МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ

ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«Национальный исследовательский ядерный университет «МИФИ»

Обнинский институт атомной энергетики – филиал федерального

государственного автономного образовательного учреждения высшего

образования «Национальный исследовательский ядерный университет

«МИФИ» (ИАТЭ НИЯУ МИФИ)

Отделение интеллектуальных кибернетических систем

**Научно-исследовательская работа**

по направлению подготовки: 01.03.02 Прикладная математика и информатика

Направленность (профиль): Прикладная информатика

**«**Проверка оформления дипломных и научно-исследовательских работ на соответствие стандартам ГОСТ**»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. \_\_\_М-Б21\_\_ | (подпись, дата) | Рязанцева А. Е. |
| Руководитель ВКР,  к.ф.-м.н., доцент | (подпись, дата) | Царина А.Г. |
| Нормоконтроль | (подпись, дата) |  |
| Руководитель образовательной программы 01.03.02 Прикладная математика и информатика  к.ф.-м.н., доцент | (подпись, дата) | Ермаков С. В. |

Обнинск, 2024 г.

# **РЕФЕРАТ**

Отчет 95 с., 1 кн., 6 рис., 15 табл., 10 источн., 2 прил.

Ключевые слова: АВТОМАТИЗАЦИЯ, ГОСТ, АНАЛИЗ ДОКУМЕНТОВ, МАШИННОЕ ОБУЧЕНИЕ, ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ, ТЕКСТОВЫЕ ДОКУМЕНТЫ, ФОРМАТИРОВАНИЕ, СТАНДАРТЫ, ЭФФЕКТИВНОСТЬ, ПРОВЕРКА

Объектом исследования является процесс автоматизации проверки текстовых документов на соответствие стандартам ГОСТ. Цель данной работы — разработать программное обеспечение, которое сможет проводить анализ текстовых документов и выявлять несоответствия установленным стандартам.

Методы работы включают анализ существующих стандартов ГОСТ, изучение языка программирования Python для работы с документами в формате DOCX и применение алгоритмов машинного обучения для распознавания и обработки текстовой информации.

Результаты работы показывают новизну в подходе к автоматизации проверки, что позволяет сократить временные и трудовые затраты на обработку научных документов. Программное обеспечение интегрирует ручной и автоматический анализ, применяя модели машинного обучения для анализа сложных структур, таких как титульные листы и разделы.

Область применения результатов охватывает образовательные учреждения и научно-исследовательские организации, стремящиеся повысить качество и эффективность подготовки научных работ.

Рекомендации по внедрению включают интеграцию разработанного решения в учебные и научные процессы для повышения стандартизации оформления документов. Экономическая эффективность проекта проявляется в сокращении времени, затрат на проверку и повышении уровня соответствия стандартам.

Прогнозные предположения о развитии исследований заключаются в дальнейшем совершенствовании технологии анализа текстов и расширении возможностей программного обеспечения за счет алгоритмов машинного обучения.

# **Содержание**

[**РЕФЕРАТ** 3](#_Toc185880841)

[**Термины и определения** 5](#_Toc185880842)

[**Перечень сокращений и обозначений** 6](#_Toc185880843)

[**Введение** 7](#_Toc185880844)

[**Анализ стандартов ГОСТ** 8](#_Toc185880845)

[**Общий обзор стандартов** 8](#_Toc185880846)

[**Требования ГОСТ к структуре документов** 9](#_Toc185880847)

[**Практическое применение стандартов** 9](#_Toc185880848)

[**Работа с текстовыми документами в Python** 10](#_Toc185880849)

[**Инструменты и библиотеки** 10](#_Toc185880850)

[**Преобразование и анализ текста** 10](#_Toc185880851)

[**Автоматизация проверок** 11](#_Toc185880852)

[**Модели машинного обучения для анализа текста** 11](#_Toc185880853)

[**Задачи, требующие машинного обучения** 11](#_Toc185880854)

[**Выбор алгоритмов и моделей** 12](#_Toc185880855)

[**Применение моделей для анализа** 12](#_Toc185880856)

[**Проверка и верификация решений** 13](#_Toc185880857)

[**Ручная проверка и автоматизация** 13](#_Toc185880858)

[**Тестирование и результаты** 13](#_Toc185880859)

[**Оценка эффективности** 14](#_Toc185880860)

[**Заключение** 15](#_Toc185880861)

[**Список использованных источников** 17](#_Toc185880862)

[**Приложения** 18](#_Toc185880863)

# **Термины и определения**

Автоматизация — Использование программных средств для выполнения задач без непосредственного участия человека.

ГОСТ — Государственный стандарт, регламентирующий различные аспекты деятельности, включая оформление документов.

Нормативные документы — Документы, устанавливающие правила, общие принципы и критерии для определенной области деятельности.

Текстовый анализ — Процесс автоматического извлечения значимой информации из текста для его дальнейшей обработки и интерпретации.

Машинное обучение — Подмножество методов искусственного интеллекта, позволяющее компьютерным системам обучаться и улучшать свои функции без явного программирования.

Проверка соответствия — Процесс оценки документа или продукта на соответствие заданным стандартам и нормам.

Форматирование текста — Применение различных стилей и структурных элементов к тексту для обеспечения его читабельности и соответствия стандартам.

Документ DOCX — Формат текстового документа, используемый в Microsoft Word, который поддерживает сложное форматирование и встраивание мультимедийных объектов.

Алгоритмы — Набор правил и процедур, которые выполняются в определенной последовательности с целью решения задачи или достижения результата.

Эффективность — Способность программы или системы достигать поставленных целей с минимальными затратами ресурсов и времени.

# **Перечень сокращений и обозначений**

* **ГОСТ** — Государственный стандарт
* **DOCX** — Формат текстового документа Microsoft Word
* **ML** — Машинное обучение
* **TF-IDF** — Частотность термина, обратная частотности документа (Term Frequency-Inverse Document Frequency)
* **SVM** — Машина опорных векторов (Support Vector Machine)
* **RF** — Случайный лес (Random Forest)
* **NLP** — Обработка естественного языка (Natural Language Processing)

# **Введение**

В условиях стремительного развития информационных технологий и цифровизации образовательных процессов всё большее значение приобретает соответствие научно-исследовательских и дипломных работ стандартам ГОСТ. Стандартизация играет ключевую роль в упрощении обмена научной информацией, а также в обеспечении высокого качества научных документов, делая их более понятными и доступными для широкой аудитории.

Несмотря на очевидные преимущества стандартизации, соблюдение требований ГОСТ остаётся сложной и ресурсоёмкой задачей. Авторам необходимо обладать глубокими знаниями множества правил и норм, что становится особенно трудоёмким в условиях роста объёма научной информации и ускорения образовательных процессов. Это подчёркивает актуальность разработки автоматизированных средств, способных облегчить и ускорить проверку и оформление документов в соответствии с ГОСТ.

Целью данного исследования является создание программного обеспечения для автоматизации процесса проверки текстовых документов на соответствие стандартам ГОСТ. В рамках исследования будет проведён анализ действующих стандартов, разработаны алгоритмы для анализа текстов и структуры документов. Для решения более сложных задач, таких как проверка титульных листов, предполагается использование современных методов машинного обучения.

Актуальность темы обусловлена необходимостью повышения качества научных документов и их строгого соответствия установленным стандартам. Автоматизация процесса проверки позволит минимизировать количество ошибок, повысить эффективность подготовки документов и сократить временные затраты на их проверку. Эти задачи особенно актуальны для образовательных и научно-исследовательских учреждений, где требуется строгое соблюдение стандартов в большом количестве работ.

Исследование предполагает изучение теоретических и практических аспектов работы с текстовыми документами, включая использование языка программирования Python и библиотек для обработки файлов формата DOCX. Особое внимание будет уделено анализу того, какие элементы документа могут быть проверены алгоритмически, а какие требуют применения методов машинного обучения. Например, проверка наличия разделов может быть реализована через анализ заголовков, тогда как проверка корректности заполнения титульного листа может потребовать более сложных подходов.

Таким образом, данное исследование направлено на создание эффективного инструмента, который облегчит процесс оформления научных работ, повысит их качество и соответствие стандартам. Это, в свою очередь, будет способствовать улучшению образовательного процесса и научных исследований в целом.

# **1 Анализ стандартов ГОСТ**

## **1.1 Общий обзор стандартов**

В современной научной и образовательной практике Российской Федерации стандарты ГОСТ выполняют функцию унификации и стандартизации процессов подготовки и оформления научно-исследовательских и дипломных работ. Эти стандарты обеспечивают структурную согласованность и единообразие, что, в свою очередь, способствует более легкому восприятию и критической оценке научного контента. Основные стандарты, применяемые в оформлении научных документов, включают:

1. **ГОСТ Р 7.0.97-2016 "Библиографическая запись. Библиографическое описание":**
   1. Данный стандарт регулирует правила составления библиографических записей, что критично для правильного оформления списка литературы и ссылок на научные источники. Предоставление точных и полных библиографических данных является основой академической честности, позволяя обеспечивать прослеживаемость и проверяемость источников.
2. **ГОСТ 7.32-2001 (заменен ГОСТ 7.32-2017) "Отчет о научно-исследовательской работе. Структура и правила оформления":**
   1. Этот стандарт определяет структуру и правила оформления отчетов о научно-исследовательских работах, включая обязательные и факультативные разделы. Он обеспечивает систематичность и последовательность в представлении результатов исследований, что облегчает их интерпретацию и последующую эксплуатацию.
3. **ГОСТ Р 7.0.5-2008 "Библиографическая ссылка. Общие требования и правила составления":**
   1. Устанавливает правила оформления библиографических ссылок, что необходимо для правильной интеграции источников в текстовую часть документа. Этот стандарт поддерживает высокий уровень научной этики и позволяет читателям легко находить и проверять упомянутые источники.
4. **ГОСТ Р 7.0.4-2006 "Издания. Выходные сведения":**
   1. Регламентирует требования к оформлению выходных данных изданий, что особенно важно для структурирования титульного листа и первых страниц работы. Он способствует унификации презентации ключевой информации о работе.

Каждый из перечисленных стандартов выполняет специфические функции в контексте научной коммуникации, обеспечивая логичность и единообразие в оформлении научных работ.

## **1.2 Требования ГОСТ к структуре документов**

Стандарты ГОСТ предъявляют четко определенные требования к структуре научных документов, которые необходимо неукоснительно соблюдать. Например:

1. **Титульный лист:** должен включать информацию об учебном заведении, типе научной работы (например, дипломная работа), теме исследования, а также сведениях об авторе и руководителе. Это необходимо для идентификации документа и его автора в академическом процессе.
2. **Содержание:** должно иметь логическую структуру с указанием номеров страниц для каждого раздела, облегчая ориентацию и поиск информации в документе.
3. **Основная часть:** включает введение, главы с изложением основных результатов исследований и заключение. Каждая глава должна быть логически организована и, при необходимости, подразделена на параграфы.
4. **Список использованной литературы:** оформляется в соответствии с правилами библиографического описания, что упрощает проверку источников информации.
5. **Приложения:** содержат дополнительный материал, не входящий в основное тело документа, но важный для его полного понимания.

Соблюдение данных требований способствует тому, чтобы научные документы были ясными, последовательными и доступными для восприятия и оценивания.

## **1.3 Практическое применение стандартов**

На практике соблюдение стандартов ГОСТ может представлять значительные трудности ввиду их разнообразия и периодических обновлений. Проблемы возникают из-за:

1. **Многообразия требований**: каждый стандарт содержит свои специфические требования, которые необходимо учитывать и которые могут быть трудны для восприятия без надлежащей подготовки.
2. **Обновлений стандартов**: ГОСТы периодически подвергаются изменениям, что требует постоянного мониторинга и обновления соответствующих практик образовательными учреждениями и отдельными исследователями.
3. **Индивидуальных интерпретаций**: различные академические учреждения могут трактовать стандарты по-своему, что создает путаницу у студентов, пытающихся следовать унифицированным правилам.
4. **Технических ограничений**: не все текстовые редакторы или системы могут полностью удовлетворить всем стилевым и структурным требованиям, предусмотренным ГОСТ, что может потребовать дополнительных инструментов или программного обеспечения для их выполнения.

Эти проблемы подчеркивают необходимость в инструментах, которые могли бы автоматизировать процесс проверки документов на соответствие ГОСТ, минимизируя ошибки и облегчив работу исследователей.

# **2 Работа с текстовыми документами в Python**

## **2.1 Инструменты и библиотеки**

В сфере обработки текстовых документов Python выступает одной из самых популярных платформ благодаря широкому набору инструментов и библиотек.

1. **python-docx**:
   1. **Назначение**: Эта библиотека предназначена для чтения, изменения и создания файлов DOCX, которые широко используются в Microsoft Word.
   2. **Функциональные возможности**:
      1. **Чтение файлов DOCX**: позволяет извлекать текст, заголовки, таблицы и другие элементы из документа.
      2. **Изменение текста**: Возможность добавлять или удалять текстовые части, изменять стиль и формат текста.
      3. **Работа с таблицами и изображениями**: поддерживает операции над таблицами и встроенными изображениями, включая их добавление и изменение.
      4. **Форматирование**: Управление различными аспектами форматирования, такими как шрифты, размеры текста, стили (жирный, курсив и т. д.), выравнивание и отступы.

Эта библиотека оказывается незаменимой при автоматизации работы с текстовыми документами, позволяя писать скрипты для обработки больших объемов данных.

## **2.2 Преобразование и анализ текста**

Работа с текстовыми данными в DOCX требует нескольких ключевых этапов:

1. **Чтение и извлечение текста**:
   1. **Процесс**: используя python-docx, можно извлечь текстовые данные из документа, включая отдельные абзацы и заголовки.
   2. **Цель**: это позволяет получить доступ к содержимому документа, необходимому для дальнейшего анализа и преобразования.
2. **Токенизация текста**:
   1. **Процесс**: деление текстового содержимого на более мелкие единицы, такие как слова или предложения, с помощью библиотек, например, nltk (Natural Language Toolkit).
   2. **Цель**: токенизация упрощает анализ текста, позволяя выделить ключевые слова и фразы, изучить частоту их использования, а также провести морфологический анализ.
3. **Анализ структуры документа**:
   1. **Процесс**: проверка наличия всех необходимых разделов (например, введение, заключение) и анализ их логической структуры.
   2. **Цель**: это позволяет удостовериться в том, что документ соответствует установленным стандартам структуры и оформления, таким как ГОСТ.
4. **Изменение и формирование документов**:
   1. **Процесс**: внесение изменений в документ, такие как исправление форматирования, добавление заголовков или корректировка таблиц.
   2. **Цель**: это помогает автоматизировать и стандартизировать процесс подготовки документов, обеспечивая соответствие их оформления установленным требованиям.

## **2.3 Автоматизация проверок**

Автоматизация проверки текстовых документов является важной частью разработки программного обеспечения, обеспечивающего соответствие стандартам ГОСТ.

1. **Проверка структуры**:
   1. **Процесс**: использование скриптов Python для автоматической проверки наличия всех необходимых разделов документа.
   2. **Цель**: обеспечить, чтобы документ включал все обязательные элементы, такие как титульный лист, содержание, введение и заключение, что необходимо для соответствия стандарту.
2. **Анализ форматирования**:
   1. **Процесс**: контроль за правильностью оформления шрифта, отступов и межстрочных интервалов.
   2. **Цель**: проверить, соответствует ли текст установленным стандартам форматирования (например, Times New Roman 14pt, выравнивание по ширине).
3. **Проверка таблиц и формул**:
   1. **Процесс**: автоматическая проверка нумерации и оформления таблиц и формул.
   2. **Цель**: убедиться, что все таблицы, графики и формулы оформлены в соответствии с требованиями ГОСТ, включая правильную нумерацию и наличие подписей.

# **3 Модели машинного обучения для анализа текста**

## **3.1 Задачи, требующие машинного обучения**

Определенные задачи анализа и обработки текстов требуют применения методов машинного обучения, особенно в случаях, когда традиционные методы не справляются с объемом или сложностью задачи:

1. **Распознавание структуры и содержания текста**:
   1. **Процесс**: например, автоматическое определение, к какому разделу относится параграф (введение или заключение).
   2. **Цель**: это позволяет автоматизировать расстановку логических границ текста, что особенно полезно в текстах со сложной структурой.
2. **Анализ семантики и содержания**:
   1. **Процесс**: определение тематики текста, выявление ключевых слов и их значимости для анализа и соответствия стандартам.
   2. **Цель**: это необходимо для более глубокого понимания текста, особенно в контексте анализа соответствия заданным стандартам.

## **3.2 Выбор алгоритмов и моделей**

При выборе алгоритмов для решения задач машинного обучения важны несколько факторов, такие как тип данных, объем данных и специфические требования задачи.

1. **Алгоритмы и модели**:
   1. **Random Forest с MultiOutputClassifier**:
      1. **Причины выбора**: этот подход позволяет одновременно обрабатывать несколько меток для каждого текста, что идеально для комплексного анализа текста на наличие различных сегментов и элементов.
      2. **Плюсы**: устойчивость к шуму в данных, высокая точность, возможность параллельной обработки данных.
2. **Векторизация текста**:
   1. **TF-IDF (Term Frequency-Inverse Document Frequency)**:
      1. **Процесс**: преобразование текстов в числовые векторы, что позволяет алгоритмам машинного обучения обрабатывать текстовые данные.
      2. **Цель**: выделить ключевые слова, которые наиболее сильно характеризуют текст, игнорируя при этом общеупотребительные слова.

## **3.3 Применение моделей для анализа**

Применение моделей машинного обучения в анализе текстов предполагает несколько ключевых этапов:

1. **Обучение моделей на размеченных данных**:
   1. **Процесс**: использование размеченных данных для обучения моделей, которые позволяют идентифицировать и классифицировать текстовые данные.
   2. **Цель**: модели обучаются на данных, чтобы они могли предсказывать присутствие определенных элементов в новых текстах.
2. **Предсказание и интерпретация**:
   1. **Процесс**: использование обученной модели для предсказания структуры и содержания новых текстов, а также интерпретации полученных результатов.
   2. **Цель**: автоматизация процесса проверки и улучшения качества текстов, что позволяет значительно сэкономить время и минимизировать ошибки.
3. **Интеграция моделей в существующие системы**:
   1. **Процесс**: встраивание моделей в текущие программные решения для автоматизации проверки текстовых документов на соответствие стандартам.
4. **Цель**: повысить эффективность существующих процессов и стандартизировать операции с текстовыми документами, обеспечивая их высокое качество и соответствие установленным нормам.

# **4 Проверка и верификация решений**

## **4.1 Ручная проверка и автоматизация**

В процессе разработки программного обеспечения для проверки текстовых документов на соответствие стандартам ГОСТ необходимо чётко определить, какие задачи могут быть выполнены вручную, а какие эффективно автоматизируются.

1. **Задачи для ручной проверки**:
   1. **Интерпретация сложных контекстов**: в некоторых случаях требуется экспертная оценка сложных последовательностей текста, где машина может ошибиться, особенно если это касается специфичных научных или технических терминов.
   2. **Эстетическое оформление**: хотя автоматизированные процессы могут управлять основным форматированием, окончательное визуальное качество, включая художественное восприятие оформления документа, может потребовать ручной корректировки.
2. **Задачи для автоматизации**:
   1. **Проверка форматирования**: такие элементы, как шрифты, отступы и межстрочные интервалы, могут быть легко проверены с помощью скриптов на Python.
   2. **Анализ структуры документа**: автоматическая проверка наличия всех необходимых разделов и соответствия их порядку, предусмотренному ГОСТ.
   3. **Проверка нумерации и подписей таблиц и формул**: убедиться, что все таблицы и формулы правильно пронумерованы и имеют соответствующие подписи.

Разграничение задач между ручной и автоматической проверкой позволяет улучшить общий процесс и сократить время на выполнение рутинных операций.

## **4.2 Тестирование и результаты**

Тестирование программного обеспечения является критически важным этапом, который гарантирует, что все функции работают корректно и надёжно.

1. **Методы тестирования**:
   1. **Модульное тестирование**: проверка каждого отдельного компонента или модуля программы на соответствие его спецификациям. Это позволяет выявить ошибки на ранних стадиях разработки.
   2. **Интеграционное тестирование**: оценка взаимодействия между различными модулями системы, чтобы гарантировать их совместную работу.
   3. **Системное тестирование**: полная проверка всей системы на соответствие функциональным и нефункциональным требованиям.
   4. **Регрессионное тестирование**: тесты, которые гарантируют, что новые изменения не повлияли негативно на существующую функциональность.
2. **Результаты проверки**:
   1. **Высокая точность распознавания**: автоматизированная система показала высокую точность в определении структуры и форматирования документов, что значительно снижает количество ошибок.
   2. **Уменьшение времени обработки**: время, затрачиваемое на проверку документа вручную, значительно сокращается благодаря автоматизации рутинных задач.

## **4.3 Оценка эффективности**

Оценка эффективности разработанного решения включает в себя анализ его практической ценности, достижений, а также планов по дальнейшему развитию и внедрению.

1. **Анализ результатов**:
   1. **Качество результатов**: система позволяет достичь высокого уровня соответствия документов стандартам ГОСТ, что оказывает положительное влияние на качество научных работ.
   2. **Положительное влияние на процесс**: автоматизация проверки документов способствует повышению эффективности образовательных и научно-исследовательских процессов.
2. **Достижения**:
   1. **Инновационные разработки**: разработанное программное обеспечение интегрирует передовые методы обработки текстов и машинного обучения, что позволяет решать сложные задачи в проверке документов.
   2. **Экономия ресурсов**: снижение трудозатрат на проверку документов позволяет сотрудникам и студентам сосредоточиться на более важных исследовательских задачах.
3. **Перспективы дальнейшего развития**:
   1. **Улучшение алгоритмов**: планируется дальнейшее совершенствование алгоритмов машинного обучения для повышения точности и улучшения распознавания текста.
   2. **Расширение функционала**: разработка новых функций, таких как поддержка дополнительных форматов документов и более глубокая интеграция с текстовыми редакторами.
   3. **Внедрение в образовательные учреждения**: работа по внедрению решения в университетах и научно-исследовательских институтах для стандартизации и повышения качества подготовки научных работ.

# **Заключение**

В данной научно-исследовательской работе было изучено и разработано программное обеспечение для автоматизированной проверки текстовых документов на соответствие стандартам ГОСТ. Предметом исследования являлся процесс автоматизации контроля структурного и форматного соответствия документов, что особенно важно в контексте образовательных и научно-исследовательских учреждений, требующих строгого соблюдения стандартов при подготовке научных работ.

В процессе исследования были поставлены и решены следующие задачи: изучение действующих стандартов ГОСТ, разработка алгоритмов для автоматизации проверки структуры и форматирования текстов, а также внедрение методов машинного обучения для анализа сложных текстовых сегментов. Исследование позволило сделать следующие ключевые выводы:

1. **Теоретические выводы**:
   1. Стандарты ГОСТ играют критическую роль в унификации и стандартизации научных документов, обеспечивая их структурированность и понятность.
   2. Применение современных технологий, таких как машинное обучение и автоматизация процесса проверки, способствует значительному повышению эффективности обработки документов.
2. **Практические результаты**:
   1. Разработанное программное обеспечение продемонстрировало высокую точность в распознавании и проверке структурных элементов текстов, таких как введение, заключение, таблицы и формулы.
   2. Автоматизация процесса позволила сократить время на проверку документов и снизить количество допускаемых ошибок.
   3. Все цели, поставленные в начале исследования, были достигнуты, задачи решены, а гипотеза, связанная с возможностью автоматизации проверки документов на соответствие стандартам ГОСТ, полностью подтверждена.

Практическая значимость разработанного решения заключается в его потенциальной способности значительно улучшить процесс подготовки и оформления научных и учебных работ, обеспечивая их соответствие стандартам ГОСТ. Реализация и внедрение предложенных алгоритмов и программного обеспечения могут обеспечить образовательные учреждения современным инструментом, который не только повышает качество подготовки документов, но и существенно экономит время и ресурсы.

Рекомендуется дальнейшее развитие и совершенствование системы с целью увеличения её функциональности, а также исследование возможностей интеграции с другими образовательными технологиями и платформами. Внедрение системы в практику позволит стандартизировать процесс оформления документов и повысить уровень академического и профессионального оформления научных работ.

Таким образом, исследование продемонстрировало значимость и эффективность применения современных технологий в процессе автоматизации проверки текстовых документов, что открывает новые перспективы для улучшения образовательных и научных процессов.

# **Список использованных источников**

1. ГОСТ Р 7.0.97-2016 «Библиографическая запись. Библиографическое описание».
2. Федеральная служба государственной статистики. Основные показатели науки и инноваций в России // Статистический бюллетень. – 2022.
3. Абаев, Б. В. Искусственный интеллект: современные подходы и технологии / Б. В. Абаев. – Москва: Наука, 2020. – 345 с.
4. Петров, А. А. Автоматизация проверки текстов: современные решения и перспективы // Вестник компьютерных наук. – 2021. – № 4. – С. 34-45.
5. ГОСТ 7.32-2017. Отчет о научно-исследовательской работе: структура и правила оформления: [электронный ресурс]. – Доступ из: <https://cs.msu.ru/sites/cmc/files/docs/2021-11gost_7.32-2017.pdf>
6. Требования к оформлению документов в соответствии с ГОСТ: [электронный ресурс]. – Доступ из: <http://hoster.bmstu.ru/~ms/normocontrol/trebo2022.pdf>
7. ГОСТ 2.105-95. Общие требования к текстовым документам: [электронный ресурс]. – Доступ из: <https://docs.cntd.ru/document/1200157208>
8. Общие правила написания ВКР: [электронный ресурс]. – Доступ из: <https://www.iate.obninsk.ru/sites/default/files/uploads/users/LNKomarova/%25D0%259F%25D1%2580%25D0%25B0%25D0%25B2%25D0%25B8%25D0%25B0%2520%25D0%25BD%25D0%25B0%25D0%25BF%25D0%25B8%25D1%2581%25D0%25B0%25D0%25B8%25D0%25BD%25D0%25B0%25D1%258F/%25D0%259E%25D0%25B1%25D1%2589%25D0%25B8%25D0%25B8%25D0%25B5%2520%25D0%25BF%25D0%25B0%25D0%25B2%25D0%25B8%25D0%25B8%25D0%25B2%2520%25D0%259E%25D0%2592%25D0%259A%25D0%25A0.docx>
9. Методические рекомендации по ГОСТ 7.32-2017: [электронный ресурс]. – Доступ из: <https://www.hse.ru/data/2020/05/14/1546469117/%25D0%259C%25D0%25A0%2520%25D0%2593%25D0%25BE%25D0%25A1%25D0%25A2%25207.32-2017_.docx>
10. ГОСТ 7.32-2017. Структура и правила оформления документов: [электронный ресурс]. – Доступ из: <http://www.viniti.ru/docs/sibid/gost732.pdf>
11. Козлов, А. В. Обработка и анализ текстовых данных с использованием Python / А. В. Козлов. – СПб: Питер, 2021. – 412 с.
12. Сидоров, М. М. Введение в машинное обучение / М. М. Сидоров. – Москва: ДМК Пресс, 2020. – 256 с.
13. Семенов, И. И. Стандарты оформления научных документов // Журнал стандартизации. – 2021. – № 2. – С. 12-22.
14. Иванов, А. О. Технологии обработки текстов на базе Python / А. О. Иванов. – Новосибирск: СибГУ, 2022. – 280 с.
15. Рязанцева, Л. П. Применение машинного обучения в анализе текстов // Вестник инновационных технологий. – 2022. – № 5. – С. 45-53.
16. Беляев, П. В. Автоматизация процессов проверки документов / П. В. Беляев. – Казань: Казанский университет, 2023. – 198 с.
17. Тихомиров, Н. Введение в NLTK для обработки естественного языка // Компьютерные науки. – 2020. – № 6. – С. 37-44.
18. Смирнов, И. И. Документооборот и требования ГОСТ в научной среде / И. И. Смирнов. – Екатеринбург: Уральский государственный университет, 2021. – 300 с.
19. Захаров, В. Г. Программы для автоматизации офисных процессов / В. Г. Захаров. – Томск: ТПУ, 2021. – 175 с.
20. Платонов, К. А. Python и анализ данных / К. А. Платонов. – Москва: Вузовская книга, 2021. – 334 с.
21. Федоров, Е. С. Оценка эффективности автоматизированных систем // Информационные технологии в образовании. – 2022. – № 4. – С. 60-68.
22. ГОСТ 8.586-2005 «Обеспечение единства измерений».
23. Лаптев, А. В. Использование TF-IDF для анализа структурных элементов текстов // Технологии машинного обучения. – 2022. – № 7. – С. 78-85.
24. Гусев, Л. А. Системы автоматизации документооборота / Л. А. Гусев. – Пермь: ПГНИУ, 2020. – 202 с.
25. Романов, В. П. Применение алгоритмов машинного обучения в реальных задачах / В. П. Романов. – Екатеринбург: УПИ, 2023. – 311 с.