

# АЛГОРИТМ ФОРМАЛИЗАЦИИ ТРЕБОВАНИЙ К ОФОРМЛЕНИЮ ДОКУМЕНТОВ ДЛЯ СЕРВИСА АВТОМАТИЗИРОВАННОГО НОРМОКОНТРОЛЯ

## THE ALGORITHM OF FORMALIZATION THE REQUIREMENTS FOR THE EXECUTION OF DOCUMENTS FOR THE AUTOMATION NORM RULE CHECKING SERVICE

**E. Kobets  
N. Nasyrov  
M. Komarov  
N. Gorlushkina**

*Summary.* The scientific paper the problem is considered the automation norm rule checking the execution of text documents. The authors investigated the process of increasing the functionality opportunities of the automated error search tool. The main stages are identified and the tasks are considered for the formalization of the requirements of various standards. The task of translating requirements from a verbal description in the algorithmic language has been solved. As a result, the algorithm was developed for checking the sequence of actions to formalize the requirements of standards that regulate and govern the execution of documentation. The results allow to expand the functionality opportunities of the automation documents norm rule checking service.

*Keywords:* algorithm, automation, norm rule checking, formalization errors, execution of text documents, standards, unification.

**Кобец Елизавета Александровна**

Ведущий инженер, аспирант, Университет ИТМО  
(Санкт-Петербург)  
www.kobets@yandex.com

**Насыров Наиль Фаизович**

Ассистент, аспирант, Университет ИТМО (Санкт-Петербург)  
pasdel@mail.ru

**Комаров Михаил Сергеевич**

Инженер, Университет ИТМО (Санкт-Петербург)  
the237th@yandex.com

**Горлушкина Наталия Николаевна**

К.т.н., с.н.с., доцент, Университет ИТМО (Санкт-Петербург)  
nagor.spb@mail.ru

*Аннотация.* В статье рассматривается проблема автоматизации нормоконтроля оформления текстовых документов. Авторами был исследован процесс расширения функциональных возможностей инструмента автоматизированного поиска ошибок. Выявлены основные этапы и рассмотрены задачи по формализации требований различных стандартов. Решена задача перевода требований из словесного описания в алгоритмический язык. В результате был разработан алгоритм проверки последовательности действий по формализации требований стандартов, регулирующих и регламентирующих оформление документации. Полученные результаты позволяют расширить функциональные возможности сервиса автоматизированного нормоконтроля документов.

*Ключевые слова:* алгоритм, автоматизация, нормоконтроль, ошибки оформления, оформление текстовых документов, стандарты, унификация.

## Введение

Оформление текстовых документов, казалось бы, задача не особо важная для информационного общества. Однако, при переходе к цифровизации экономики и цифровой трансформации общества невыполнение стандартов вполне может привести к непоправимым ошибкам [1]. Именно переход к цифровизации особенно четко требует выполнения правил, которые едины и прописаны в стандартах. При этом проблема осуществления качественного нормоконтроля при сокращении времени на рутинную работу стоит перед разработчиками достаточно давно. И ее актуальность со временем не уменьшается. Практиче-

ски во всех странах существуют стандарты оформления документации. Ряд работ российских ученых посвящены этой проблеме в последние годы, например, [2–4]. На проблему проверки оформления документации указывают и зарубежные авторы [5–8].

К решению проблемы авторы обращались и ранее, рассматривая только оформление выпускных квалификационных работ (ВКР). Это решение — разработанный сервис автоматизированного нормоконтроля документов и обучения оформлению документации — представлено в статьях [9,10] и описано ниже. Именно внедрение и тестирование сервиса послужило продолжению исследовательской работы для расширения

возможностей и качества работы сервиса. Текстовые документы образовательных учреждений не ограничиваются только текстами выпускных квалификационных работ, поэтому возникла необходимость подстройки сервиса к различным ГОСТам оформления документации.

В работе рассматривается проблема нормоконтроля [11] оформления отчетной документации образовательного учреждения в соответствии с:

- ◆ «ГОСТ 7.32–2017. Межгосударственный стандарт. Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления»,
- ◆ «ГОСТ Р 7.0.11–2011. Система стандартов по информации, библиотечному и издательскому делу. Диссертация и автореферат диссертации. Структура и правила оформления».

Выбор ГОСТ 7.32–2017 и ГОСТ Р 7.0.11–2011 обусловлен тем, что на них основываются требования к оформлению ВКР студентов, отчетов о научно-исследовательских работах (НИР), а также оформление кандидатских и докторских диссертаций и авторефератов диссертаций по всем отраслям знаний. Важность области нормоконтроля, с точки зрения практики применения, доказывается положениями ГОСТов:

#### 1. ГОСТ 7.32–2017:

- a. «устанавливает общие требования к структуре и правилам оформления отчетов о научно-исследовательских, проектно-конструкторских, конструкторско-технологических и проектно-технологических работах (далее — отчетов о НИР)», напрямую указывается что: «единая процедура оформления будет содействовать обмену информацией, совершенствуя обработку отчета в информационной системе» (см. раздел 1 Область применения ГОСТ 7.32–2017),
- b. реализацией деятельности нормоконтролера и соблюдением всех нормативных требований: «Основная задача нормоконтролера — проверка соблюдения норм и требований, установленных настоящим стандартом, соблюдение всех нормативных требований, соблюдения единообразия в оформлении структурных элементов и правил оформления отчета о НИР» (п. 5.2.3 ГОСТ 7.32–2017),
- c. является основополагающим и обязательным документом к применению при проверке ВКР, НИР,
- d. устанавливает общие требования к структуре проверяемых документов, правилам оформления,
- e. объясняет в рамках некоторых примеров различные случаи применения.

#### 2. ГОСТ Р 7.0.11–2011:

- a. «устанавливает общие требования к оформлению кандидатских и докторских диссертаций и авторефератов диссертаций по всем отраслям знаний» (см. раздел 1 Область применения ГОСТ Р 7.0.11–2011),
- b. является основополагающим и обязательным документом к применению при проверке диссертации в виде рукописи, в виде научного доклада, автореферата диссертации (см. разделы 4–9 ГОСТ Р 7.0.11–2011),
- c. устанавливает общие требования к структуре проверяемых документов, правилам оформления, кандидатских и докторских диссертаций и авторефератов диссертаций,
- d. объясняет в рамках некоторых примеров различные случаи применения.

#### 1 Описание сервиса.

##### Общие идеи разработки

Как было сказано выше, в Университете ИТМО реализуется «Сервис автоматизированного нормоконтроля документов и обучения оформлению документации». Автоматизированная проверка соответствия документации установленным требованиям с помощью сервиса позволила существенно сократить время на проверку ВКР и соответственно повысить качество их оформления. Подробно о сервисе описано в работах [9,10].

Основные этапы работы сервиса и поток данных между ними выражены в следующем виде:

1. загрузка и сохранение нового документа на сервер, проверка его оформления,
2. сохранение документа на сервере (вход: аутентификационная информация, docx документ; выход: docx, закрепленный за пользователем, идентификатор документа),
3. получение свойств параграфов документа (вход: docx; выход: свойства параграфов в формате csv файла),
4. классификация параграфов (вход: свойства параграфов в формате csv; выход: последовательность меток, обозначающих классы),
5. повышение точности классификатора с помощью методики меток (вход: последовательность меток; выход: модифицированная последовательность меток),
6. проверка документа на соответствие выбранным требованиям (вход: docx документ, последовательность меток, идентификатор выбранных требований; выход: json с результатами проверки),
7. сохранение последних результатов проверки для документа (вход: идентификатор документа, результаты проверки, выбранные требования;

выход: идентификатор исправлений, идентификатор документа, json с результатами проверки).

В основу разработки системы автоматизированной проверки соответствия требованиям оформления была положена клиент-серверная архитектура [12]. Необходимо было предусмотреть реализацию архитектуры с учетом непрерывности процессов создания и проверки оформления документа. Для этого один из клиентских модулей системы был разработан на базе технологии надстроек Office (MS Word Add-in). Основные идеи:

- ◆ реализация проверки оформления электронных документов единого наиболее распространенного формата электронных документов — docx,
- ◆ в качестве клиентской стороны используется созданный add-in для microsoft word,
- ◆ клиент отправляет файл в исходном виде на сервер,
- ◆ на сервере запускается ряд последовательных процессов по проверке документа на соответствие требованиям оформления соответствующего стандарта (включая модуль машинного обучения и модуль проверки),
- ◆ итогом работы серверной части является список абзацев и список ошибок к ним (corrections.json),
- ◆ список ошибок выводится в add-in.

Реализованный сервис автоматизированного нормоконтроля позволяет выполнять следующие функциональные требования в контексте проверки оформления электронных документов формата docx:

- ◆ выполняется процесс проверки оформления электронных документов формата docx по ГОСТ 7.32–2017,
- ◆ разработан add-in (встроенное приложение для microsoft word), который позволяет взаимодействовать с сервисом,
- ◆ по результатам проверки пользователь получает наглядный список ошибок, который выводится в разработанный add-in.

Разработанный инструмент позволяет проводить автоматизированный нормоконтроль оформления документов в соответствии с требованиями ГОСТ 7.32–2017 (веб-сайт, RESTful API, подсистема самооценки и взаимной оценки работ пользователями системы, подсистема автоматизированной проверки работ на соответствие требованиям стандартов по оформлению).

В процессе развития сервиса, его внедрения, тестирования было выявлено то, что есть возможность расширения функций, одна из них — реализация поиска ошибок оформления электронных документов в соответствии с иными стандартами.

## Цель исследования

Решение проблемы оформления документации различного уровня в соответствии с ГОСТами и нормативными актами расширенной номенклатуры на примере ГОСТ Р 7.0.11–2011.

## Разработка алгоритма

При расширении номенклатуры документов и ГОСТов необходимо учитывать:

- ◆ разнообразие типовых элементов структуры документа,
- ◆ множество видов оформления,
- ◆ разнообразие способов идентификации и проверки каждого отдельного элемента.

Поскольку электронный документ состоит из множества структурных элементов. Под структурными элементами авторами понимается любое содержимое одного абзаца документа будь то обычный параграф, картинка, таблица и т.д. Проверка на соответствие требованиям определенного стандарта сводится к проверке каждого такого элемента. Таким образом, в ходе реализации сервиса по нормоконтролю был выделен общий подход к логике процедуры проверки. Для этого необходимо выполнение двух задач: идентификации каждого структурного элемента на соответствие какому-либо классу, а также формализации требований ГОСТа к оформлению документов и программной реализации алгоритмов проверки каждого структурного элемента. Настоящее исследование направлено на решение второй задачи.

Необходимым условием в процессе реализации сервиса автоматизированного нормоконтроля является работа с текстами соответствующих ГОСТов на предмет выявления формальных требований к оформлению документов. В ходе решения задачи по расширению функциональности сервиса автоматизированного нормоконтроля была выстроена логика работы к исследуемым текстам ГОСТов и была применена следующая методология исследования: наблюдение, сравнительный и контент-анализ, формализация, классификация, метод классов.

В последующем, проводились аналитические работы по ознакомлению, изучению и сравнению текстов основных и сопутствующих ГОСТов, на которые были ссылки в текстах основных ГОСТов в контексте выявления правил оформления текстовых документов.

На первом этапе исследования был изучен опыт наших коллег в рамках задачи “Анализ ГОСТ 7.32–2017” [9,10], систематизированы требования к оформлению

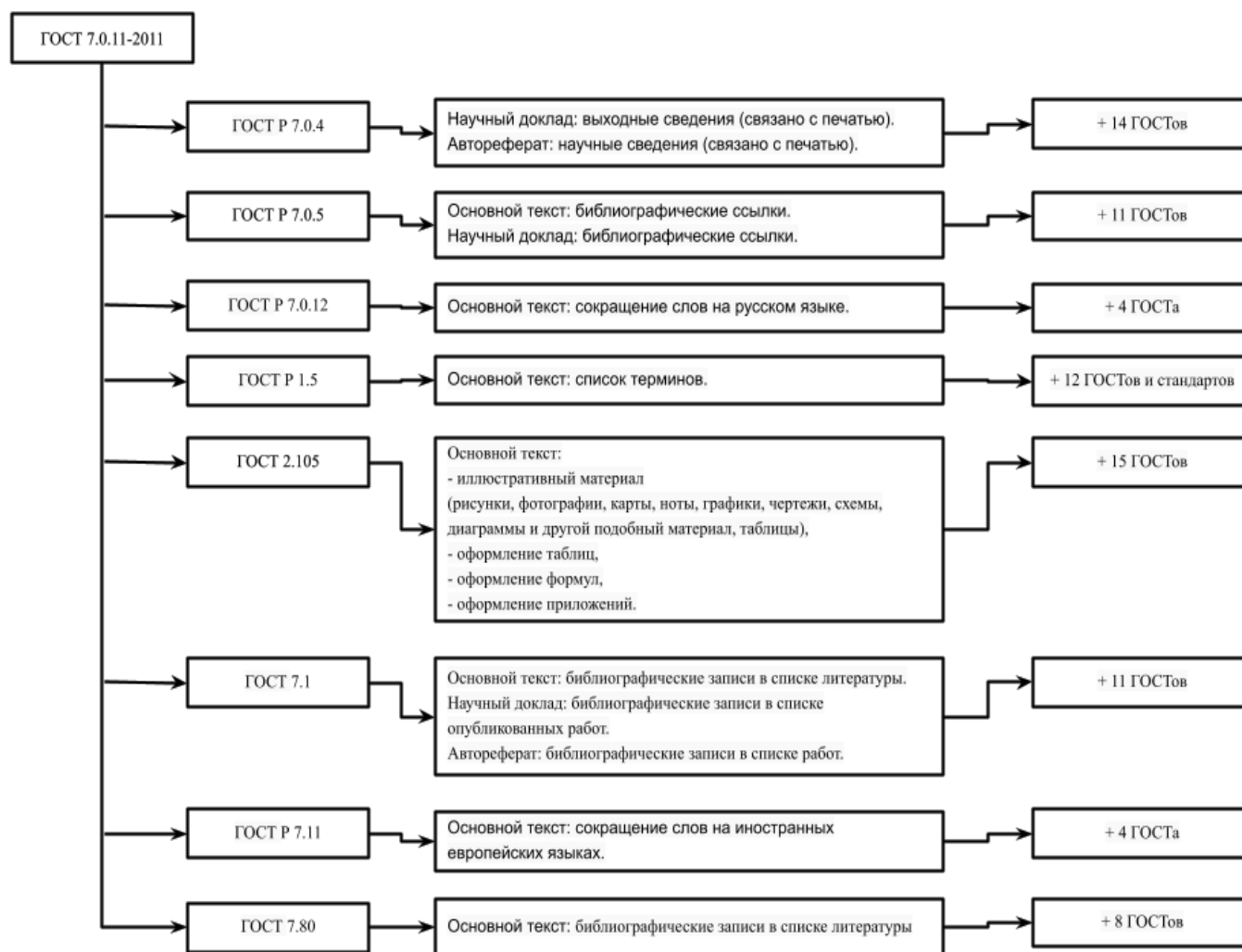


Рис. 1. Логика исследования ГОСТ Р 7.0.11–2011

текстовых элементов и реализован алгоритм проверки электронных документов формата docx.

На втором этапе исследования был проанализирован ГОСТ Р 7.0.11–2011 и составлена классификация преемственности вложенных ГОСТов (см. рисунок 1) в количестве 91 ГОСТ (основной ГОСТ Р 7.0.11–2011, и 8 вложенных ГОСТов, которые были упомянуты в описании, разделах и пунктах основного изучаемого ГОСТа, также были изучены сопутствующие ГОСТы во вложенных ГОСТах — еще 82 ГОСТа).

На основании накопленного опыта в результате процесса разработки был предложен универсальный алгоритм: “по формализации требований ГОСТа” (см. рисунок 2). Этот алгоритм позволяет осуществлять перенос требований различных ГОСТов и стандартов по оформлению электронных документов к программным алгоритмам проверки. Процесс работы по алгоритму можно описать следующим образом:

- ♦ определить основной ГОСТ, проверка требований к оформлению которого будет автоматизирована,
- ♦ последовательно изучить все разделы и пункты основного ГОСТа:
- ♦ при наличии в данном разделе требований к оформлению, перенести их в рабочий документ,
- ♦ при наличии ссылок на вложенные ГОСТы перейти к их изучению и переносу правил оформления в рабочий документ,
- ♦ провести классификацию структурных элементов текстовых документов,
- ♦ соотнести полученный массив требований с выделенными классами,
- ♦ формализовать требования к оформлению для дальнейшей разработки программного обеспечения:
- ♦ оставить краткое и емкое описание требований,
- ♦ на основании выбранной технологии разработки (для взаимодействия с файлами электронных

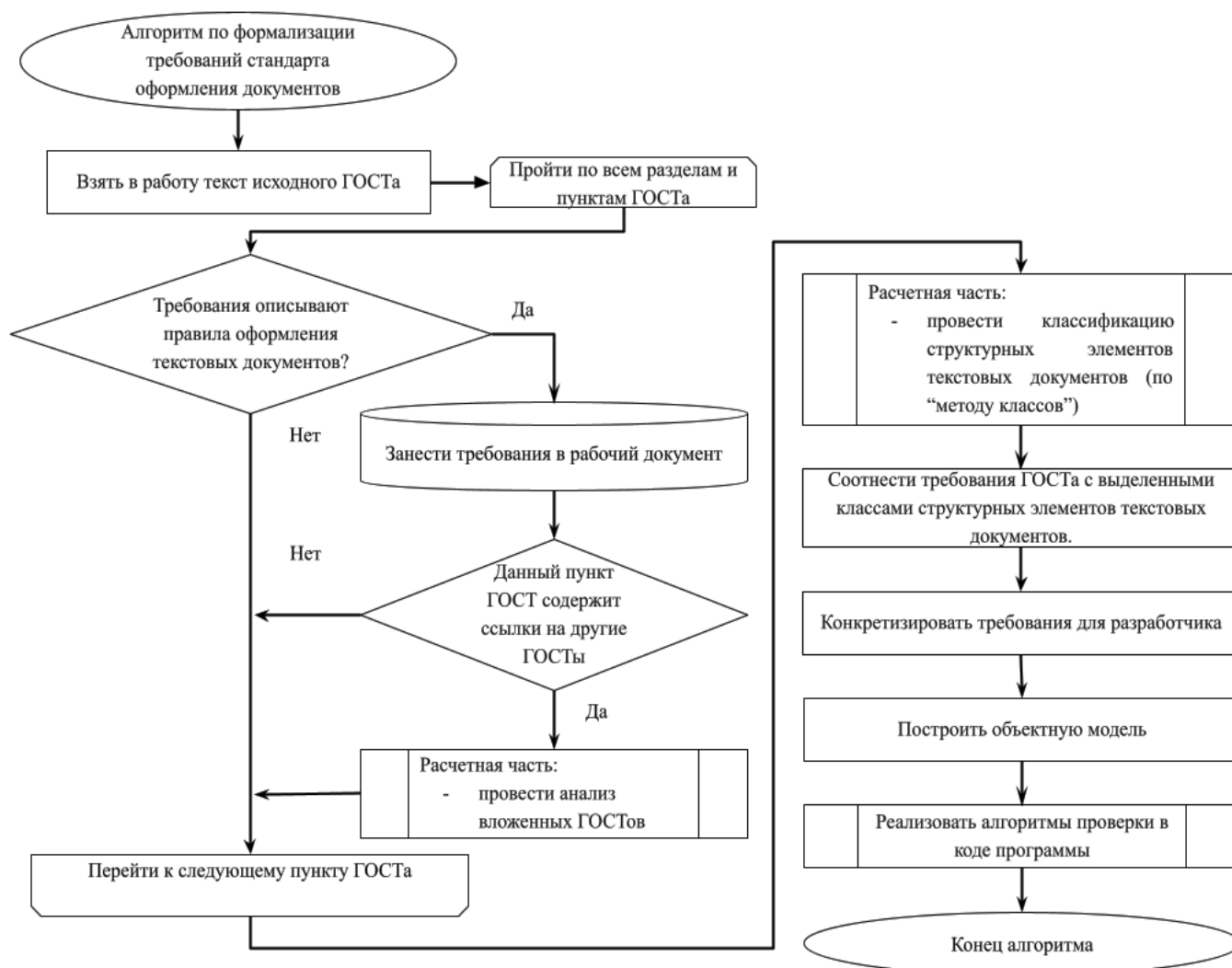


Рис. 2. Алгоритм проверки последовательности действий по формализации требований ГОСТа

документов) и проведенной классификации, построить объектную модель требований,

- ◆ реализовать алгоритмы проверки требований исходного ГОСТа в коде программы.

В предложенном алгоритме (см. рисунок 2) входными данными является текст какого-либо ГОСТа и стандарта, содержащий совокупность разделов и правил, включающий правила оформления текстовых документов. Результатом работы алгоритма является формализованные правила проверки оформления, выраженные в программном коде для каждого структурного элемента документа. Раскроем этапы реализации алгоритма.

На третьем этапе происходит ознакомление с текстом какого-либо стандарта, последовательный анализ каждого раздела и пункта и выделение требований, указывающих на проверку оформления текстовых документов. При наличии ссылок на иные нормативные

документы, ГОСТы и стандарты (например, “Формулы оформляют в соответствии с требованиями ГОСТ 2.105.” [текст ГОСТа Р 7.0.11]) происходит их анализ и дополнение выделенных правил оформления документов.

На четвертом этапе необходима работа с выборкой документов, соответствующих исследуемому стандарту: требуется выделить и классифицировать все типовые структурные элементы документов. Авторами предлагается сопоставлять выделенные классы с названиями или метками (например, “заголовок первого уровня” — a1, элемент “обычный абзац”- b1, и так далее [см. таблицу 1]).

Затем, на пятом этапе, проводится сопоставление полученных массивов правил оформления к выделенным классам, и полученный результат необходимо формализовать в каком-либо виде, например, в виде объектной модели, это необходимо для визуализации

Таблица 1. Определение классов структурных элементов документа

Структурный элемент документа		Класс	Особенности
Заголовок	заголовок первого уровня	b1	Выравнивание: по середине
			Полужирное начертание
	...	...	
Заголовок	заголовок второго уровня	b2	Отступ первой строки: 1.25cm
			Содержит номер
	...	...	
Абзац	обычный абзац	c1	Выравнивание: по ширине
			Межстрочный интервал: Полуторный (1.5)
	...	...	
Абзац	абзац перед списком	c2	Последний символ: “.”
			...
	...	...	
Подпись к рисунку	как отдельный абзац	e1	Расположение: под рисунком
			Содержание: «Рисунок N — Пояснение»
	...	...	
...	...	...	...

и систематизации требований для разработчика — человека, который переведет собранный массив правил в программный код.

Далее, на шестом этапе, необходимо сопоставить выделенные элементы (классы) конкретным правилам оформления. Процесс классификации, выделения классов из поступающего документа может быть выполнен различными способами. В ходе разработки сервиса автоматизированного нормоконтроля для данных целей применялись алгоритмы машинного обучения.

На завершающем седьмом этапе предлагается перевести требования оформления в код программы, реализующей проверку каждого элемента (в нашем случае — на языке C#). Таким образом, реализуется ключевой этап автоматизированного нормоконтроля — выполняется поиск ошибок в текстовом документе.

## Заключение

В ходе исследования проведена работа по подготовке правил для разработчика для последующей реализации автоматизированного поиска ошибок в текстах, оформленных в соответствии с ГОСТами. В ходе реализации объектной модели были выделены классы, необходимые для реализации алгоритмов машинного обучения, целью которых является классификация структурных элементов, на основании которых выполняется дальнейшая проверка элементов требованиям отдельного класса.

Таким образом подтверждена и апробирована на практике гипотеза о том, что “Сервис автоматизированного нормоконтроля документов и обучения оформлению документации” способен находить и исправлять ошибки оформления текстов в соответствии с различными стандартами при некоторой адаптации сервиса, что показывает расширение его функционала.

## ЛИТЕРАТУРА

1. Стратегия цифровой трансформации: написать, чтобы выполнить // URL: <https://strategy.cdto.ranepa.ru/> (даты обращения: 01.11.2020–25.05.2021)
2. Байбаков В., Клименко Э. Опыт нормоконтроля. Техническое задание на разработку автоматизированной системы // Стандарты и качество — М., 2012. — № 7 С: 42–47
3. Соколов А.М. Нормоконтроль в техническом вузе: проблемы и пути их решения // Высшее образование сегодня. -. 2015. — № 8. — С. 26–28.
4. Соколов А.А., Дворянkin А.М., Ужва А.Ю. Разработка метода автоматизации процесса нормоконтроля технической документации. // Известия ВолгГТУ. 2013. № 22. С. 114–118
5. The Excelsior Online Writing LAB // URL: <https://owl.excelsior.edu/> (даты обращения: 16.01.2020–25.05.2021)
6. Purdue Online Writing Lab // URL: [https://owl.purdue.edu/owl/purdue\\_owl.html](https://owl.purdue.edu/owl/purdue_owl.html) (даты обращения: 16.01.2020–25.05.2021)
7. Citation Styles: APA, MLA, Chicago, Turabian, IEEE: Home // URL: <https://pitt.libguides.com/citationhelp> (даты обращения: 16.01.2020–25.05.2021)
8. APA Style // URL: <https://apastyle.apa.org/> (даты обращения: 16.01.2020–25.05.2021)

9. Berezhkov A.V., Nasyrov N.F., Valitova Y.O., Ivanov S.E., Gorlushkina N.N., Kobets E.A. Organizational models of student peer assessment // 1st International Conference on Computer Technology Innovations dedicated to the 100th anniversary of the Gorky House of Scientists of Russian Academy of Science (ICCTI-2020) — 2020, pp. 40–45
10. Насыров Н.Ф., Кобец Е.А., Горлушкина Н.Н. Автоматизированная генерация учебных подзадач на основе методики тегов и критериев // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки — 2020. — № 3. — С. 102–107
11. А.В. Бережков, Ю.О. Валитова, А.И. Клименко, Д.Д. Пономарев Опыт повышения качества оформления выпускных квалификационных работ студентов технического вуза // Педагогический журнал. — 2020. — Т. 10. — № 1–1. — С. 367–375
12. Nail Nasyrov, Mikhail Komarov, Petr Tartynskikh, Nataliya Gorlushkina. Automated formatting verification technique of paperwork based on the gradient boosting on decision trees // Procedia Computer Science, Volume 178, 2020, pp. 365–374, ISSN1877–0509, <https://doi.org/10.1016/j.procs.2020.11.038>

© Кобец Елизавета Александровна ( [www.kobets@yandex.com](mailto:www.kobets@yandex.com) ), Насыров Наиль Фаизович ( [pasdel@mail.ru](mailto:pasdel@mail.ru) ),  
Комаров Михаил Сергеевич ( [the237th@yandex.com](mailto:the237th@yandex.com) ), Горлушкина Наталия Николаевна ( [nagor.spb@mail.ru](mailto:nagor.spb@mail.ru) ).  
Журнал «Современная наука: актуальные проблемы теории и практики»

