|  |
| --- |
| МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ 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

**«**Название работы**»**

|  |  |  |
| --- | --- | --- |
| Выполнил: студент гр. \_\_\_М-Б21\_\_ | (подпись, дата) |  |
| Руководитель ВКР,  к.ф.-м.н., доцент | (подпись, дата) | Царина А.Г. |
| Нормоконтроль | (подпись, дата) |  |
| Руководитель образовательной программы 01.03.02Прикладная 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.

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 предлагает решения для ручной проверки таких элементов, как разделы, и применения моделей машинного обучения для анализа более сложных структур, например, титульных листов.

Область применения результатов охватывает образовательные учреждения и научно-исследовательские организации, которые стремятся 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 2](#_Toc183212140)

[Содержание 3](#_Toc183212141)

[Термины и определения 4](#_Toc183212142)

[Перечень сокращений и обозначений 5](#_Toc183212143)

[Введение 6](#_Toc183212144)

[**Анализ стандартов ГОСТ** 7](#_Toc183212145)

[**Общий обзор стандартов** 7](#_Toc183212146)

[**Требования ГОСТ к структуре документов** 7](#_Toc183212147)

[**Практическое применение стандартов** 7](#_Toc183212148)

[**Работа с текстовыми документами в Python** 7](#_Toc183212149)

[**Инструменты и библиотеки** 7](#_Toc183212150)

[**Преобразование и анализ текста** 7](#_Toc183212151)

[**Автоматизация проверок** 7](#_Toc183212152)

[**Модели машинного обучения для анализа текста** 7](#_Toc183212153)

[**Задачи, требующие машинного обучения** 7](#_Toc183212154)

[**Выбор алгоритмов и моделей** 7](#_Toc183212155)

[**Применение моделей для анализа** 7](#_Toc183212156)

[**Проверка и верификация решений** 7](#_Toc183212157)

[**Ручная проверка и автоматизация** 7](#_Toc183212158)

[**Тестирование и результаты** 7](#_Toc183212159)

[Оценка эффективности 8](#_Toc183212160)

[Заключение 8](#_Toc183212161)

[Список использованных источников 9](#_Toc183212162)

[Приложения 10](#_Toc183212163)

# 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

# Введение

В условиях стремительного развития информационных технологий и цифровизации образовательных процессов все большее значение приобретает соответствие научно-исследовательских и дипломных работ стандартам ГОСТ. Стандартизация не только упрощает обмен научной информацией, но и обеспечивает качество научных документов, делая их более понятными и доступными для широкой аудитории.

Несмотря на очевидные преимущества, соблюдение стандартов ГОСТ остается сложной и ресурсоемкой задачей. Это требует от авторов глубокого понимания множества правил и норм, что становится особенно трудоемким в условиях возрастающего объема научной информации и ускорения темпов образования. В связи с этим актуальной задачей становится разработка автоматизированных средств, способных облегчить и ускорить процесс проверки и оформления документов по ГОСТ.

Целью данного исследования является разработка программы, способной автоматизировать процесс проверки текстовых документов на соответствие стандартам ГОСТ. В рамках работы предполагается изучение существующих стандартов, разработка алгоритмов для анализа текстов и структуры документов, а также применение современных методов машинного обучения для более сложных задач, таких как проверка титульных листов.

Актуальность темы обусловлена необходимостью повышения качества научных документов и их соответствия установленным стандартам. Автоматизация процесса позволит минимизировать количество ошибок, повысить эффективность подготовки документов и сократить временные затраты на их проверку. Это особенно важно в образовательных и научно-исследовательских учреждениях, где большое количество работ требует строгого соблюдения стандартов.

Исследование предполагает изучение теории и практики работы с текстовыми документами, включая использование языка программирования Python и библиотек для обработки файлов формата DOCX. Особое внимание будет уделено анализу, какие элементы документа могут быть проверены вручную, а какие требуют применения алгоритмов машинного обучения. Например, в то время как проверка наличия разделов может быть выполнена по названиям, проверка корректности заполнения титульного листа, возможно, потребует более сложных подходов.

Таким образом, данная работа направлена на создание эффективного инструмента, который облегчит процесс оформления научных работ, повысит их качество и соответствие стандартам, что будет способствовать улучшению образовательного процесса и научных исследований.

# **Анализ стандартов ГОСТ**

## **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»**:
   * 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»**:
   * 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»**:
   * 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»**:
   * 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.

## **Требования ГОСТ к структуре документов**

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

* **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.
* **Содержание**: должно иметь логическую структуру с указанием номеров страниц для каждого раздела, облегчая ориентацию и поиск информации в документе.
* **Основная часть**: включает введение, главы с изложением основных результатов исследований и заключение. Каждая глава должна быть логически организована и, при необходимости, подразделена на параграфы.
* **Список использованной литературы**: оформляется в соответствии с правилами библиографического описания, что упрощает проверку источников информации.
* **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.

## **Практическое применение стандартов**

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

* **Многообразия требований**: каждый стандарт содержит свои специфические требования, которые необходимо учитывать и которые могут быть трудны для восприятия без надлежащей подготовки.
* **Обновлений стандартов**: ГОСТы периодически подвергаются изменениям, что требует постоянного мониторинга и обновления соответствующих практик образовательными учреждениями и отдельными исследователями.
* **Индивидуальных интерпретаций**: различные академические учреждения могут трактовать стандарты по-своему, что создает путаницу у студентов, пытающихся следовать унифицированным правилам.
* **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.

# **Работа с текстовыми документами в Python**

## **Инструменты и библиотеки**

Обзор используемых библиотек для обработки текстовых документов, таких как python-docx.

## **Преобразование и анализ текста**

Описание процессов чтения, изменения и анализа текстовых данных в файлах DOCX.

## **Автоматизация проверок**

Примеры автоматизированной проверки элементов текста на соответствие стандартам ГОСТ.

# **Модели машинного обучения для анализа текста**

## **Задачи, требующие машинного обучения**

Объяснение, какие задачи анализа текста требуют использования моделей машинного обучения.

## **Выбор алгоритмов и моделей**

Описание выбранных алгоритмов и подходов для анализа текста и почему они были выбраны.

## **Применение моделей для анализа**

Как модели применяются для анализа текстов в документах и улучшения качества автоматизированной проверки.

# **Проверка и верификация решений**

## **Ручная проверка и автоматизация**

Разграничение между задачами, которые могут быть выполнены вручную, и теми, которые автоматизируются.

## **Тестирование и результаты**

Описание методов тестирования программного решения и результатов проверки.

## Оценка эффективности

Анализ результатов, достижения и перспективы дальнейшего развития и внедрения разработанного решения.

# Заключение

# 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%25BB%25D0%25B0%2520%25D0%25BD%25D0%25B0%25D0%25BF%25D0%25B8%25D1%2581%25D0%25B0%25D0%25BD%25D0%25B8%25D1%258F/%25D0%259E%25D0%25B1%25D1%2589%25D0%25B8%25D0%25B5%2520%25D0%25BF%25D1%2580%25D0%25B0%25D0%25B2%25D0%25B8%25D0%25BB%25D0%25B0%2520%25D0%25BD%25D0%25B0%25D0%25BF%25D0%25B8%25D1%2581%25D0%25B0%25D0%25BD%25D0%25B8%25D1%258F%2520%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>

# Приложения