. Science Index. URL: https://www.elibrary.ru/projects/science_index/info_org.pdf#page=24&zoom=100,53,508 (дата обращения: 24.01.2025)
- НИИ механики МГУ. 2018. Импорт данных из системы eLibrary.ru. Москва. URL: <https://istina.msu.ru/help/import/elibrary/> (дата обращения: 24.01.2025)
- ПервыйБИТ. 2017. Импорт публикаций из портала научной электронной библиотеки. URL: <http://www.almazovcentre.ru/wp-content/uploads/Импорт-публикаций-из-РИНЦ-в-БИТ-1.pdf> (дата обращения: 24.01.2025)
- Попов А.Ю., Ремез М.В., Жилина Е.В., Ожиганова М.И. 2022. Парсинг электронных ресурсов. Библиотека selenium или fake useragent? *Информатизация в цифровой экономике*, 3(4):197–210. DOI 10.18334/ide.3.4.115219
- Резниченко О.С. 2025. Приложение к статье «Algorithms and tools for transformation of scientific organizations' articles data which downloading from the eLIBRARY.RU portal». GitHub. URL: https://github.com/leo-phoenix/elibrary_xml_to_xls (дата обращения: 24.01.2025)
- Сахалинский государственный университет. 2024. Базовая инструкция по работе в системе 1С: Университет ПРОФ. Выгрузка из ELibrary. Южно-Сахалинск. URL: http://sakhgu.ru/wp-content/uploads/page/record_121377/2024_10/ELibrary_1С_Университет_ПРОФ.pdf (дата обращения: 24.01.2025)



- Чуриков Д.В. 2018. Конвертация метаданных научных статей из файла XML стандарта РИНЦ для импорта записей в форматах «CrossRef» (DOI) и WordPress. Свидетельство о государственной регистрации программы для ЭВМ № 2018665773 от 11.12.2018. Москва: Федеральная служба по интеллектуальной собственности. URL: <https://www.elibrary.ru/item.asp?edn=aoexus> (дата обращения: 24.01.2025)
- Шишкин Д.Н., Сибилева Н.С., Логунова О.С. 2021. Программа для импорта данных о публикационной активности из наукометрических систем. Свидетельство о государственной регистрации программы для ЭВМ № 2021663719 от 23.08.2021. Москва: Федеральная служба по интеллектуальной собственности. URL: <https://www.elibrary.ru/item.asp?id=46602190> (дата обращения: 24.01.2025)
- Biniok Jan. 2025. Tampermonkey. URL: <https://www.tampermonkey.net> (дата обращения: 24.01.2025)
- JetBrain. 2025. PyCharm: The Python IDE for data and web professionals. URL: <https://www.jetbrains.com/pycharm/> (дата обращения: 24.01.2025)
- Microsoft. 2025. Explore Microsoft 365 for free. URL: <https://www.microsoft.com/en-us/microsoft-365/try> (дата обращения: 24.01.2025)
- Microsoft. 2025. Import XML data. URL: <https://support.microsoft.com/en-us/office/import-xml-data-beca3906-d6c9-4f0d-b911-c736da817fa4> (дата обращения: 24.01.2025)
- Van den Bossche Joris at al. 2024. pandas.DataFrame. URL: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html> (дата обращения: 24.01.2025)
- van Rossum Guido at al. 2025. Python. URL: <https://www.python.org> (дата обращения: 24.01.2025)

References

- Guskov A.E., Kosyakov D.V. 2020. National Fractional Calculations and Evaluating Organization's Science Efficiency. *Scientific and Technical Libraries*, 1(9):15–42. (in Russian). DOI: 10.33186/1027-3689-2020-9-15-42
- Zhelnov Pavel. 2022. Sredstvo e`ksporta bibliograficheskix zapisej iz eLIBRARY.RU (SEBZER) [Tool for exporting bibliographic records from eLIBRARY.RU (SEBZER)]. GitHub. Available at: <https://github.com/p1m-ortho/xs-sebzer> (accessed: 24 January 2025)
- Schmidt Institute of physics of the earth of the Russian Academy of Sciences. 2018. Instrukciya po vneseniyu spiska publikacij iz RINCz v sistemu ISTINA [Manual for entering a list of articles from the Russian Science Citation Index into the “ISTINA” system]. Available at: https://ifz.ru/documents/All/Instrukcija_po_dobavleniju_v_ISTINU_spiska_publicacii_iz_RINC_fi_n_03.pdf (accessed: 24 January 2025)
- Kondrat'ev Andrey, Bondyakov Aleksey. 2024. Avtomatizirovannyi sbor i sistematizatsiya publikatsii [Automated collection and systematization of articles]. Open Systems for Information Systems Architects, 2024(9) Available at: <https://www.osp.ru/os/2024/01/13058272> (accessed: 24 January 2025)
- Mavrin Vladislav. 2020. Ehksport iz RINTS v Gefest [Export from Russian Science Citation Index to "Gefest" system]. Available at: <https://www.youtube.com/watch?v=rImH64dUpqo> (accessed: 24 January 2025)
- Scientific Electronic Library. 2024. Science Index. Interfeis dlya predstavatelya organizatsii [Science Index. Interface for a company representative]. Available at: https://www.elibrary.ru/projects/science_index/info_org.pdf#page=24&zoom=100,53,508 (accessed: 24 January 2025)
- Scientific Electronic Library. 2024. Nauchnaya elektronnyaya biblioteka eLIBRARY.RU [Scientific electronic library eLIBRARY.RU]. Available at: <https://elibrary.ru/defaultx.asp> (accessed: 24 January 2025)
- Scientific Electronic Library. 2024. Interfeys programmirovaniya prilozheniy API [Application Programming Interface API]. Available at: https://www.elibrary.ru/projects/api/api_info.asp (accessed: 24 January 2025)
- Institute of Mechanics. 2018. Import dannykh iz sistemy eLibrary.ru [Importing data from the eLibrary.ru system]. Available at: <https://istina.msu.ru/help/import/elibrary/> (accessed: 23 January 2025)
- Perviy Bit. 2017. Import publikatsii iz portala nauchnoi ehlektronnoi biblioteki [Importing articles from the scientific electronic library portal]. Available at: <http://www.almazovcentre.ru/wp-content/uploads/Импорт-публикаций-из-РИНЦ-в-БИТ-1.pdf> (accessed: 24 January 2025)

- Popov A.Y., Remez M.V., Zhilina E.V., Ozhiganova M.I. 2022. Parsing of electronic resources. Selenium library or fake useragent? *Informatization in the Digital Economy*, 3(4): 197–210. (in Russian) DOI: 10.18334/ide.3.4.115219
- Reznichenko Oleg. 2025. Appendix to article "Algorithms and tools for transformation of scientific organizations' articles data which downloading from the eLIBRARY.RU portal". GitHub. Available at: https://github.com/leo-phoenix/elibrary_xml_to_xls (accessed: 24 January 2025) (in Russian)
- Sakhalin State University. 2024. Bazovaya instruktsiya po rabote v sisteme 1S: Universitet Prof. Vygruzka iz ELibrary [Basic manual for working with "1C: Univertitet Prof" system. Uploading from ELibrary]. Available at: http://sakhgu.ru/wp-content/uploads/page/record_121377/2024_10/ELibrary_1C_Университет_ПРОФ.pdf (accessed: 24 January 2025)
- Churikov D.V. 2018. Konvertatsiya metadannykh nauchnykh statei iz faila XML standarta RINTS dlya importa zapisei v formatakh «CrossRef» (DOI) i WordPress [Converting scientific articles metadata from the XML file of the Russian Science Citation Index standard to import records in the CrossRef (DOI) and WordPress formats]. The Certificate on Official Registration of the Computer Program in Russia. No. 2018665773. Available at: <https://www.elibrary.ru/item.asp?edn=aoexus> (accessed: 24 January 2025)
- Shishkin D.N., Sibileva N.S., Logunova O.S. 2021. Programma dlya importa dannykh o publikatsionnoi aktivnosti iz naukometricheskikh system [Software for importing publication activity data from scientometric systems]. The Certificate on Official Registration of the Computer Program in Russia. No. 2021663719. Available at: <https://www.elibrary.ru/item.asp?id=46602190> (accessed: 24 January 2025)
- Biniok Jan. 2025. Tampermonkey. Available at: <https://www.tampermonkey.net> (accessed: 24 January 2025)
- JetBrain. 2025. PyCharm: The Python IDE for Professional Developers. Available at: <https://www.jetbrains.com/pycharm/> (accessed: 24 January 2025)
- Microsoft. 2025. Explore Microsoft 365 for free. Available at: <https://www.microsoft.com/en-us/microsoft-365/try/> (accessed: 24 January 2025)
- Microsoft. 2025. Import XML data. Available at: <https://support.microsoft.com/en-us/office/import-xml-data-6eca3906-d6c9-4f0d-b911-c736da817fa4> (accessed: 24 January 2025)
- Van den Bossche Joris at al. 2024. pandas.DataFrame. Available at: <https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.html> (accessed: 24 January 2025)
- van Rossum Guido at al. 2025. Python. Available at: <https://www.python.org> (accessed: 23 January 2025)

Конфликт интересов: о потенциальном конфликте интересов не сообщалось.

Conflict of interest: no potential conflict of interest related to this article was reported.

Поступила в редакцию 25.01.2025

Received January 25, 2025

Поступила после рецензирования 02.03.2025

Revised March 02, 2025

Принята к публикации 04.03.2025

Accepted March 04, 2025

ИНФОРМАЦИЯ ОБ АВТОРЕ

INFORMATION ABOUT THE AUTHOR

Резниченко Олег Сергеевич, старший преподаватель кафедры прикладной информатики и информационных технологий, Белгородский государственный национальный исследовательский университет, г. Белгород, Россия

Oleg S. Reznichenko, Senior lecturer of the Department of Applied Informatics and Information Technology, Belgorod State National Research University, Belgorod, Russia