

DOI: 10.32517/0234-0453-2024-39-2-78-91

АВТОМАТИЗАЦИЯ ПРОЦЕССОВ НОРМОКОНТРОЛЯ ОТЧЕТНОЙ ДОКУМЕНТАЦИИ СТУДЕНТОВ: КОНЦЕПЦИЯ ПРОГРАММНОГО РЕШЕНИЯ

И. С. Шахова¹ ✉¹ Казанский федеральный университет, г. Казань, Республика Татарстан, Россия

✉ is@it.kfu.ru

Аннотация

Отчетная документация является одним из оценочных средств, используемых для проверки освоения студентами ряда универсальных, общепрофессиональных и профессиональных компетенций, отражающих способность осуществлять поиск, критический анализ и синтез информации, а также умение оформлять методические материалы и разрабатывать технические документы, связанные с профессиональной деятельностью. Освоение соответствующих компетенций оценивается по результатам анализа содержательной части текста отчета и контроля соответствия документов требованиям к оформлению по набору стандартных проверок. Подобный нормоконтроль не предполагает вариативности в оценивании, а значит, имеет перспективы автоматизации. Внедрение программного решения, направленного на автоматизацию процессов нормоконтроля, позволит сократить временные затраты на проверку отчетной документации и формирование рекомендаций для студентов по исправлению обнаруженных несоответствий.

В статье проанализированы наиболее часто встречающиеся ошибки оформления отчетной документации студентов по результатам прохождения практики, проведена их классификация и предложены методы обнаружения каждого из типов несоответствий требованиям к оформлению документов. Концепция программного решения автоматизации нормоконтроля отчетной документации, которая предлагает реализацию предложенных методов и интерфейсы взаимодействия с пользователями, создана с использованием принципа модульности. Результаты исследования, представленного в статье, демонстрируют потенциал применения методов автоматизации для задач проверки отчетной документации студентов.

Ключевые слова: нормоконтроль, проверка оформления, электронный текстовый документ, отчетная документация студента, автоматизация, программное решение.

Для цитирования:

Шахова И. С. Автоматизация процессов нормоконтроля отчетной документации студентов: концепция программного решения. *Информатика и образование*. 2024;39(2):78–91. DOI: 10.32517/0234-0453-2024-39-2-78-91.

THE AUTOMATION OF THE PROCESSES OF STUDENTS' REPORTING DOCUMENTATION STANDARD CONTROL: THE SOFTWARE SOLUTION CONCEPT

I. S. Shakhova¹ ✉¹ Kazan Federal University, Kazan, The Republic of Tatarstan, Russia

✉ is@it.kfu.ru

Abstract

Reporting documentation is one of the assessment tools used to verify the students' mastery of several universal, general professional, and vocational competencies, reflecting the ability to search, critically analyze, and synthesize information, as well as the ability to design methodological materials and develop technical documents related to professional activities. Mastering the relevant competencies can be assessed by the results of analyzing the report text and controlling the compliance of documents with requirements to the design, accompanied by a set of standard checks. Such standard control does not imply any variability in the assessment and therefore has prospects for automation. The introduction of a software solution aimed at automating the processes of standard control will reduce the time spent on checking the reporting documentation and the formation of recommendations for students to correct the detected inconsistencies.

The article analyzes the most common errors in the design of students' reporting documentation on the results of internships, compiles a classification of errors, and suggests methods for detecting each type of non-compliance with the requirements to the design of documents. The concept of the software solution for automation of standard reporting documentation control, which offers a realization of the proposed methods and creates interfaces for interaction with users, is created using the principle of modularity.

© Шахова И. С., 2024

The results of the research presented in the article demonstrate the potential of applying automation methods to the tasks of checking students' reporting documents.

Keywords: standard control, formatting check, electronic text document, student's reporting documentation, automation, software solution.

For citation:

Shakhova I. S. The automation of the processes of students' reporting documentation standard control: The software solution concept. *Informatics and Education*. 2024;39(2):78–91. (In Russian.) DOI: 10.32517/0234-0453-2024-39-2-78-91.

1. Введение

Отчетные документы обучающихся высших учебных заведений используются в качестве средств оценки уровня освоения компетенций и демонстрируют способности студентов проводить анализ и синтез информации, разрабатывать методические материалы и документацию, связанную с профессиональной деятельностью. Одним из примеров отчетной документации, наиболее часто используемой в качестве оценочного средства, является набор отчетных документов по практике.

Согласно федеральному государственному образовательному стандарту высшего образования по направлению подготовки 09.03.04 «Программная инженерия»¹ объем блока «Практика» должен составлять не менее 20 зачетных единиц. Количество практик, их типы и объем практик каждого типа устанавливает организация, реализующая образовательную программу. Так, на момент проведения исследования образовательные программы по направлению подготовки 09.03.04 «Программная инженерия», реализуемые в высших учебных заведениях из списка ста лучших вузов России по данным рейтингового агентства RAEX («РАЭК-Аналитика») за 2023 год², в среднем включают в себя по 6 практик. Требования к отчетности о прохождении практики регламентируются локальными нормативными актами учебных заведений, а потому перечень документов может варьироваться. Однако основным документом студента, отражающим выполненную на практике работу и приобретенные компетенции, является отчет по практике. Таким образом, по результатам завершения каждой практики обучающийся предоставляет руководителю практики от образовательной организации отчет и сопроводительные документы, которые оцениваются руководителем согласно критериям, указанным в программе соответствующей практики.

Процесс оценивания включает в себя анализ содержательной части отчета и проверку соответствия предоставленного документа требованиям к оформлению. В отличие от рассмотрения текста с точки

зрения полноты и корректности его содержания, **контроль оформления (нормоконтроль) состоит из набора стандартных проверок и не предполагает вариативности в оценивании, что позволяет говорить о перспективах автоматизации данного процесса.**

Автоматизация нормоконтроля является актуальной темой исследований и разработок в российском научном сообществе. За последние пять лет в научных изданиях, включенных в Российский индекс научного цитирования, было опубликовано не менее 30 статей, посвященных автоматизации проверки текстовых документов на соответствие требованиям к оформлению³. В основной части статей, посвященных применению автоматических методов для задач нормоконтроля (около 75 %), авторы исследуют процессы автоматизации нормоконтроля студенческих документов в учебных заведениях, что позволяет делать выводы о значимости рассматриваемых проблем, сопровождающих процесс проверки соответствия отчетной документации студентов требованиям к оформлению. В работах исследователей представлены описания программных средств автоматизации проведения процедуры нормоконтроля выпускных квалификационных работ, научно-исследовательских и учебно-практических работ студентов, а также текстовых документов произвольного типа. Предложенные авторами программные средства предоставляют возможность проверки входного текстового файла на соответствие заданным требованиям к оформлению.

Г. А. Нестеров, А. О. Шестакова, Ю. Е. Карякин в статье «Автоматизация проверки научно-технической документации» [1] представляют прототип клиент-серверного приложения для автоматизированной проверки оформления научно-технической документации. В этом приложении реализована проверка научно-исследовательских работ в формате Word-документа на предмет соблюдения отдельных требований к оформлению (нумерации страниц, цвета текстовых абзацев, межстрочного интервала, верхнего и нижнего интервалов у определенных абзацев), а проверяемые свойства хранятся в коде программы. В. Е. Зобнина и Н. А. Вихорь в статье «Автоматизация проверки форматирования текстового документа» [2] предлагают анализировать документы Word с помощью макроса, созданного с использованием языка Visual Basic for Application (VBA), интегрированного в пакет Microsoft Office.

¹ Приказ Министерства образования и науки Российской Федерации от 19 сентября 2017 года № 920 (ред. от 08.02.2021) «Об утверждении федерального государственного образовательного стандарта высшего образования — бакалавриат по направлению подготовки 09.03.04 «Программная инженерия»». https://fgosvo.ru/uploadfiles/FGOS%20VO%203++/Bak/090304_B_3_15062021.pdf

² Рейтинг лучших вузов России RAEX-100, 2023 год. RAEX. Rating Review. 07.06.2023. https://raex-rr.com/education/russian_universities/top-100_universities/2023

³ Эти публикации можно найти, оформив поисковый запрос на сайте eLIBRARY (<https://elibrary.ru>): ключевое слово «нормоконтроль», сортировка по теме «Народное образование. Педагогика».

Разработанная авторами программа анализирует установленные в документе значения полей страницы, наименования и размера шрифта, начертания и выравнивания символов. Однако **В. Н. Поздин и Е. А. Выймова** считают, что метод на основе языка VBA не может быть использован как основной, поскольку макрос работает с документом автономно, а, следовательно, такой подход не позволяет формировать общую статистику. В статье **«Алгоритм автоматизированного нормоконтроля работ обучающихся образовательного учреждения»** [3] они предлагают конвертировать документы в формат HTML и использовать метод с применением парсинга на основе регулярных выражений, где правила проверки задаются при помощи кастомизированных шаблонов. Аналогичный подход к настройке правил проверки предлагают **Б. Е. Стариченко и М. А. Устинов** в статье **«Программа автоматизации контроля оформления текстовых документов»** [4]. В своей работе авторы представляют приложение MAPDoc, алгоритмы проверки которого реализованы путем последовательного сопоставления элементов форматирования каждого абзаца проверяемого документа с заранее заданным эталоном. Авторы из Национального исследовательского университета «Высшая школа экономики» предлагают иной способ задания правил проверки: он основан на использовании элементов пользовательского интерфейса разработанных авторами программных инструментов [5, 6]. Все указанные научные работы ориентированы на обработку документов в редактируемом формате.

Публикации научной группы Национального исследовательского университета ИТМО в области автоматизации нормоконтроля текстовых документов посвящены анализу возможностей парсинга электронных текстовых документов [7], разработке алгоритма формализации требований к оформлению документа [8], построению объектной модели эталонного текстового документа [9], созданию методов [10] и инструментов [11] обработки файлов разных форматов (ODT, DOCX и PDF), что позволяет расширить область применения предложенных решений.

В качестве целевого пользователя в отмеченных ранее работах рассматривается проверяющий, которым в зависимости от особенностей организации процессов нормоконтроля может являться как студент, так и преподаватель или сотрудник кафедры. Авторы из Томского государственного университета систем управления и радиоэлектроники в работах [12, 13] предлагают расширить список пользователей информационной системы, позволяющей автоматизировать процесс нормоконтроля учебной документации, включив в него представителей учебного управления, сотрудников кафедры, преподавателей и студентов. Автоматизации процессов нормоконтроля с точки зрения взаимодействия участников этого процесса посвящена статья **А. Н. Сливина и В. А. Шакуры «Информационная система автоматизированного контроля оформления выпускных квалификационных работ»** [14]. Авторы разработали информа-

ционную систему, направленную на ускорение проверки выпускных квалификационных работ путем систематизации типовых ошибок и предоставления проверяющему возможности выбрать обнаруженное в работе несоответствие требованиям к оформлению из заранее заданного списка с последующей отправкой отчета об обнаруженных ошибках оформления студенту.

Основным объектом исследований зарубежных авторов в области обработки документов является анализ структуры PDF-документа. Предложены размеченные наборы документов для моделей машинного обучения (см., например, [15]) и многочисленные методы семантической сегментации структурных компонентов текстовых документов в формате PDF:

- на основе механизмов анализа пикселей изображений [16];
- на базе нейросетевых подходов [17], в том числе с использованием полносвязных нейронных сетей [18] и нейронных сетей на основе архитектуры U-Net [19];
- на основе комплексного применения технологий обработки естественного языка и компьютерного зрения [20].

Помимо публикаций по теме исследования, имеются зарегистрированные в Роспатенте программные средства, направленные на решение обозначенной проблемы. Они ориентированы на автоматизацию нормоконтроля выпускных квалификационных работ¹ и курсовых проектов², а также предоставляют

¹ Свидетельство о государственной регистрации программы для ЭВМ № 2021680767 Российская Федерация. Модуль анализа документов для программного комплекса автоматизированной проверки документов на соответствие требованиям к оформлению и структуре «Нормоконтроль — просто»: № 2021680148: заявлено 06.12.2021: опубликовано 14.12.2021 / Булатицкий Д. И., Питикин А. А.; правообладатель ФГБОУ ВО «Брянский государственный технический университет». Зарегистрировано в реестре программ для ЭВМ. EDN: BWYLEO. Режим доступа: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2021680767&TypeFile=html; Свидетельство о государственной регистрации программы для ЭВМ № 2023681890 Российская Федерация. Программный комплекс организации и проверки нормоконтроля выпускных квалификационных работ: № 2023616083: заявлено 30.03.2023: опубликовано 20.04.2023 / Ткачев А. В., Логачев М. С., Никишина И. Н.; правообладатель Ткачев А. В., Логачев М. С., Никишина И. Н. Зарегистрировано в реестре программ для ЭВМ. EDN: BLQMTV. Режим доступа: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2023618190&TypeFile=html; Свидетельство о государственной регистрации программы для ЭВМ № 2023681109 Российская Федерация. Система автоматического нормоконтроля текстовых документов: № 2023669730: заявлено 26.09.2023: опубликовано 10.10.2023 / Вагарина Н. С., Семенова Е. А.; правообладатель ФГБОУ ВО «Саратовский государственный технический университет имени Гагарина Ю. А.». Зарегистрировано в реестре программ для ЭВМ. EDN: DDNEGO. Режим доступа: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2023681109&TypeFile=html

² Свидетельство о государственной регистрации программы для ЭВМ № 2015615893 Российская Федерация. Нормоконтроль студенческих работ: № 2015612356: заявлено

возможность проверки соответствия произвольного текстового файла шаблонному документу¹.

Результаты анализа существующих методов и программных средств автоматизации процессов нормоконтроля отчетной документации студентов продемонстрировали значительный задел в исследуемой области и обнаружили направление потенциального развития — обеспечение возможности адаптации программного решения к особенностям образовательного процесса с точки зрения взаимодействия участников, а также типов документов и требований. **Целью статьи является** представление концепции и методов реализации программного решения по автоматизации нормоконтроля отчетной документации студентов, предусматривающего возможность масштабирования и модификации с учетом особенностей процесса, в котором предполагается его использование.

2. Анализ ошибок оформления документов

Для проведения анализа отчетной документации на предмет наиболее часто допускаемых ошибок оформления были выбраны исходные варианты 320 комплектов отчетных документов, подготовленных студентами Института информационных технологий и интеллектуальных систем Казанского (Приволжского) федерального университета (КФУ) после прохождения практики в 2022/2023 учебному году. Согласно положению о практике обучающихся КФУ² **комплект отчетных документов включает в себя:**

- **отчет по практике** — основной документ, отражающий организацию работы в процессе практики и содержащий описание поставленных задач, методов их решения и полученных результатов;
- **дневник практиканта** — документ, включающий в себя краткое описание содержания выполненной работы и сроков ее выполнения;

31.03.2015: опубликовано 20.06.2015 / Кашупина И. А., Луковенкова О. О., Кудринская О. В.; правообладатель ФГБОУ ВПО «Камчатский государственный университет имени Витуса Беринга». Зарегистрировано в реестре программ для ЭВМ. EDN: UCMSJY. Режим доступа: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2015615893&TypeFile=html

¹ Свидетельство о государственной регистрации программы для ЭВМ № 2019661229 Российская Федерация. Программа автоматизированной проверки оформления офисных текстовых документов MAPDoc: № 2019617118; заявлено 17.06.2019; опубликовано 23.08.2019 / Стариченко Б. Е., Устинов М. А.; правообладатель ФГБОУ ВО «Уральский государственный педагогический университет». Зарегистрировано в реестре программ для ЭВМ. EDN: UKFBPW. Режим доступа: https://www.fips.ru/registers-doc-view/fips_servlet?DB=EVM&DocNumber=2019661229&TypeFile=html

² Положение о практике обучающихся, осваивающих основные профессиональные образовательные программы высшего образования. Принято решением ученого совета Казанского (Приволжского) федерального университета от 23 апреля 2020 года. <https://kpfu.ru/portal/docs/F535492349/Polozhenie.2020.Novoe.pdf>

- **индивидуальное задание** (при прохождении практики в основных структурных подразделениях КФУ) — документ, содержащий рабочий график (план) проведения практики;
- **путевка обучающегося-практиканта с индивидуальным заданием** (при прохождении практики в профильных организациях) — документ, включающий в себя информацию о поставленных задачах, приобретенных компетенциях и отзыве руководителя о работе практиканта.

Поскольку путевка обучающегося-практиканта предоставляется руководителю практики от образовательной организации исключительно в бумажном виде, данный документ был исключен из рассмотрения. Таким образом, для анализа были использованы 320 отчетов о прохождении практики, 320 дневников практиканта и 262 индивидуальных задания:

- 110 отчетов и дневников, 104 индивидуальных задания студентов 1-го года обучения;
- 103 отчета и дневника, 79 индивидуальных заданий студентов 2-го года обучения;
- 107 отчетов и дневников, 79 индивидуальных заданий студентов 3-го года обучения.

Процесс анализа был выполнен в ручном режиме: каждый комплект документов был последовательно рассмотрен на предмет соблюдения требований к заполнению документации, установленных в упомянутом ранее положении о практике, и рекомендаций руководителя практики по тексту отчета, основанных на наборе государственных стандартов, регламентирующих общие правила оформления документации³.

По результатам анализа было выявлено, что около 30 % комплектов документов составлены корректно:

- на 1-м году обучения — 34 %;
- на 2-м году обучения — 27 %;
- на 3-м году обучения — 28 %.

Оставшиеся 70 % комплектов содержат 51 разновидность несоответствий требованиям к оформлению, которые были классифицированы на 7 типов. Из дальнейшего рассмотрения были исключены недочеты содержательной части отчета: недостаточно подробное описание результатов выполнения задач и иные содержательные ошибки (обозначены в исследовании как «ошибки типа Т7»), — поскольку автоматизация обнаружения таких недочетов в полной мере не представляется возможной.

Таким образом, было выявлено, что 577 ошибок 49 разновидностей представляют собой несоответствия требованиям к оформлению (что составляет около 94 % всех выявленных ошибок в документации по практике). Типология несоответствий требованиям к оформлению документов приведена в таблицах 1–6.

³ Шахова И. С. Методические указания по выполнению и защите курсовой работы. Казань: Казанский (Приволжский) федеральный университет; 2019. 17 с. https://shelly.kpfu.ru/e-ksu/docs/F494780150/Shakhova_course_work_method.pdf

Таблица 1 / Table 1

Тип T1: ошибки заполнения полей документов**Type T1: errors in filling out document fields**

| ID | Несоответствие | Частота встречаемости | | |
|-------|--|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T1E1 | Неполный комплект документов | 0 | 1 | 3 |
| T1E2 | Не указана должность руководителя практики | 6 | 6 | 7 |
| T1E3 | Ошибка в заполненном поле «ФИО» | 2 | 1 | 0 |
| T1E4 | Отсутствуют отдельные поля | 1 | 2 | 0 |
| T1E5 | Некорректные даты прохождения практики | 2 | 3 | 2 |
| T1E6 | Некорректное место прохождения практики | 5 | 7 | 10 |
| T1E7 | Не указан номер группы | 3 | 6 | 4 |
| T1E8 | Некорректные даты сдачи отчета | 4 | 3 | 5 |
| T1E9 | Заполнено поле с оценкой | 3 | 6 | 2 |
| T1E10 | Некорректные шаблоны документов | 6 | 6 | 2 |
| T1E11 | Некорректное ФИО и/или должность руководителя практики | 0 | 2 | 1 |

Таблица 2 / Table 2

Тип T2: ошибки оформления содержания отчета**Type T2: errors in report content formatting**

| ID | Несоответствие | Частота встречаемости | | |
|------|---|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T2E1 | Несоответствие номеров страниц в оглавлении страницам в тексте | 3 | 6 | 3 |
| T2E2 | Страницы разделов содержания указаны в формате «первая страница раздела — последняя страница раздела» | 3 | 0 | 0 |
| T2E3 | Номера страниц указаны под названием раздела или через пробел после него | 3 | 5 | 3 |
| T2E4 | Среди пунктов содержания присутствует заголовок «Содержание» | 0 | 2 | 6 |
| T2E5 | Номера страниц в оглавлении не указаны или указаны не для всех компонентов текста | 3 | 2 | 1 |

Таблица 3 / Table 3

Тип T3: ошибки оформления текста отчета**Type T3: errors in report text formatting**

| ID | Несоответствие | Частота встречаемости | | |
|------|-------------------------------------|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T3E1 | Разный межстрочный интервал | 13 | 18 | 15 |
| T3E2 | Пустые строки между абзацами | 19 | 15 | 13 |
| T3E3 | Разное выравнивание обычного текста | 16 | 21 | 15 |
| T3E4 | Пропущены отступы красной строки | 17 | 15 | 11 |

Окончание таблицы 3 /
End of the table 3

| ID | Несоответствие | Частота встречаемости | | |
|-------|---|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T3E5 | На титульном листе указан номер страницы | 1 | 0 | 0 |
| T3E6 | Пропущены точки в конце предложений | 2 | 0 | 3 |
| T3E7 | Выделены цветом отдельные фрагменты текста | 2 | 1 | 1 |
| T3E8 | Пустые страницы | 2 | 1 | 2 |
| T3E9 | Разные шрифты (размер, семейство) у однотипных элементов | 3 | 12 | 10 |
| T3E10 | Разный цвет текста | 0 | 2 | 8 |
| T3E11 | Некорректное оформление фрагментов программного кода | 0 | 2 | 2 |
| T3E12 | Некорректная верстка статических компонентов отчетных документов, не предполагающих изменения | 6 | 8 | 2 |
| T3E13 | Глава начинается с рисунка | 1 | 1 | 0 |
| T3E14 | Отсутствие отступа между параграфами | 1 | 0 | 1 |

Таблица 4 / Table 4

Тип 4: ошибки оформления рисунков**Type 4: errors in images formatting**

| ID | Несоответствие | Частота встречаемости | | |
|------|---|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T4E1 | Подпись к рисунку перенесена на следующую страницу | 2 | 4 | 4 |
| T4E2 | Рисунок выходит за поля страницы | 1 | 0 | 3 |
| T4E3 | Подпись к рисунку размещена не по центру | 2 | 1 | 3 |
| T4E4 | Нет подписей к рисункам (подписи не ко всем рисункам, подпись неполная) | 4 | 11 | 4 |
| T4E5 | Рисунок размещен не по центру | 3 | 3 | 4 |
| T4E6 | Некорректная нумерация рисунков (дублирование, пропуски и др.) | 3 | 4 | 2 |
| T4E7 | Точка в конце названия рисунка | 1 | 0 | 0 |
| T4E8 | Отсутствие упоминания рисунков в тексте | 4 | 8 | 6 |

Таблица 5 / Table 5

Тип 5: ошибки оформления заголовков**Type 5: errors in headers formatting**

| ID | Несоответствие | Частота встречаемости | | |
|------|--------------------------------------|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T5E1 | Заголовки со строчной буквы | 1 | 0 | 0 |
| T5E2 | Заголовки с точкой в конце | 3 | 1 | 1 |
| T5E3 | Заголовки без полужирного начертания | 1 | 1 | 1 |

Окончание таблицы 5 /
End of the table 5

| ID | Несоответствие | Частота встречаемости | | |
|------|--|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T5E4 | Текст заголовка не соответствует заголовку в содержании | 2 | 1 | 1 |
| T5E5 | Главы начинаются не с новой страницы | 1 | 2 | 3 |
| T5E6 | Поставлен номер перед заголовком, который не нумеруется (введение, заключение, список использованных источников) | 1 | 0 | 2 |
| T5E7 | Некорректная нумерация глав (дублирование, пропуски и др.) | 3 | 3 | 0 |
| T5E8 | Пропущен заголовок | 0 | 1 | 0 |
| T5E9 | Одинаковые названия глав | 0 | 1 | 0 |

Таблица 6 / Table 6

Тип T6: ошибки оформления списка использованных источников
Type T6: errors in references formatting

| ID | Несоответствие | Частота встречаемости | | |
|------|--|-----------------------|------------------|------------------|
| | | 1-й год обучения | 2-й год обучения | 3-й год обучения |
| T6E1 | Несоответствие требованиям к оформлению списка использованных источников | 10 | 20 | 9 |
| T6E2 | Отсутствие упоминания источников в тексте | 4 | 8 | 6 |

По результатам проведенного анализа было выявлено, что в отчетных документах студентов преобладают ошибки оформления текста отчета (T3), в частности ошибки T3E1–T3E4, касающиеся форматирования компонентов основного текста. Следующим по частоте недочетом оформления является несоответствие требованиям к оформлению списка использованных источников (T6E1). Суммарно на пять упомянутых несоответствий приходится около 40 % всех обнаруженных ошибок. Десять наиболее часто встречающихся в отчетной документации несоответствий требованиям к оформлению — это ошибки T1E2, T1E6, T3E1–T3E4, T3E9, T4E4, T4E8, T6E1. Они составляют 57 % от общего числа ошибок. Двадцать наиболее часто встречающихся несоответствий (T1E2, T1E6–T1E10, T2E1, T2E3, T3E1–T3E4, T3E9–T3E10, T3E12, T4E1, T4E4, T4E8, T6E1–T6E2) покрывают 79 % ошибок.

Диаграмма суммарной частоты встречаемости несоответствий требованиям к оформлению в разрезе разновидностей ошибок приведена на рисунке 1.

При анализе с разбивкой по годам обучения наблюдается аналогичная тенденция распределения ошибок в отчетной документации студентов (рис. 2).

Полученные результаты позволяют приоритизировать задачи по автоматизации обнаружения несоответствий согласно частоте их встречаемости в отчетной документации студентов.

3. Методы поиска несоответствий требованиям к оформлению документации

Для масштабируемости создаваемого программного решения в качестве входного файла был принят текстовый документ. Реализация первой версии программных механизмов направлена на обработку файлов в формате DOCX.

Особенности различных типов ошибок оформления накладывают ограничения на методы, применяемые для их поиска. **Эффективным подходом**, который обеспечивает обнаружение до 90 % ошибок оформления (ошибки типа T1–T4, T6), **является метод на основе правил**¹. Он заключается в анализе входного файла с применением заранее заданных условий для обнаружения искомых синтаксических конструкций.

Так, например, для проверки корректности заполнения фиксированных полей документа задается набор ключевых слов, по которым необходимо проводить поиск. При обнаружении указанных слов проводится проверка соответствия последовательности символов, следующих за указанными ключевыми

¹ Метод на основе правил (англ. Rule Based) — один из методов извлечения информации, основанный на применении набора правил, которые определяют отношения вида «условие — действие».

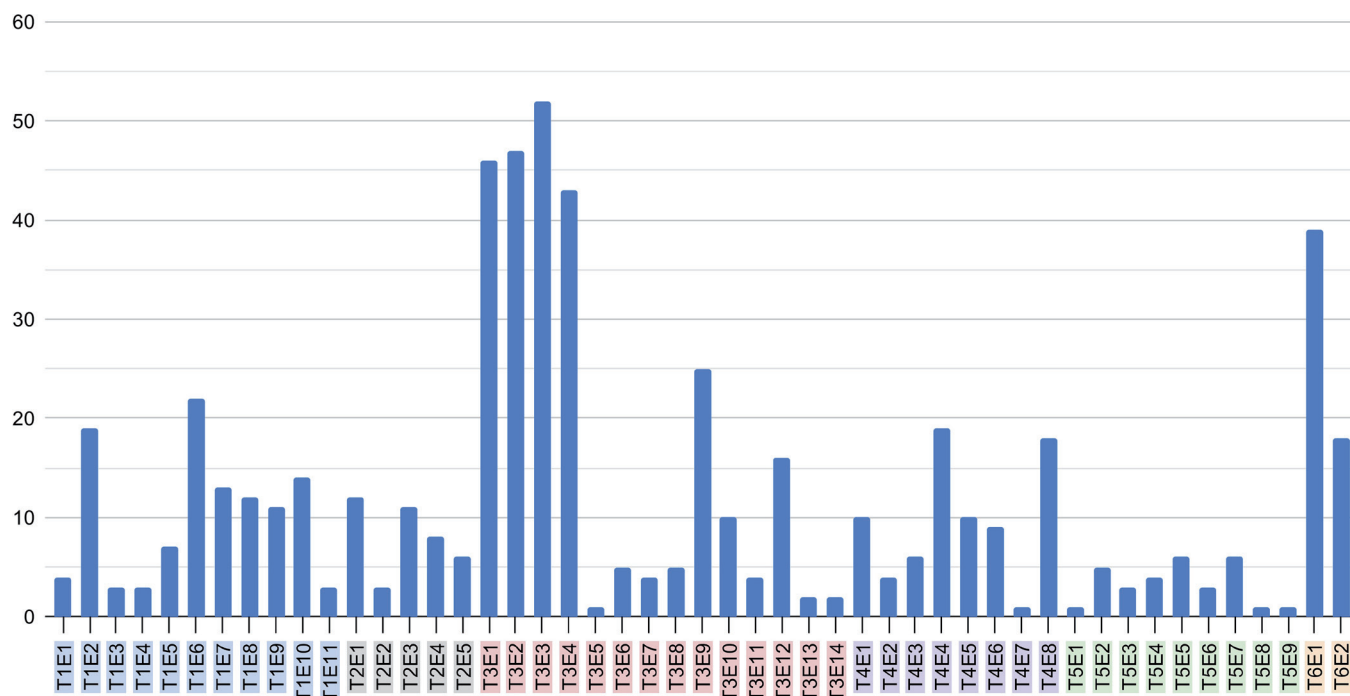


Рис. 1. Диаграмма суммарной частоты встречаемости ошибок оформления по разновидностям

Fig. 1. The diagram of the total frequency of formatting errors by kind

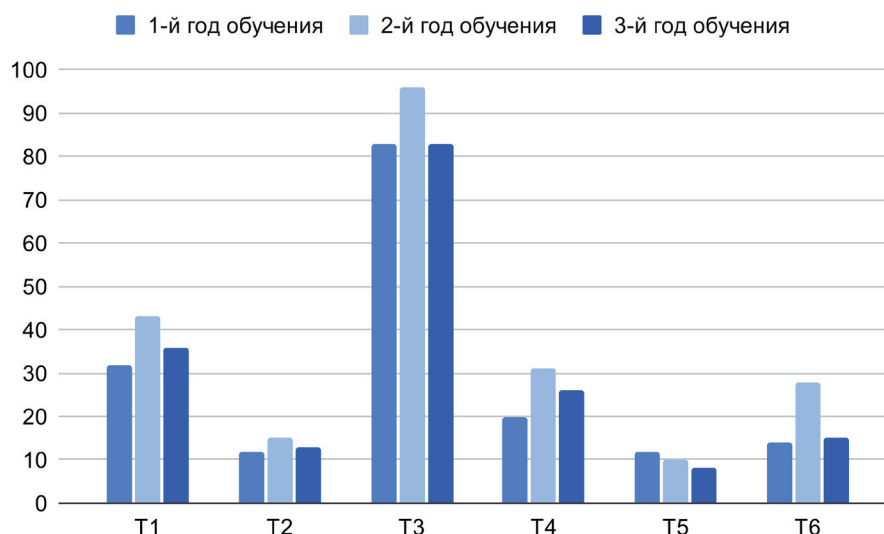


Рис. 2. Диаграмма частоты встречаемости ошибок оформления по типам ошибок и годам обучения

Fig. 2. The diagram of the frequency of formatting errors by error type and year of study

словами, строке, заданной в пользовательском интерфейсе программного решения. Аналогичные алгоритмы используются для обнаружения ошибок в содержании, оформлении текста отчета, подписей к рисункам и списка использованных источников информации.

Несмотря на то что метод, основанный на заранее заданных правилах, позволяет обнаружить подавляющее большинство несоответствий требованиям к оформлению, его применение с учетом первоначальной структуры DOCX-файла может

негативно сказываться на трудоемкости расширения функциональных возможностей программы, в частности, на внедрении семантического анализа текста. Для оптимизации трудоемкости и обеспечения масштабируемости программного решения было предложено графовое представление текста отчета по практике. Тело документа предлагается рассматривать в виде графа, где вершинами являются структурные компоненты текста разного уровня иерархии, а ребрами задаются связи типа «родитель — потомок» и принадлежность атрибутов

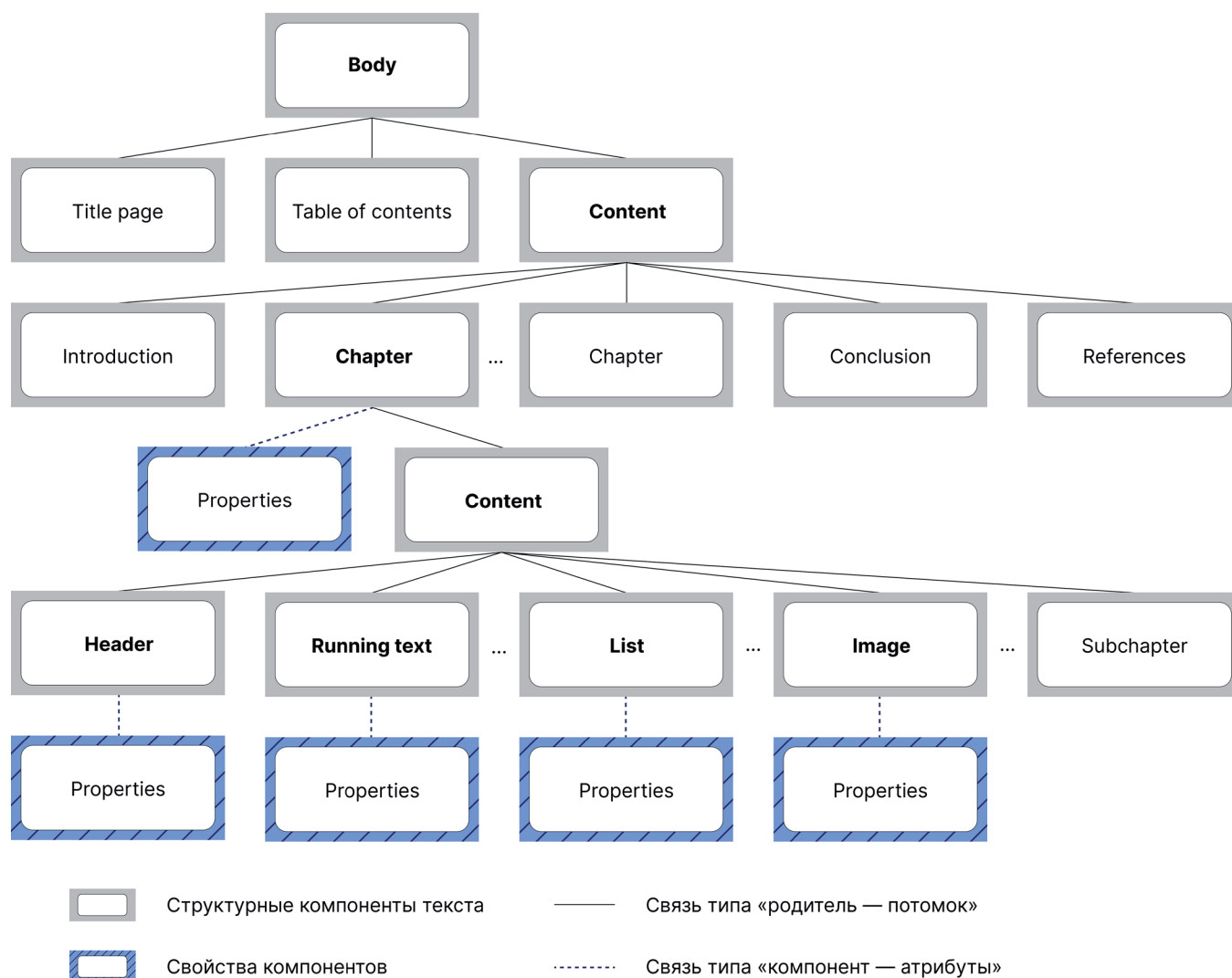


Рис. 3. Модифицированное графовое представление текста

Fig. 3. Modified graph representation of a text

соответствующему компоненту текста. Упрощенный вариант предложенного графового представления приведен на рисунке 3. Структурные компоненты текста, названия которых выделены полужирным начертанием на рисунке, детализированы до терминальной вершины хотя бы по одному из потомков, для других компонентов приведены исключительно узлы первого уровня вложенности.

Построение графа, представленного на рисунке 3, предполагает наличие программного механизма поиска заголовков глав и параграфов текста, который позволяет определять структурные компоненты типа “Chapter” и “Subchapter”. Данный механизм также является необходимым для обнаружения несоответствий требованиям к оформлению заголовков (ошибки типа Т5). Для разработки такого механизма был проведен количественный анализ 623 уникальных заголовков первого, второго и третьего уровней, используемых в отчетах студентов трех лет обучения. По результатам анализа было выявлено, что заголовки

не содержат глаголов, а медианное значение длины заголовков, измеряемое в количестве токенов (слов), варьируется в пределах двух. Так, медианное значение равно 3 для всех заголовков второго и третьего уровней, 4 — для заголовков первого уровня в отчетах студентов 1-го и 2-го года обучения, 5 — для заголовков первого уровня в отчетах студентов 3-го года обучения. Максимальным значением (без учета выбросов) длины заголовка является 11 токенов (рис. 4).

Выявленные количественные характеристики, а также особенность заголовка, требующая наличия символа переноса строки до и после заголовка при условии отсутствия опознавательных символов элемента списка, позволяют составить комплекс метрик для обнаружения заголовков в тексте отчета.

Таким образом, предложенные методы (метод на основе правил и метод выявления заголовков в тексте на основе количественных характеристик) обеспечивают обнаружение ошибок оформления всех шести типов.

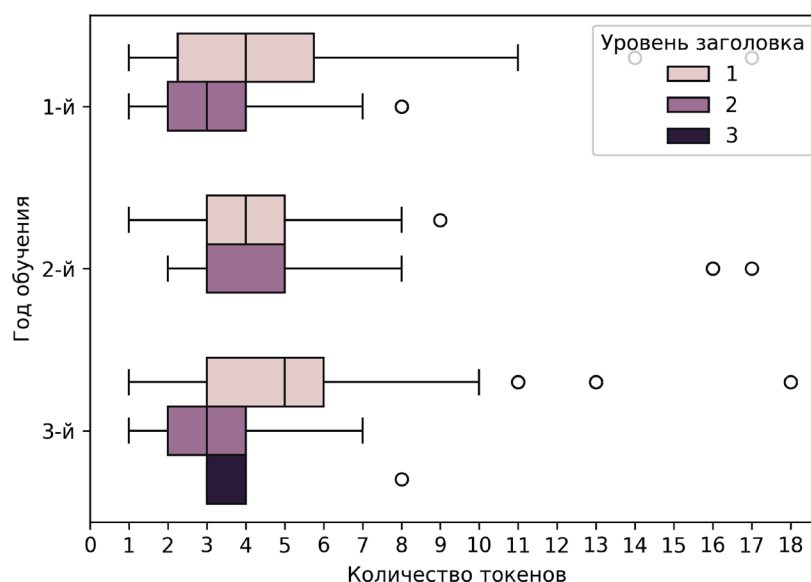


Рис. 4. Диаграмма размаха распределения длины заголовков (в токенах)

Fig. 4. The box plot of the distribution of headers length (in tokens)

4. Концепция программного решения автоматизации нормоконтроля отчетной документации

В основе концепции предложенного программного решения автоматизации нормоконтроля отчетной документации лежит принцип модульности, обеспечивающий снижение затрат на замену отдельных компонентов программы при ее дальнейшем развитии. Так, модуль обработки входного текстового файла отделен от модуля обнаружения ошибок, что позволяет расширять набор допустимых форматов. Модуль обнаружения несоответствий требованиям к оформлению, в свою очередь, может дополняться программными механизмами анализа текста документации на предмет иных нарушений. **Обособленность модуля, инкапсулирующего пользовательские интерфейсы программного решения, позволяет модифицировать подходы и средства его реализации в процессе тестирования или эксплуатации без необходимости вносить изменения в другие модули программы.**

Программное решение подразумевает наличие пользователей двух типов:

1. **Пользователь 1**, который загружает текстовую документацию на проверку и получает обратную связь о необходимости внесения изменений (для задачи автоматизации нормоконтроля отчетной документации обучающихся таким пользователем является **студент**).
2. **Пользователь 2**, который имеет доступ к результатам проверки документации на наличие несоответствий требованиям к оформлению и непосредственно к самой документации для возможности ее проверки на предмет иных недочетов, которые требуют исправления

(для рассматриваемой задачи это **преподаватель**).

Начальным взаимодействием с программным решением является настройка требований к оформлению. Это действие производится пользователем 2 единожды, заданные настройки сохраняются до дальнейшей их корректировки.

Пользователь 2 указывает следующие данные:

- межстрочный интервал;
- форматирование обычного текста: семейство шрифтов, кегль, выравнивание (по левому краю/по центру/по правому краю/по ширине), допустимость полужирного начертания (да/нет), допустимость курсивного начертания (да/нет), допустимость наличия выделения цветом (да/нет), допустимость разного цвета фрагментов текста (да/нет);
- оформление подписи к рисунку: регулярное выражение;
- форматирование заголовка: семейство шрифтов, кегль, полужирное начертание (да/нет), курсивное начертание (да/нет), выравнивание (по левому краю/по центру/по правому краю/по ширине), интервал перед заголовком, интервал после заголовка;
- оформление списка использованных источников информации: регулярное выражение;
- главы с новой страницы (да/нет).

Остальные атрибуты компонентов текста, проверка которых осуществляется в процессе поиска несоответствий в первой реализации программного решения, заданы по умолчанию.

Принципы взаимодействия пользователей с программным решением в рамках процесса проверки отчетной документации на соответствие требованиям к оформлению представлены на рисунке 5.

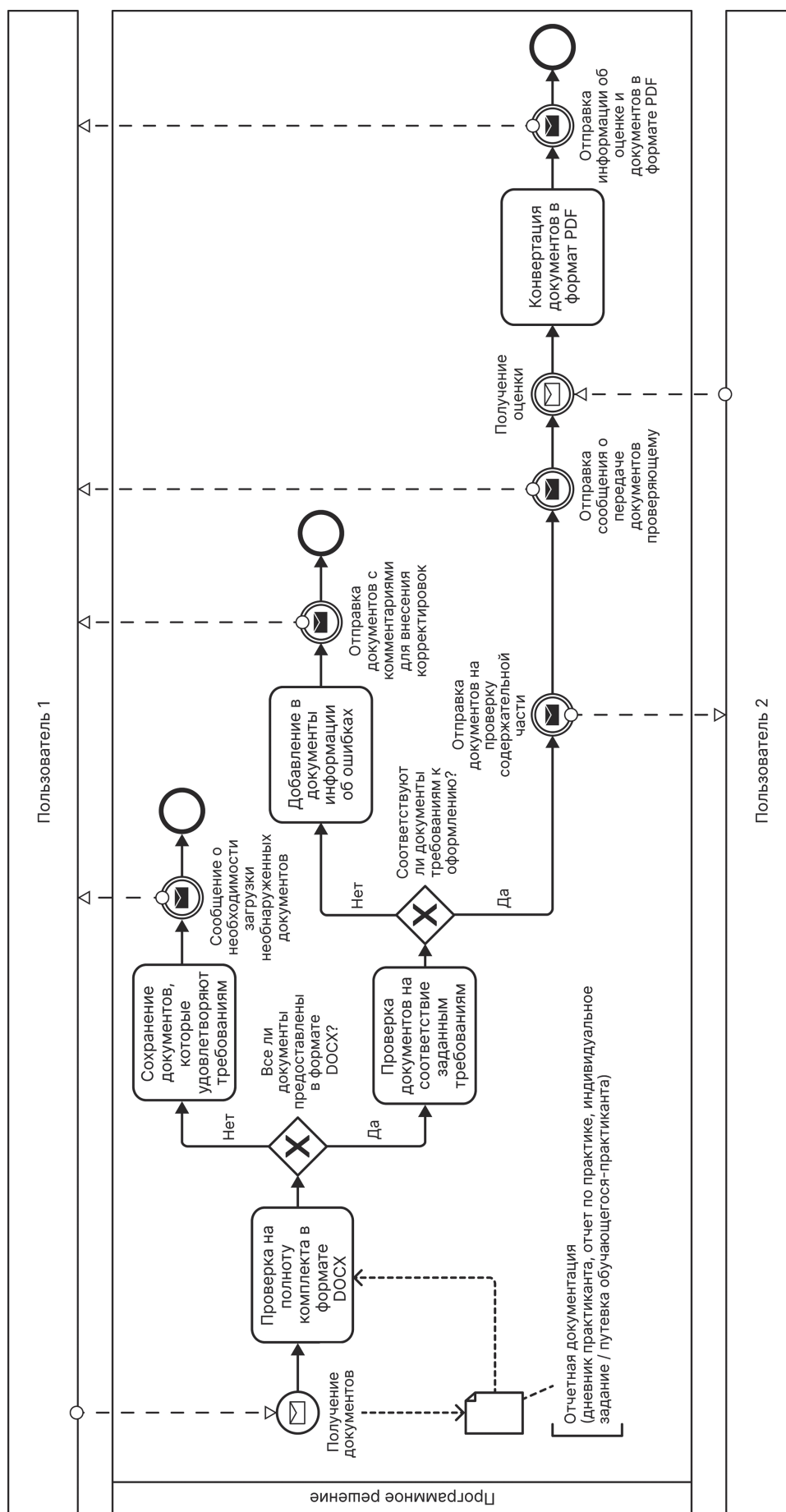


Рис. 5. Диаграмма процесса проверки отчетной документации
Fig. 5. The diagram of the process of checking reporting documents

После загрузки пользователем 1 текстовой документации в систему осуществляется проверка загруженного комплекта файлов на наличие всех необходимых документов в требуемом формате DOCX. Если комплект неполный и/или документы загружены в некорректном формате, пользователь 1 получает сообщение о необходимости повторной загрузки тех документов, которые не были обнаружены в формате DOCX. Данная последовательность действий производится циклично до тех пор, пока проверка комплекта документов на полноту не завершится успешно. С этого момента у пользователя 2 появляется возможность производить оценку загруженных документов даже в случае, если они не соответствуют требованиям к оформлению.

По результатам обработки загруженных документов строится графовое представление текстов, производится проверка соответствия документов требованиям к оформлению, которые заданы пользователем 2 и установлены по умолчанию. В случае обнаружения несоответствий в документ добавляются комментарии с сообщением об обнаруженной ошибке и рекомендациями по ее исправлению. Для применения предложенной программы в сторонней предметной области или с целью решения иных задач упомянутый программный механизм добавления комментариев может быть заменен на механизм автоматического исправления обнаруженных несоответствий.

Документы, в которых были обнаружены ошибки оформления, возвращаются пользователю 1 с сопроводительным сообщением о необходимости внести указанные в комментариях корректировки и загрузить документ в систему повторно. Данный процесс выполняется до тех пор, пока все требования не будут соблюдены. Если пользователь 1 считает, что на этапе проверки произошла ошибка, у него есть возможность связаться с пользователем 2 для пояснения ситуации.

После проверки содержательной части документа пользователь 2 выставляет оценку работе, а пользователь 1 получает сообщение об итоговых результатах и свои документы, сконвертированные в формат PDF для печати.

5. Заключение

Результаты анализа ошибок, наиболее часто встречающихся в отчетной документации по практике обучающихся высших учебных заведений, демонстрируют значительное преобладание несоответствий требованиям к оформлению над недостатками, касающимися содержательной части. Несответствия требованиям к оформлению документов можно разделить на типы по связи с компонентами документации, в оформлении которых они были обнаружены. Были выявлены 6 типов ошибок:

- 1) заполнения полей документов;
- 2) оформления содержания отчета по практике;
- 3) форматирования текста отчета;

- 4) оформления рисунков;
- 5) оформления заголовков;
- 6) оформления списка использованных источников.

Полученные результаты позволяют наблюдать сходные тенденции распределения количества ошибок соответствующих типов по годам обучения и приоритизировать задачи по реализации методов поиска несоответствий с учетом частоты их встречаемости.

В статье предложены методы обнаружения выявленных несоответствий, разработанные с учетом особенностей ошибок разных типов. Ключевым является метод на основе правил, направленный на поиск несоответствий отдельных компонентов текста заранее заданным условиям. Указанный подход позволяет автоматизировать обнаружение до 90 % ошибок оформления. Поиск остальных 10 % ошибок оформления требует применения алгоритма классификации заголовков, который основан на предложенных метриках, сформированных по результатам анализа текстов заголовков. В целях оптимизации процесса поиска предложено модифицированное графовое представление текста, ориентированное в числе прочего на расширение функциональных возможностей описанного программного решения в сторону семантического анализа документа.

В основу концепции предложенного программно-го решения автоматизации нормоконтроля отчетной документации положен принцип модульности, который поможет снизить затраты на замену отдельных компонентов программы при ее дальнейшем развитии.

Результаты проведенной работы демонстрируют потенциал применения методов автоматизации для задач проверки отчетной документации студентов и обнаруживают перспективы масштабирования предложенного решения для его применения в сторонних предметных областях.

Список источников / References

1. Нестеров Г. А., Шестакова А. О., Карякин Ю. Е. Автоматизация проверки научно-технической документации. *Математическое и информационное моделирование. Материалы Всероссийской конференции молодых ученых*. Тюмень: Тюменский государственный университет; 2021;19:97–101. EDN: IAINCM.

[Nesterov G. A., Shestakova A. O., Karyakin Yu. E. Automation of verification of scientific and technical documentation. *Mathematical and Information Modeling. Proc. All-Russian Conf. of Young Scientists*. Tyumen, University of Tyumen; 2021;19:97–101. (In Russian.) EDN: IAINCM.]

2. Зобнина В. Е., Вихорь Н. А. Автоматизация проверки форматирования текстового документа. *Избранные доклады 65-й Юбилейной университетской научно-технической конференции студентов и молодых ученых. Сборник докладов*. Томск: Томский государственный архитектурно-строительный университет; 2019:880–881. EDN: WOXXEH.

[Zobnina V. E., Vikhor N. A. Automation of checking the formatting of a text document. *Selected Reports of the 65th Anniversary University Scientific and Technical Conf. of Students and Young Scientists. Collection of Reports*.

Tomsk, Tomsk State University of Architecture and Building; 2019:880–881. (In Russian.) EDN: WOXVEH.]

3. Поздин В. Н., Выймова Е. А. Алгоритм автоматизированного нормоконтроля работ обучающихся образовательного учреждения. *Общество. Наука. Инновации (НПК-2021). Сборник статей XXI Всероссийской научно-практической конференции*. Киров: Вятский государственный университет; 2021;2:584–588. EDN: HMPOFQ.

[Pozdin V. N., Vyimova E. A. Algorithm for automated standard control of the work of students of an educational institution. *Society. Science. Innovations (Scientific and Practical Conference-2021). Collection of articles 21th All-Russian Scientific and Practical Conf.* Kirov, Vyatka State University; 2021;2:584–588. (In Russian.) EDN: HMPOFQ.]

4. Стариченко Б. Е., Устинов М. А. Программа автоматизации контроля оформления текстовых документов. *Педагогическое образование в России*. 2018;(8):163–168. EDN: UYIHPU. DOI: 10.26170/po18-08-22.

[Starichenko B. E., Ustinov M. A. Program of automation verification of text documents formatting. *Pedagogical Education in Russia*. 2018;(8):163–168. (In Russian.) EDN: UYIHPU. DOI: 10.26170/po18-08-22.]

5. Карманова К. А., Кудрявцева А. Г., Рыбникова В. В. Исследование и создание сервиса проверки оформления и форматирования документов. *Материалы научно-технической конференции студентов, аспирантов и молодых специалистов имени Е. В. Армского*. М.: Московский институт электроники и математики НИУ ВШЭ; 2022:113–116. EDN: ZBXIVU. Режим доступа: <https://miem.hse.ru/data/2022/09/02/1700621959/MIEM-HSE-2022-final.pdf>

[Karmanova K. A., Kudryavtseva A. G., Rybnikova V. V. Research and creation of a service for checking the design and formatting of documents. *Proc. Scientific and Technical Conf. of Students, Postgraduate Students and Young Specialists named after E. V. Armensky*. Moscow, HSE Tikhonov Moscow Institute of Electronics and Mathematics; 2022:113–116. (In Russian.) EDN: ZBXIVU. Available at: <https://miem.hse.ru/data/2022/09/02/1700621959/MIEM-HSE-2022-final.pdf>]

6. Zhigalova M., Sukhov A. System of automated checking of textual document design rules. *Information Content and Processing*. 2015;2(2):164–180. Available at: <http://www.foibg.com/ijicp/vol02/ijicp02-02-p05.pdf>

7. Марцинкевич В. И., Ларионова Г. С., Терещенко В. В., Ситникова К. А., Горлушкина Н. Н. Анализ возможностей парсинга электронных текстовых документов для автоматизации нормоконтроля. *Экономика. Право. Инновации*. 2022;(3):39–49. EDN: JLQANQ. DOI: 10.17586/2713-1874-2022-3-47-57.

[Martsinkevich V. I., Larionova G. S., Tereshchenko V. V., Sitnikova K. A., Gorlushkina N. N. Analysis of the possibilities of parsing electronic text documents for the automation of standard control. *Economics. Law. Innovation*. 2022;(3):39–49. (In Russian.) EDN: JLQANQ. DOI: 10.17586/2713-1874-2022-3-47-57.]

8. Кобец Е. А., Насыров Н. Ф., Комаров М. С., Горлушкина Н. Н. Алгоритм формализации требований к оформлению документов для сервиса автоматизированного нормоконтроля. *Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки*. 2021;(7):89–95. EDN: INOAIN. DOI: 10.37882/2223-2966.2021.07.20.

[Kobets E. A., Nasyrov N. F., Komarov M. S., Gorlushkina N. N. The algorithm of formalization the requirements for the execution of documents for the automation norm rule checking service. *Sovremennaya Nauka: Aktual'nyye Problemy Teorii i Praktiki. Seriya: Yestestvennyye i Tekhnicheskiye Nauki*. 2021;(7):89–95. (In Russian.) EDN: INOAIN. DOI: 10.37882/2223-2966.2021.07.20.]

9. Кобец Е. А., Насыров Н. Ф., Тартынский П. С., Горлушкина Н. Н. Построение объектной модели эталонного текстового документа для сервиса автоматизированного

нормоконтроля. *Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки*. 2021;(8):57–63. EDN: BVPPMU. DOI: 10.37882/2223-2966.2021.08.14.

[Kobets E. A., Nasyrov N. F., Tartynskikh P. S., Gorlushkina N. N. Creation the object model of the text benchmark document for automation the norm rule checking service. *Sovremennaya Nauka: Aktual'nyye Problemy Teorii i Praktiki. Seriya: Yestestvennyye i Tekhnicheskiye Nauki*. 2021;(8):57–63. (In Russian.) EDN: BVPPMU. DOI: 10.37882/2223-2966.2021.08.14.]

10. Kozyreva A., Nazarenko U., Berezhkov A., Nasyrov N. Analysis of the text of the FQP for automated standard control of documents. *Journal of Physics: Conference Series*. 2021;2131(2):1–7. DOI: 10.1088/1742-6596/2131/2/022102.

11. Martsinkevich V., Tereshchenko V., Larionova G., Berezhkov A., Kobets E., Nasyrov N., Gorlushkina N. Web tool for automated document formatting verification. *Proc. 33rd Conf. of FRUCT Association*. Helsinki, Finland, FRUCT Oy; 2023;(2):375–381. Available at: <https://fruct.org/publications/volume-33/acm33/files/Mar.pdf>

12. Захарова А. А., Аверьянова А. М., Глухих К. Д. Автоматизированная информационная система нормоконтроля в вузе: моделирование процессов управления. *Инновационные технологии в машиностроении. Сборник трудов XIV Международной научно-практической конференции*. Томск: Национальный исследовательский Томский политехнический университет; 2023:149–152. EDN: XOZDBC. Режим доступа: <https://earchive.tpu.ru/handle/11683/76559>

[Zakharova A. A., Averyanova A. M., Glukhikh K. D. Automated information system for normative control at a university: Modeling of management processes. *Innovative Technologies in Mechanical Engineering. Proc. of the 14th Int. Scientific and Practical Conf.* Tomsk, National Research Tomsk Polytechnic University; 2023:149–152. (In Russian.) EDN: XOZDBC. Available at: <https://earchive.tpu.ru/handle/11683/76559>]

13. Захарова А. А., Аверьянова А. М., Глухих К. Д. Web-приложение для автоматизации процесса проведения нормоконтроля учебной документации. *Современное образование: интеграция образования, науки, бизнеса и власти. Трансформация образования, науки и производства — основа технологического прорыва. Материалы международной научно-методической конференции*. Томск: Томский государственный университет систем управления и радиоэлектроники; 2023:260–263. EDN: XLTURO.

[Zakharova A. A., Averyanova A. M., Glukhikh K. D. Web application for automating process of conducting standard control of educational documentation. *Modern Education: Integration of Education, Science, Business and Government. Transformation of Education, Science and Manufacturing — the Basis of a Technological Breakthrough. Proc. Int. Scientific and Methodological Conf.* Tomsk, Tomsk State University of Control Systems and Radioelectronics; 2023:260–263. (In Russian.) EDN: XLTURO.]

14. Сливин А. Н., Шакура В. А. Информационная система автоматизированного контроля оформления выпускных квалификационных работ. *Измерения, автоматизация и моделирование в промышленности и научных исследованиях (ИАМП-2022). Материалы XVII Всероссийской научно-технической конференции студентов, аспирантов и молодых ученых с международным участием*. Бийск: Алтайский государственный технический университет им. И. И. Ползунова; 2022:11–12. EDN: JVHVXU.

[Slivin A. N., Shakura V. A. Information system for automated control of final qualification papers design. *Measurements, Automation and Modeling in Industry and Scientific Research (MAMI-2022). Proc. 17th All-Russian Scientific*

and Technical Conf. of Students, Postgraduate Students and Young Scientists with International Participation. Biysk, Polzunov Altai State Technical University; 2022:11–12. (In Russian.) EDN: JVHVXU.]

15. Pfitzmann B., Auer C., Dolfi M., Nassar A. S., Staar P. DocLayNet: A large human-annotated dataset for document layout segmentation. *Proc. 28th ACM SIGKDD Conf. on Knowledge Discovery and Data Mining*. New York, NY, USA, Association for Computing Machinery; 2022:3743–3751. DOI: 10.1145/3534678.3539043.

16. Bhowmik S., Kundu S., Sarkar R. BINYAS: A complex document layout analysis system. *Multimedia Tools and Applications*. 2021;80(6):8471–8504. DOI: 10.1007/s11042-020-09832-3.

17. Grijalva F., Santos E., Acuña B., Rodríguez J. C., Larco J. C. Deep learning in time-frequency domain for document layout analysis. *IEEE Access*. 2021;9:151254–151265. DOI: 10.1109/ACCESS.2021.3125913.

18. Wu X., Ma T., Du X., Hu Z., Yang J., He L. DRFN: A unified framework for complex document layout analysis. *Information Processing & Management*. 2023;60(3):1–16. DOI: 10.1016/j.ipm.2023.103339.

19. Wang B., Zhou J., Zhang B. MSNet: A multi-scale segmentation network for documents layout analysis. *Learning Technologies and Systems. Proc. 19th Int. Conf. on Web-Based Learning (ICWL 2020) and 5th Int. Symposium on Emerging*

Technologies for Education (SETE 2020). Springer-Verlag, Berlin, Heidelberg; 2021:225–235. DOI: 10.1007/978-3-030-66906-5_21.

20. Zhang P., Li C., Qiao L., Cheng Z., Pu S., Niu Y., Wu F. VSR: A unified framework for document layout analysis combining vision, semantics and relations. *Document Analysis and Recognition. Proc. 16th Int. Conf. (ICDAR 2021)*. Berlin, Heidelberg, Germany, Springer-Verlag; 2021:115–130. DOI: 10.1007/978-3-030-86549-8_8.

Информация об авторе

Шахова Ирина Сергеевна, старший преподаватель кафедры программной инженерии, Институт информационных технологий и интеллектуальных систем, Казанский федеральный университет, г. Казань, Республика Татарстан, Россия; *ORCID*: <https://orcid.org/0000-0003-1591-5767>; *e-mail*: is@it.kfu.ru

Information about the author

Irina S. Shakhova, Senior Lecturer at the Department of Software Engineering, Institute of Information Technology and Intelligent Systems, Kazan Federal University, Kazan, The Republic of Tatarstan, Russia; *ORCID*: <https://orcid.org/0000-0003-1591-5767>; *e-mail*: is@it.kfu.ru

Поступила в редакцию / Received: 29.03.24.

Поступила после рецензирования / Revised: 08.04.24.

Принята к печати / Accepted: 09.04.24.