

*И. А. Кшенин<sup>1\*</sup>, Е. В. Шевчук<sup>1</sup>*

## **Автоматизация процесса формирования электронной документации по научно-исследовательской работе кафедры вуза**

<sup>1</sup> Сибирский государственный университет геосистем и технологий, г. Новосибирск,  
Российская Федерация  
e-mail: fen0omen@yandex.ru

**Аннотация.** В данной работе описывается комплекс мероприятий, направленных на создание эффективной системы управления и хранения научно-исследовательских материалов кафедры в электронном формате. Основные цели автоматизации данного процесса включают повышение доступности информации для сотрудников и обучающимися, повышение качества управления научной деятельностью кафедры, а также увеличение эффективности работы кафедры в целом. В статье описаны методы использования современных средств информационных технологий и программных решений для автоматизации документооборота кафедры, в том числе генерацию различных отчетных форм о научных исследованиях и публикационной активности сотрудников кафедры.

**Ключевые слова:** автоматизация бизнес-процессов кафедры, цифровая трансформация образования, электронный документооборот, Elibrary, Python, Selenium Webdriver, PyQt

*I. A. Kshenin<sup>1\*</sup>, E. V. Shevchuk<sup>1</sup>*

## **Automation of the Process of Generating Electronic Documentation for Research Work of a University Department**

<sup>1</sup> Siberian State University of Geosystems and Technologies, Novosibirsk, Russian Federation  
e-mail: fen0omen@yandex.ru

**Abstract.** This paper describes a set of measures aimed at creating an effective system for managing and storing the department's research materials in electronic form. The main goals of automating this process include increasing the availability of information for employees and students, improving the quality of management of scientific departments, as well as increasing the efficiency of departments as a whole. The article describes methods of using modern information technology tools and software solutions to automate the department's document flow, including the generation of various reporting forms on scientific research and publication activity of department employees.

**Keywords:** automation of department business processes, digital transformation of education, electronic document management, Elibrary, Python, Selenium Webdriver, PyQt

### ***Введение***

Указом Президента РФ от 07.05.2024 N 309 «О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года» в качестве целевых показателей национальной цели «Цифровая трансформация государственного и муниципального управления, экономики и социальной сферы» определено достижение в 2030 году «цифровой зрелости» сферы об-

разования, предполагающее в том числе развитие процессов автоматизации, технологий обработки больших объемов данных [1].

Вопросы цифровой трансформации в рамках обеспечения качества образования являются достаточно широко обсуждаемыми [2–5].

В рамках процессов цифровизации образовательные организации все шире развивают системы электронного документооборота, автоматизируя работу с большими объемами данных. Парсинг информации из сети интернет является одной из самых распространенных практик при автоматизации поиска, обработки, а также сохранения данных.

В настоящей статье представлен процесс разработки специального программного обеспечения для парсинга и дальнейшей работы с данными портала eLibrary.

Портал eLibrary.Ru [6] – это сетевая библиотека, которая интегрирована с РИНЦ (Российский индекс научного цитирования [7]). На сегодняшний день в базе eLibrary содержится информация о более чем 50 миллионах научных публикаций. Создан данный портал в конце 1999 года, в целях создания платформы для доступа российским исследователям к зарубежным научным публикациям. В начале 2005 года портал работает и с публикациями на русском языке, и на сегодняшний день является ведущей библиотекой с электронным доступом в русскоязычном сегменте.

На портале eLibrary имеется возможность поиска авторов и публикаций, в том числе и с применением фильтров:

- фильтрация по фамилии автора (причем, можно указать как полное ФИО, так и только фамилию, или фамилию и имя/отчество);
- фильтрация по городу/стране;
- добавление фильтра по организации, которая находится в нормативном списке организаций РИНЦ.

Как правило, этих фильтров достаточно, чтобы найти нужного автора и его публикации.

Сама работа с порталом происходит через веб-браузер пользователя (взаимодействие с веб-интерфейсом): для простого просмотра и поиска достаточно не авторизовываться, но, чтобы найти и просмотреть, например, файлы PDF, пользователь должен пройти регистрацию и войти в систему.

### ***Исследование предметной области***

Чтобы найти публикации автора, нужно нажать на ссылку, которой является отображаемое количество публикаций. Соответственно, парсер должен иметь возможность находить этот элемент на странице, а также открывать ссылку с публикациями. Рис. 1 демонстрирует страницу с найденными публикациями автора.

Рис. 1. Найденные публикации автора

На странице находится список публикаций автора, без фильтрации по параметрам из формы «Параметры» (отмечен «1»). Каждая публикация (отмечена «2») имеет такие элементы:

- порядковый номер публикации;
- ссылка на скачивание PDF файла (опционально, ее может не быть, и это нужно учитывать);
- название публикации;
- автор (или несколько авторов);
- издание, в котором была размещена публикация;
- год;
- страницы, на которых находится данная публикация.

Каждый из перечисленных выше элементов должен быть обработан программой (парсером): на выходе, в файле .xlsx в каждой ячейке должны быть элементы из каждой публикации (каждая на своей строке).

### Результаты

Для описания возможных вариантов работы с разрабатываемой программой использован такой инструмент, как use-case диаграмма.

На рис. 2 показана данная диаграмма.

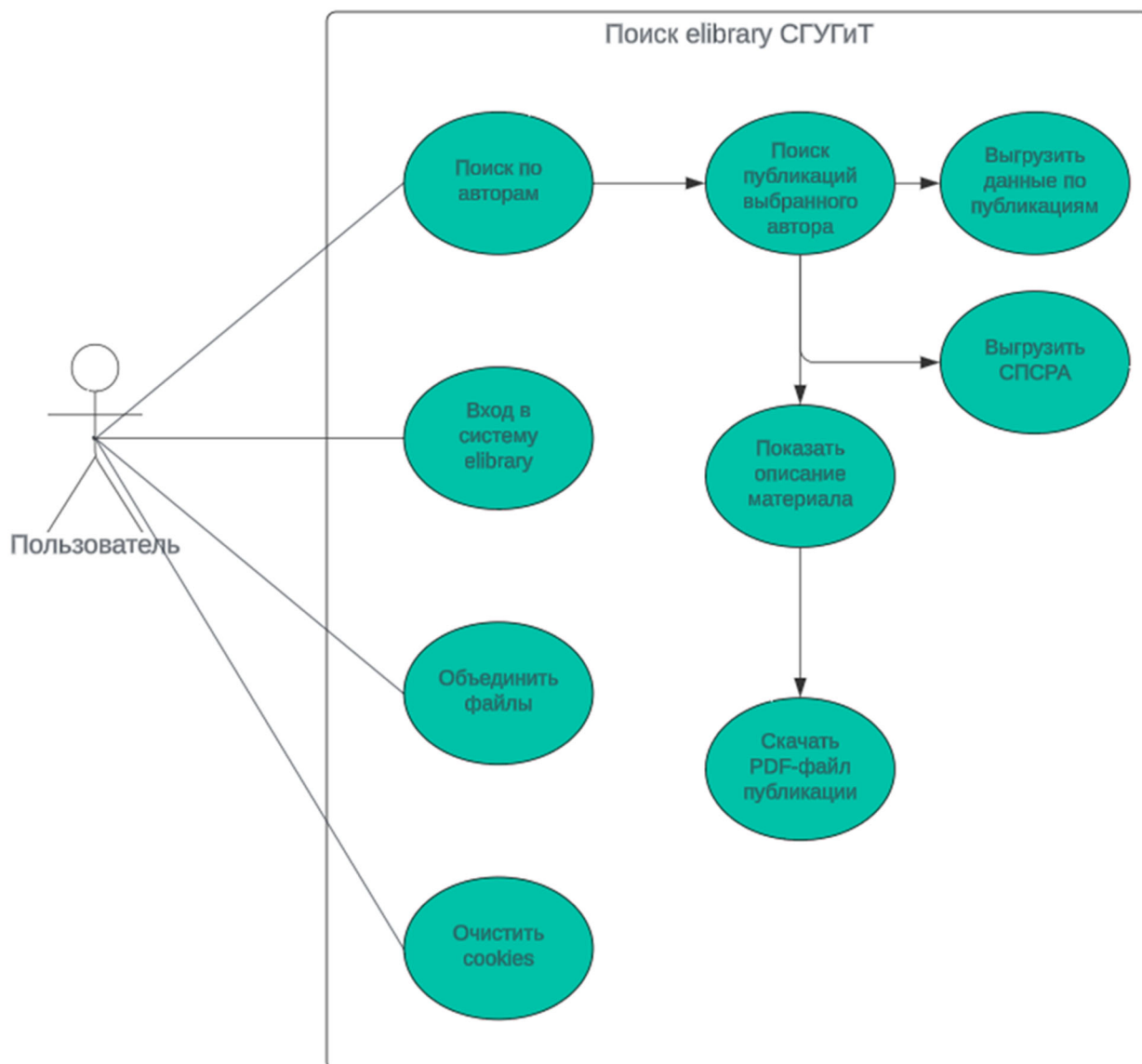


Рис. 2. Use-case диаграмма взаимодействия пользователя с разрабатываемым программным обеспечением

Всего существует 4 глобальных варианта использования программы:

- поиск по авторам (по введенной фамилии);
- вход в систему eLibrary;
- объединение созданных ранее Excel-файлов;
- очистка cookie.

Причем для первого пункта также возможны варианты использования:

- поиск публикаций выбранного автора (на этом этапе можно произвести выгрузку данных по публикациям, либо выгрузить список публикаций, ссылающихся на работы автора);
- просмотреть описание публикации;
- просмотреть/скачать PDF-файл публикации.

При разработке программного обеспечения будут заложены эти сценарии работы с пользователем (нужно подчеркнуть, что это самые основные варианты работы программы, и поэтому внутренняя работа каждого из вариантов здесь не показана, а будет описана на каждом этапе разработки программного обеспечения).

Поскольку требовалось, чтобы программное обеспечение имело удобный интерфейс для взаимодействия с пользователем, было решено для разработки использовать графический интерфейс.

Языком программирования, на котором написан модуль программы, был выбран Python.

Среда разработки, в которой проводилось написание, тестирование и отладка кода – PyCharm Community.

В качестве фреймворка, на котором реализована графическая часть программы, был выбран PyQt [8]. Были использованы такие инструменты данного фреймворка:

- утилита для создания прототипа GUI-интерфейса программы Qt Designer;
- инструменты для работы с потоками QThread и QRunnable [9];
- инструмент для просмотра веб-страниц QWebEngineView.

Для работы с парсингом сайта и доступа к html-элементам был выбран инструмент Selenium Webdriver [10]. Он позволяет автоматизировать работу с движком веб-браузера, а также находить нужные элементы на странице сайта. Плюсом также является то, что сам движок встроен в библиотеку Selenium и этот движок не нужно дополнительно скачивать и конфигурировать его в коде программы.

Выгрузка данных была реализована с помощью библиотеки XlsxWriter. Главными достоинствами этой библиотеки являются гибкая работа с Excel-файлами, максимально «глубокая» работа с Excel (вплоть до создания формул, графиков и т. п.).

Также было решено использовать ряд других библиотек для корректной работы модуля программы (pickle для создания и загрузки cookie-файлов, web browser для работы с PDF-файлами и открытия их в браузере пользователя, datetime для формирования уникального имени выгруженных файлов, библиотека re для работы с регулярными выражениями).

### ***Тестирование программного продукта***

Было проведено ручное тестирование разработанного программного обеспечения. Тестировались следующие части программного обеспечения: авторизация в системе eLibrary, поиск по базе авторов, поиск по выбранному автору его публикаций, просмотр публикации автора, просмотр PDF-файла публикации. Также были проведены тесты выгрузки (для списка публикаций, загруженного в окне программы и для списка публикаций, ссылающихся на работы автора).

1. Успешно авторизовавшись в системе eLibrary, в окне программы появилось соответствующее сообщение.

2. Вход в систему необходим, так как иначе eLibrary не позволяет работать со многими функциями, в частности:

- просмотр PDF-файлов публикации,
- просмотр списка публикаций, ссылающихся на работы автора,

Вход в систему – главный этап, после которого можно переходить к тестированию работы других частей программы.

3. Для тестирования поиска автора достаточно ввести его фамилию, нажать на кнопку «Поиск авторов» и просмотреть загруженный в окно программы список.

Здесь мы видим ФИО авторов и относящийся к ним ВУЗ. Из списка авторов возможно выбрать только одного, и уже по нему производить поиск и работу с публикациями (при выборе более одного автора программа выдаст предупреждение и не позволит продолжить работу).

На рис. 3 показано окно, в котором произведен поиск публикаций по выбранному автору (для этого был выбран определенный автор и нажата кнопка «Найти публикации выбранного автора»).

4. Тестирование работы с публикациями похоже на логику работы с авторами: можно выбрать только одну публикацию и по ней просмотреть описание.

Для того, чтобы просмотреть описание была выбрана третья публикация из списка и нажата кнопка «Показать описание материала».

Описание публикации показывается прямо в окне разработанной программы. При необходимости его можно скопировать.

После того, как был протестирован просмотр описания публикации можно провести тестирование открытия PDF-файла у данной публикации. Для этого нужно нажать на кнопку «Просмотр/скачивание PDF-файла».

На рис. 3 показаны совмещенные скриншоты сигнализирования программы о том, что PDF-файл был открыт, а в нижней части – сам открытый файл в браузере пользователя (браузер открывается автоматически).

В случае, если ссылка на скачивание PDF-файла будет недоступна, либо файл будет отсутствовать в принципе, то программа не будет открывать браузер, а вместо этого просто выдаст соответствующее сообщение.

Теперь приведем описание тестирования выгрузки тех публикаций, которые были ранее найдены и загружены в окно программы. Нажатием на кнопку «Выгрузить данные по публикациям» начнется процесс выгрузки. Результат успешной выгрузки, а также сгенерированный Excel-файл показаны на рис. 4.

На рис. 4, в верхней части показано информирование пользователя о том, что выгрузка прошла успешно, а в нижней части показан фрагмент выгруженного файла.

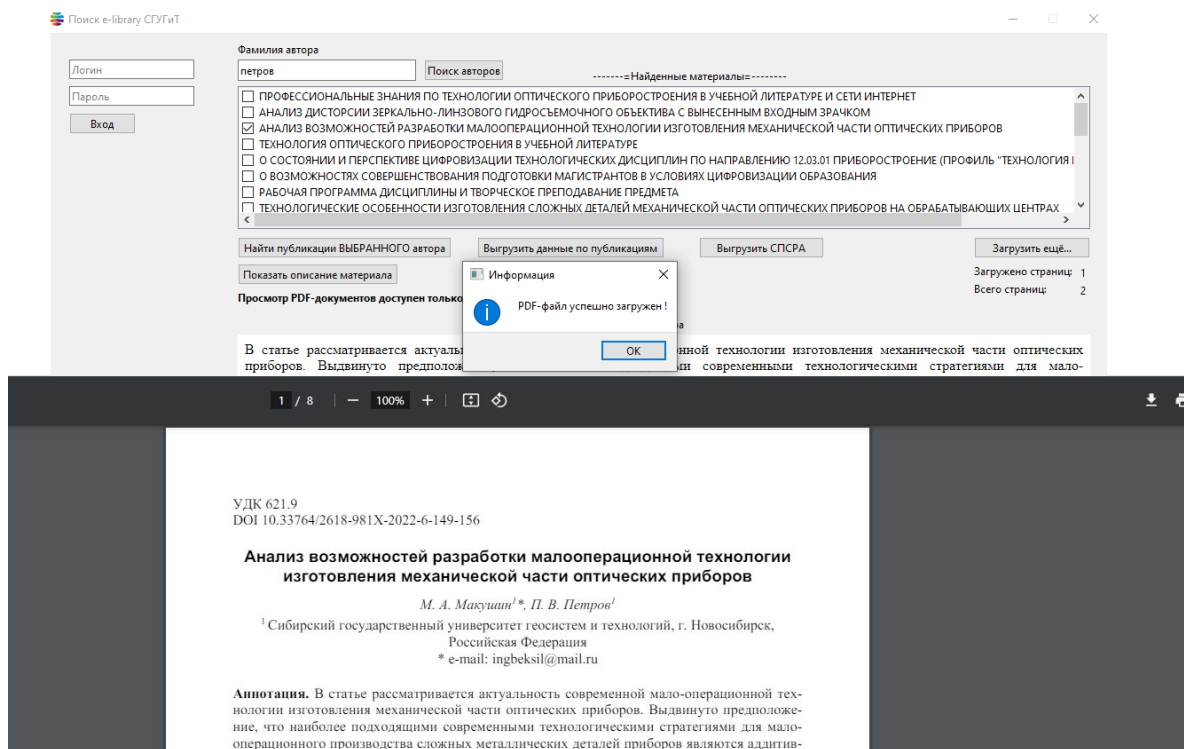


Рис. 3. Сообщение об успешном открытии PDF-файла и сам открытый файл в браузере пользователя

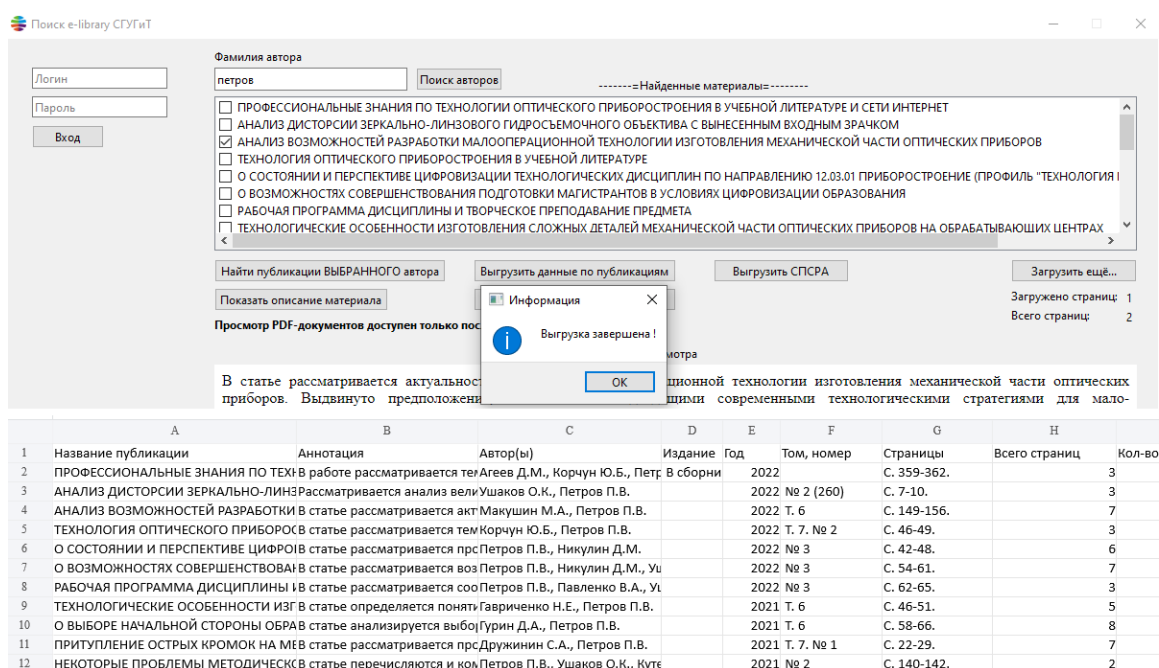


Рис. 4. Тестирование выгрузки публикаций автора

### Заключение

В рамках представленного исследования был описан процесс разработки приложения для автоматизации процесса формирования электронной докумен-

тации, которое сотрудниками кафедры может использоваться в различных целях, таких как, составление библиографии, проведение обзоров литературы, генерация различных видов отчетности по публикационной активности сотрудников.

### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Российская Федерация. Президент. О национальных целях развития Российской Федерации на период до 2030 года и на перспективу до 2036 года: Указ Президента Российской Федерации от 07.05.2024 N 309. – Текст : электронный // Техэксперт. – URL: <http://metrolog.kodeks.ru/docs02/>. – Режим доступа: по подписке (дата обращения: 14.05.2024).

2. Шевчук, Е. В. Цифровая трансформация управления качеством образовательных бизнес-процессов / Е. В. Шевчук, А. В. Шпак // Вестник Российского университета дружбы народов. Серия: Информатизация образования. – 2023. – Т. 20, № 2. – С. 159-175. – DOI 10.22363/2312-8631-2023-20-2-159-175. – EDN JQUMLD.

3. Григоренко, О. В. Независимая оценка качества образования: опыт и перспективы / О. В. Григоренко, А. С. Рванова, А. А. Таджигитов // Актуальные вопросы образования. – 2018. – № 1. – С. 34-38. – EDN XWDPKX.

4. Петрова, М. А. Цифровые компетенции: цифровая дидактика / М. А. Петрова // Актуальные вопросы образования. – 2023. – № 3. – С. 116-123. – EDN JHIEXP.

5. Логачева, О. М. О некоторых примерах внедрения цифровых технологий в преподавании математических дисциплин / О. М. Логачева, А. В. Логачев // Актуальные вопросы образования. – 2022. – № 3. – С. 81-86. – EDN CUVRED.

6. О проекте eLIBRARY.RU [Электронный ресурс]. URL: [https://www.elibrary.ru/elibrary\\_about.asp/](https://www.elibrary.ru/elibrary_about.asp/) (дата обращения: 19.04.2024).

7. Российский индекс научного цитирования. – URL: [https://ru.wikipedia.org/wiki/Российский\\_индекс\\_научного\\_цитирования/](https://ru.wikipedia.org/wiki/Российский_индекс_научного_цитирования/) (дата обращения: 19.04.2024).

8. Тарланов, А. Т. Знакомство с библиотекой PyQT : учебно-методическое пособие / А. Т. Тарланов, Е. С. Карбова. — Москва : РТУ МИРЭА, 2021. — 81 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/176525> (дата обращения: 21.04.2024). — Режим доступа: для авториз. пользователей.

9. QThreadPool — Qt for Python. – URL: <https://doc.qt.io/qtforpython-5/Py-Side2/QtCore/QThreadPool.html> (дата обращения: 19.04.2024).

10. Web Scraping Using Selenium And Python. – URL: <https://www.scrapingbee.com/blog/selenium-python/> (дата обращения: 19.04.2024).

© И. А. Кушенин, Е. В. Шевчук, 2024