

УДК 004.912

DOI: 10.22213/2410-9304-2025-2-93-103

## Современные подходы к оптимизации технической документации: вызовы и решения для беспилотных авиационных систем

Е. В. Исаева, кандидат филологических наук, доцент, зав. кафедрой английского языка профессиональной коммуникации, Пермский государственный национальный исследовательский университет, Пермь, Россия

А. В. Ходасевич, студент, Национальный Исследовательский Технологический Университет МИСИС, Москва, Россия

Я. М. Исаева, студент, Пермский государственный национальный исследовательский университет, Пермь, Россия

*В статье рассматривается вопрос автоматизации в реструктуризации технической документации, необходимой для сертификации беспилотных авиационных систем (БАС). Сертификация – важнейший нормативный процесс, обеспечивающий соответствие национальным и международным стандартам безопасности. Подготовка технической документации для сертификации является трудоемким, сопряженным с ошибками и дорогостоящим процессом, при этом ручные методы значительно снижают эффективность. Эти недостатки создают проблемы для массовой сертификации БАС, которая становится все более актуальной в связи с быстрым ростом отрасли. В статье рассматриваются современные технологии, которые могут быть использованы для решения этих задач, включая машинное обучение, обработку естественного языка (NLP), графы знаний и семантический анализ. Рассматриваются такие инструменты, как EasyOCR, OnToCode, и продвинутые модели, такие как BERT, с точки зрения их пригодности для распознавания текста, извлечения данных и автоматической реструктуризации сложных технических документов. Кроме того, в исследовании приведены примеры возможного практического применения этих методов, сделан акцент на их способности стандартизировать подготовку документов, уменьшить количество человеческих ошибок и поддержать массовую сертификацию БАС. Также обсуждаются такие проблемы, как отсутствие унификации форматов документов, интеграция устаревших систем и динамичность норм регулирования. В исследовании делается вывод о том, что, несмотря на открывающиеся возможности автоматизации для оптимизации процессов сертификации, ее практическая реализация требует дальнейшего изучения и сотрудничества с регулирующими органами. Предлагаемый обзор служит концептуальной основой, подчеркивающей потенциальные преимущества и ограничения интеграции автоматизации в рабочие процессы сертификации БАС и открывающей путь для будущих достижений в этой критически важной области.*

**Ключевые слова:** беспилотные авиационные системы, стандартизация БАС, автоматизация, сертификация БАС, техническая документация, реструктуризация текста, машинное обучение, графы знаний, обработка естественного языка.

### Введение

Сертификация беспилотных авиационных систем (БАС) – важнейший регуляторный процесс, направленный на обеспечение безопасности, надежности и соответствия национальным и международным стандартам. В соответствии с Федеральными авиационными правилами (<https://favt.gov.ru/dokumenty-federalnye-pravila/>) и другими нормативными документами, такими как Воздушный кодекс Российской Федерации (№ 60-ФЗ 1997, редакция от 08.08.2024), Нормы летной годности (<https://favt.gov.ru/sertifikaciya-avia-tehniki-serti-fikaciya-tipa-avia-tehnika-zakony-pravila-dokumenty/?id=5498>), Приказ Минтранса РФ от 18.10.2024 № 367, сертификация является обязательной для БАС, предназначенных для коммерческой эксплуатации и работ, связанных с авиацией. Этот процесс также распространяется на организации, ответственные за обслуживание сертифицированных БАС или их компонентов.

Процесс сертификации требует подготовки и представления большого количества документов, включая технические спецификации, чертежи общего вида, сертификационные базисы и оценки соответствия. Традиционным способом формирования сертификационного досье является «ручной» сбор и форматирование документации: «В настоящий момент поиск и проверка требований нормативно-технической документации выполняется вручную, что влечет существенное количество ошибок» [1, р. 256]. Этот метод является трудоемким, поскольку требует значительных человеческих усилий и времени. Каждый этап подготовки к сертификации БАС сопряжен с задержками и дополнительными расходами из-за возможных ошибок в документации, таких как несоответствие форматирования, отсутствие данных или отклонение от нормативных стандартов. Любые выявленные проблемы могут привести к возврату

документов на доработку, что увеличивает сроки и финансовую нагрузку на заявителей. Такая неэффективность создает значительные препятствия для массовой сертификации БАС, замедляя развитие отрасли и инноваций.

Для упрощения процесса подготовки технической документации следует обратиться к автоматизированным подходам. Автоматизация поможет снизить риск человеческой ошибки, обеспечить соответствие установленным стандартам и минимизировать время и ресурсы, затрачиваемые на внесение изменений.

Мы нацелены рассмотреть существующие технологии автоматизации реструктуризации технической документации с позиции перспективы их применения для упрощения задач сертификации БАС.

#### **Методы для решения отдельных задач автоматической реструктуризации технической документации**

Реструктуризация документов предполагает извлечение информации и преобразование документов из различных источников и форматов в заранее определенную структуру. Это необходимо для облегчения ориентации и поиска информации в тексте, упрощения аналитики и извлечения точных данных. Структурирование документов подразумевает использование стандартов документации, классификаций и стандартизированных терминологий [2]. В настоящее время для решения этой задачи все чаще используются методы обработки естественного языка и машинного обучения для анализа текстов. В данном обзоре мы сфокусируемся на таких методах, доступных для реализации посредством языка Python.

Одним из перспективных подходов является система поиска требований и ассоциации документов (RRDA), которая была успешно применена в строительной отрасли для извлечения неявных требований, связанных с безопасностью, из проектной документации. Эта методология может быть адаптирована для беспилотных авиационных систем (БАС) с целью повышения соответствия сертификационным требованиям. Система RRDA объединяет два ключевых метода NLP – семантический анализ текста и анализ тональности (Sentiment-анализ) – для выявления сегментов, связанных с требованиями, в неструктурированных документах. В частности, она использует ассоциативные правила «требование-документ» для сопоставления технических документов с извлеченными требованиями, обеспечивая высокую степень соответствия между нормативными

требованиями и содержанием документов. RRDA состоит из нескольких ключевых этапов: Предварительная обработка (стандартизация форматов текста, удаление стоп-слов и лемматизация), извлечение требований (идентификация сущностей, поведения и объектов требований с помощью алгоритмов анализа текста), ассоциация документов (применение ассоциативных правил для связи извлеченных требований с соответствующими разделами технической документации), проверка и оптимизация (сравнение извлеченных требований со структурированными базами данных для более точного извлечения). Система продемонстрировала точность более 91% [3].

Аналогичная задача решается в исследовании, посвященном автоматическому распознаванию требований к конструкции из текстов строительных контрактов, которые могут включать как существенную, так и вспомогательную информацию, что требует «ручной вычитки». Была разработана модель классификации сегментов текста, относящихся и не относящихся к формальным требованиям, с точностью 95% (F1-мера). В работе представлены алгоритмы предобработки текста (удаление стоп-слов, лемматизация, анализ n-грамм) и метода опорных векторов для реализации классификации [4]. Модель может быть экстраполирована на нашу область в том числе для идентификации требований Сертификационного базиса.

Для анализа текстов документации формата PDF требуется технология OCR (optical character recognition, оптическое распознавание символов) – технология автоматизации извлечения данных из PDF-файла с преобразованием текста в машиночитаемую форму. Технология OCR позволяет работать как с печатным текстом, так и письменным, с отсканированными страницами документов, с изображениями [5]. Для автоматизации этого процесса есть целый комплекс доступных эффективных решений, таких как Tesseract, Easy OCR, Paddle OCR и др.

Tesseract (<https://github.com/tesseract-ocr/tesseract>) – это широко используемый механизм OCR с открытым исходным кодом, который постоянно развивается с момента своего появления в 1985 году. Как указано в описании проекта, он поддерживает более 100 языков и содержит такие полезные модули, как анализатор макета и анализ связанных компонентов для определения положения слов в печатном тексте. Для распознавания символов Tesseract использует комбинацию традиционных алгоритмов обработки изображений и машинного обучения. В версии 4 в Tesseract появился механизм OCR

на основе сети Long Short-Term Memory (LSTM), который обрабатывает входные изображения построчно. Каждая строка делится на блоки символов, которые затем поступают в LSTM-сеть для распознавания. Такой подход повысил точность распознавания текста, особенно больших фрагментов.

Tesseract нашел применение в различных предметных областях. В публикациях представлены примеры его использования для распознавания медицинских текстов на медицинских рецептах [6] и этикетках лекарств [7, 8], обработки счет-фактур [9] и преобразования отсканированных таблиц в структурированные файлы Excel [10].

Однако, хотя Tesseract эффективно распознает обычный и высококонтрастный печатный текст, его производительность снижается в сценариях с нерегулярными шрифтами, вертикальным текстом, низкоконтрастным фоном или изогнутыми поверхностями. В сравнительном исследовании с моделью на основе глубокого обучения (R-CNN) Tesseract показал более низкую точность (Precision = 38,3%, Recall = 61,3%, F1 Score = 47,1) при распознавании текстовых символов, напечатанных различными методами печати на изогнутой фармацевтической упаковке [8].

Несмотря на эти недостатки, Tesseract остается популярным выбором для задач OCR благодаря простоте использования, доступности с открытым исходным кодом и широкой языковой поддержке. При реструктуризации технической документации Tesseract можно эффективно использовать для простых задач извлечения текста или в сочетании с другими моделями глубокого обучения для повышения точности.

Easy OCR (<https://github.com/JaidedAI/EasyOCR>) – это библиотека с технологией OCR с открытым исходным кодом, поддерживающая более 80 языков, в том числе русский. Easy OCR обладает рядом преимуществ: простота в использовании (Easy OCR разработан таким образом, чтобы быть удобным для пользователя, и предоставляет простой API, который позволяет пользователям выполнять задачи распознавания текста с минимальными затратами кода и конфигурации), глубокое обучение (Easy OCR построен на базе алгоритмов глубокого обучения; при выполнении обнаружения символов используется алгоритм CRAFT – Character Region Awareness For Text detection – распознавание области символов для распознавания текста [11]; модель распознавания использует алгоритмы глубоких сверточных нейронных сетей –

CRNN, доказавшие высокую эффективность в тестах распознавания текста [12]), быстрота и эффективность (программа Easy OCR оптимизирована для повышения скорости и эффективности, что позволяет быстро обрабатывать изображения и извлекать текст, что делает ее подходящей для приложений реального времени и крупномасштабных задач распознавания текста). При их сравнении многие исследования показывают превосходство Easy OCR перед Tesseract, заключающееся в скорости обработки текста и более точных результатах в обработке текста с «шумом», таким как непроизвольное искривление отображения данных, заключающееся в особенностях реального мира [13]. [https://www.researchgate.net/publication/375535379-Gradual\\_OCR\\_An\\_Effective\\_OCR\\_Approach\\_Based\\_on\\_Gradual\\_Detection\\_of\\_Texts](https://www.researchgate.net/publication/375535379-Gradual_OCR_An_Effective_OCR_Approach_Based_on_Gradual_Detection_of_Texts).

PaddleOCR (<https://github.com/PaddlePaddle/PaddleOCR>) – это современная система оптического распознавания символов, использующая методы глубокого обучения. Она была разработана для высокоточного обнаружения и распознавания текста на разных языках и типах документов. В отличие от традиционных систем OCR, таких как Tesseract, PaddleOCR использует модели глубокого обучения, такие как сверточные рекуррентные нейронные сети (CRNN – Convolutional Recurrent Neural Networks) и трансформеры для распознавания текста на сложном фоне (SVTR – Scene Text Recognition Transformers [14]. Согласно последним исследованиям PaddleOCR демонстрирует устойчивость к шумам и искажениям текста и изображения [15]. При сравнении PaddleOCR и Tesseract для оптического распознавания русских символов было выявлено, что более высокую точность показывает Tesseract (98% против 93% у PaddleOCR), но PaddleOCR имеет более компактную модель (2 МБ против 23 МБ у Tesseract) и более высокую скорость обработки (2,07 секунды против 3,8 секунды на изображение) при использовании GPU. Таким образом, авторы делают вывод, что у обеих библиотек есть потенциал применения для оптического распознавания текстов на русском языке [16].

Для преобразования текстов из неструктурированных форматов, удобных для человеческого восприятия информации (PDF, Excel, HTML), в машиночитаемый формат (JSON) применяются современные методы, представляющие собой интегрированный пакет библиотек Python. Пакет включает в себя функциональные модули для извлечения, реструктуризации, обработки и переноса данных. Для импорта содержания

и метаданных веб-страниц, извлечения заголовков, текста и таблиц, сохраненных в HTML, используются специальные библиотеки, такие как Scrapy (<https://scrapy.org>), Selenium (<https://www.selenium.dev>), BeautifulSoup (<https://www.crummy.com/software/BeautifulSoup>), Html Unit (<https://htmlunit.sourceforge.io>) и Jsoup (<https://jsoup.org>), широко используемые в различных исследованиях и приложениях [17–19]. BeautifulSoup, например, отличается высокой эффективностью при низкой загрузке памяти и процессора, что делает библиотеку наиболее подходящей для более простых задач [20]. Для извлечения текста и метаданных из PDF-файлов – библиотеки: PyPDF2 (<https://pypdf2.readthedocs.io/en/3.x>), pdfminer (<https://pdfminersix.readthedocs.io/en/latest>), PyMuPDF (<https://pymupdf.readthedocs.io/en/latest>) и pdfplumber ([https://python.langchain.com/docs/integrations/document\\_loaders/pdfplumber](https://python.langchain.com/docs/integrations/document_loaders/pdfplumber)); для графических PDF – OCR (например, EasyOCR); для чтения содержимого ячеек таблицы Excel и преобразования их в атрибутно-значимые пары – библиотека xlrd (<https://xlrd.readthedocs.io/en/latest>). Извлеченный контент преобразуется в формат JSON с использованием структуры «ключ-значение»; табличные данные сохраняются в виде вложенных структур, а данные с подкатегориями сохраняются в виде иерархических структур. На этапе обработки данных устраняются ошибки форматирования, такие как избыточные пробелы, неправильный порядок слов; данные визуализируются, например, с помощью облака слов для просмотра ключевых слов. Далее структурированные данные могут быть сохранены в базах данных [21].

Примером реструктуризации специализированных документов, сфокусированной на извлечении и сохранении знаний, может послужить статья «A Knowledge Graph Approach towards Re-structuring of Scientific Articles» [22]. В работе представлен метод, основными вехами которого являются создание графов знаний на основе извлеченных ключевых слов, кластеризация графов, сортировка, приоритизация и визуализация данных. Сначала с помощью NLP-алгоритмов и инструментов, таких как Babel Net (<https://babelnet.org/guide>), из текста извлекаются ключевые слова, которые становятся узлами графа знаний, а их связи, например семантическая близость, совместная встречаемость, – ребрами графа. Веса ребер определяются в зависимости от частоты совместной встречаемости ключевых слов. Далее узлы графа группируются в кластеры на основе метрик сходства, отра-

жающих концептуальные взаимосвязи между ключевыми словами. Количество кластеров регулируется методом OClustR (<https://github.com/maurodl/OClust-R>), который создает пересекающиеся кластеры и за счет модульности оптимизирует их количество [23]. Модульность понимается как матрица, которая оценивает плотность разделения сети на модули и используется для измерения структуры сетей или графов. Сети с высокой степенью модульности считаются кластерами с плотными связями, а сети с низкой степенью модульности – сетями с разреженными связями. Модульность рассчитывается для каждого кластера, чтобы сократить число кластеров до 20. Полученные кластеры упорядочиваются по релевантности, а ключевые слова внутри кластеров сортируются по их значимости в тексте источнике.

После извлечения знаний и представления их в виде графов или онтологий данные должны быть обратно зашифрованы и приведены к шаблонному виду. Автоматизация редактирования текстов с помощью онтологически-управляемых систем «обеспечивает реализацию таких процессов, как структурирование и систематизация информации, интеграция распределенных информационных моделей и систем на основе использования семантических свойств, агрегация различных информационных ресурсов, визуализация необходимой информации» [24, с. 166]. В данном контексте стоит обратить внимание на программу OnToCode (<https://gitlab.com/dlr-dw/ontocode>), которая использует шаблонно-ориентированную генерацию кода из онтологий, как кросс-платформенных источников знаний. Программа OnToCode была разработана в рамках исследований по созданию автоматизированной платформы для передачи данных о компонентах космических аппаратов. Исследовательский коллектив разработал онтологию, фиксирующую всю формальную спецификацию деталей, которая потом закладывалась в основу программных прототипов. Таким образом, в качестве примера использования данной программы можно привести создание формальных моделей спецификаций летательных аппаратов для этапа проектирования [25].

Реструктуризации текста может потребовать сравнения схожести исходного или полученного документа с требуемым шаблоном. В этом случае полезным представляется метод MASSAlign, позволяющий выравнивать и аннотировать сравниваемые документы. На этапе выравнивания текстов с помощью методов Bag-of-Words и TF-IDF создается матрица схожести

для абзацев или предложений. Алгоритм ищет оптимальный способ выравнивания отрезков текста, начиная с самого подходящего исходного пункта в матрице (с порогом схожести  $\alpha \geq 0,2$ ), итеративно оценивая соседние точки и выбирая наиболее подходящие пары.

Выравненные предложения далее анализируются с помощью синтаксических деревьев для определения операций преобразования текста, таких как удаление избыточных слов и фраз, замена на синонимы или аналогичные по смыслу выражения, добавление новых слов и изменения порядка слов [26]. Такое выравнивание может быть использовано для создания параллельных корпусов на уровне абзацев и предложений, которые, как мы полагаем, могут быть адаптированы для генерации специализированных данных для обучения моделей реструктуризации текста с помощью алгоритмов машинного обучения.

Автоматизация форматирования документов может быть достигнута с помощью программного обеспечения и методов, описанных в статье «Автоматическая верстка и оформление научной и программной документации» [27]. В статье представлен функционал инструмента GOSTdown (<https://gitlab.iaaras.ru/iaaras/gost-down>) для автоматической верстки и оформления научной и программной документации в соответствии с ГОСТ 7.32 и ГОСТ 19, в том числе нумерация глав, таблиц, рисунков и формул, автоматическое создание оглавления и библиографии. В основе разработки лежат язык разметки Markdown (<https://www.markdownguide.org>), программное обеспечение для создания форматированных библиографических списков – BibTeX (<https://www.bibtex.org>) и универсальный конвертер Pandoc (<https://pandoc.org>), обеспечивающий такие функции, как автоматическое связывание, создание библиографии и совместная работа. Использование этих инструментов позволяет автоматизировать оформление и форматирование документов и сократить время редактирования [28]. В программе используются технологии разметки текста Pandoc Markdown: генерация текстовых файлов в формате .docx обеспечивается использованием шаблонов DOCX для построения колонтитулов и стилей; управление документом производится через скрипты на Powershell для постобработки текста с использованием объектной модели компонентов – COM (Component Object Model). Представленный алгоритм позволяет добиться автоматического форматирования документа в соответствии с ГОСТ.

Полагаем, что этот алгоритм может быть адаптирован для форматирования технической документации в соответствии с другими заданными параметрами.

На решение аналогичной задачи направлена система Doxygen (<https://www.doxygen.nl>), предназначенная для генерации технической документации на основе исходного кода и приведения документации к требованиям ГОСТ. Инструмент генерирует иерархическую структуру документации, включая описание файлов, классов, методов и переменных, а также визуализирует зависимости между компонентами с помощью встроенного пакета Graphviz [29]. Несмотря на то что программа Doxygen написана не на языке Python, его архитектура может быть использована для разработки алгоритмов автоматизации обработки технической документации для реализации с использованием Python-библиотек. Интеграция Doxygen с системами контроля версий Git, позволяет автоматически обновлять документацию при изменении исходных кодов. Кроме того, Doxygen предусматривает перекрестные ссылки между документацией и кодом, чтобы читатель документа мог легко перейти к фактическому коду, что является преимуществом перед такими аналогами, как Sphinx (<https://www.sphinx-doc.org/en/master>), Pydoc (<https://docs.python.org/3/library/pydoc.html>), ReadtheDocs (<https://docs.readthedocs.com/platform/stable/index.html>). Версирование документации может быть полезно при подготовке к сертификации и формировании документов с учетом динамики отрасли БАС и национальных стандартов. Таким образом, представленный подход демонстрирует потенциал для применения в задачах реструктуризации и стандартизации документов новой высокотехнологичной отрасли, но требует адаптации для учета особенностей сертификационных процессов Росавиации.

#### **Практическое применение: задачи сертификации БАС**

В момент написания данной статьи мы наблюдаем достаточно сложную ситуацию с сертификацией БАС. В ходе опроса разработчиков и производителей беспилотных аппаратов, участвовавших в проектно-образовательных интенсивах «Архипелаг» в 2023 и 2024 г. (<https://архипелаг2035.рф>), мы выяснили, что далеко не все респонденты официально регистрируют свои разработки. При этом ужесточаются требования к наличию сертификационных документов на «беспилотные авиационные системы и (или) их элементы, за исключением беспилотных авиационных систем ... с максимальной взлетной массой 30 килограммов и менее»

(см. Воздушный кодекс Российской Федерации, № 60-ФЗ 1997, редакция от 08.08.2024). В соответствии с законодательством РФ БАС, предназначенные для коммерческих перевозок или авиационных работ подлежат обязательной сертификации, а юридические лица и индивидуальные предприниматели, осуществляющие техническое обслуживание БАС и их элементов, подлежащих обязательной сертификации, обязаны получить соответствующий сертификат (Приказ Минтранса РФ от 18.10.2024 № 367). Процесс сертификации строго регламентирован Федеральными авиационными правилами (<https://favt.gov.ru/dokumenty-federalnye-pravila>) и Нормами летной годности.

Данный процесс включает в себя несколько этапов: подготовка заявки и документации по БАС (сопровождается спецификацией БАС, чертежом, проектом сертификационного базиса и др.), подготовка макета БАС (анализ сертификационного базиса, выбранных методик оценки со-

ответствия и перечня комплектующих изделий), проведение и документирование испытаний (доказательная документация формируется на основе аналитики, расчетов и результатов испытаний, включая стендовые, наземные и летные тесты, проводимые под надзором экспертов рабочей группы и сертификационного центра), получение сертификата типа (удостоверяет соответствие изделия требованиям Сертификационного базиса, сопровождается протоколами испытаний, спецификацией, эксплуатационной документацией и информацией о держателе Сертификата типа, типовой конструкции, Сертификационном базисе, условиях и ограничениях, при выполнении которых обеспечивается соответствие изделия требованиям Сертификационного базиса, указанного в Сертификате типа) (Приказ Министерства транспорта Российской Федерации от 17.06.2019 № 184. 2019). Рисунок иллюстрирует типовые формы для заполнения при подготовке к сертификации БАС.

СЕРТИФИКАТ  
ЛЕТНОЙ ГОДНОСТИ БЕСПИЛОТНОГО ГРАЖДАНСКОГО ВОЗДУШНОГО СУДНА  
С МАКСИМАЛЬНОЙ ВЗЛЕТНОЙ МАССОЙ БОЛЕЕ 30 КИЛОГРАММОВ

РОССИЙСКАЯ ФЕДЕРАЦИЯ  
МИНИСТЕРСТВО ТРАНСПОРТА РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ АГЕНТСТВО ВОЗДУШНОГО ТРАНСПОРТА  
СЕРТИФИКАТ ЛЕТНОЙ ГОДНОСТИ  
БЕСПИЛОТНОГО ГРАЖДАНСКОГО ВОЗДУШНОГО СУДНА С МАКСИМАЛЬНОЙ  
ВЗЛЕТНОЙ МАССОЙ БОЛЕЕ 30 КИЛОГРАММОВ  
N \_\_\_\_\_

1. Тип и назначение беспилотного гражданского воздушного судна	2. Национальный и регистрационный знаки	3. Серийный (заводской) номер, идентификационный номер
4. Беспилотное гражданское воздушное судно принадлежит (полное наименование юридического лица, фамилия, имя, отчество (при наличии) физического лица, в том числе индивидуального предпринимателя); адрес юридического лица в пределах места нахождения, адрес регистрации по месту жительства (пребывания) физического лица, адрес регистрации по месту жительства индивидуального предпринимателя); идентификационный номер налогоплательщика)		
5. Эксплуатационные ограничения и характеристики гражданского беспилотного воздушного судна (наименование) содержатся в карте данных (приложение к настоящему сертификату)		
Должность, подпись, фамилия, инициалы ответственного лица уполномоченного органа Дата выдачи: _____ МП		

КАРТА  
ДАННЫХ БЕСПИЛОТНОГО ГРАЖДАНСКОГО ВОЗДУШНОГО СУДНА  
С МАКСИМАЛЬНОЙ ВЗЛЕТНОЙ МАССОЙ БОЛЕЕ 30 КИЛОГРАММОВ

1. Двигатель (двигатели) (марка, заводской номер)	
мощность, кВт	
максимальные обороты (количество оборотов в минуту)	
2. Воздушный винт (марка, заводской номер)	
3. Топливо	
4. Максимальная взлетная масса беспилотного гражданского воздушного судна, кг	
5. Центровка:	
предельно-передняя, %	
предельно-задняя, %	
6. Грузоподъемность, кг	
7. Ограничения по скорости:	
максимально допустимая, км/ч	
минимально допустимая, км/ч	
скорость отрыва при взлете, км/ч	
посадочная скорость, км/ч	
8. Состав экипажа, чел.	
9. Пункт управления беспилотного гражданского воздушного судна	
10. Метеоусловия для выполнения полетов:	
высота нижнего края облаков, м	
горизонтальная видимость, м	
ветер:	

а) на взлете:	
встречный, м/с	
попутный, м/с	
под 90 градусов, м/с	
б) при посадке:	
встречный, м/с	
попутный, м/с	
под 90 градусов, м/с	
11. Температура окружающего воздуха, градус С°:	
12. Запрещены полеты:	
(эксплуатационные ограничения).	
Разрешены полеты:	
(днем, ночью, правила визуальных полетов, правила полетов по приборам)	
Должность, подпись, фамилия ответственного лица уполномоченного органа Дата выдачи: _____ МП	

Пример типовых форм сертификации БАС  
Example of typical forms for UAS certification

Этот трудоемкий процесс может осложняться и затягиваться ввиду ошибок на разных этапах работы с документацией. На этапе подготовки заявки и документации ошибки, связанные с несоответствием формата требованиям, возникающие в связи с «человеческим фактором», могут быть минимизированы с помощью систем автоматизации для генерации стандартизированных документов (например, спецификаций, чертежей и базиса). На этапе проверки и отслеживания документации внедрение программ для автоматизированной проверки соответствия Федеральным авиационным правилам и нормам летной годности упростит анализ сложной технической документации.

Приведем наше видение того, как автоматизированные инструменты помогают реинжинирингу технической документации в соответствии с регуляторными стандартами. Реструктуризация технической документации БАС предполагает прохождение трех этапов работы с текстом: дешифрование документа (распознавание текста в человекочитаемом формате и его преобразование в машиночитаемый формат), извлечение информации и преобразование ее в знания, шифрование документа, включающего полученные знания, в документ с новой заданной структурой в человекочитаемом формате. На этапе дешифрования важно учитывать, что техническая документация в области БАС вклю-

чает мультимодальные документы со сложными структурированными данными, такими как таблицы, схемы, таксономии, тексты с большим количеством технических терминов и аббревиатур. Для автоматизации анализа, распознавания, обработки и классификации этих документов наиболее удобными инструментами представляются EasyOCR и PaddleOCR. Эти инструменты позволяют интегрироваться с системами семантического анализа, что помогает автоматизировать проверку на соответствие стандартам и нормативам, ускоряя процесс подготовки документации. Встроенные алгоритмы классификации и предобработки минимизируют распространенные ошибки, такие как дублирование или пропуск данных. Благодаря гибкости (возможность работы с различными форматами документов), масштабируемости (интеграция в проекты разного масштаба, включая крупные сертификационные процессы) и простоте использования (удобный интерфейс), EasyOCR и PaddleOCR способны стать основой для построения специализированных систем автоматизации подготовки и проверки технической документации в сфере БАС, что повышает общую эффективность и снижает затраты на сертификацию.

В качестве альтернативного решения можно использовать библиотеки Python для извлечения данных из разных форматов (например, технические отчеты в PDF или таблицы спецификаций в Excel) и преобразования их в формат JSON и последующее сохранение в виде унифицированных баз данных с четкой структурой. Преобразованные данные могут быть интегрированы с технологией графов знаний, в которых узлами являются технические термины, а связи отражают их взаимоотношения. Такое представление данных позволит упростить поиск информации при подготовке стандартизированных документов для сертификации моделей БАС. Кроме того, JSON-структуры могут быть интегрированы с инструментами NLP-анализа для автоматической проверки соответствия нормативным требованиям.

Этап извлечения информации и знаний может быть реализован с применением методов создания графов знаний, позволяющих извлекать ключевые термины из технической документации (спецификации, стандарты и нормативы). На основе графа знаний может быть построена онтология, которая связывает термины и их атрибуты, такие как взаимозависимости между элементами системы. Последующая кластеризация графа позволит выделить домены или предметные области (напри-

мер, аэродинамика, электроника) и поддомены (подсистемы/подразделы предметных областей, которые могут указывать на соответствующие разделы технической документации), что упростит анализ и реструктуризацию сложной документации. Сортировка и приоритизация данных обеспечат доступ к наиболее релевантной информации, которая не должна быть утеряна при реструктуризации документа.

На этапе шифрования данных может использоваться онтологический подход для извлечения отраслевых терминов через SPARQL-запросы и стандартизированного употребления терминологии БАС в технической документации. Шаблонная генерация с помощью OneToCode и Jinja2 позволит создать коды для автоматической обработки документации, включая анализ, настройку шаблонов, адаптацию к регуляторным требованиям и проверки соответствия нормативам. Рассматриваемое программное обеспечение обеспечит расширяемость и кроссплатформенность разработанных программных компонентов.

После шифрования текста может потребоваться приведение документа к определенной размерности (например, ограничение блоков информации по длине), сравнение полученного документа с заданным эталоном и аннотирование изменений между версиями для минимизации несоответствий в документации. В этом случае полезным будет метод выравнивания и аннотирования текста. Алгоритмы аннотирования могут автоматически выявлять дублирующиеся, избыточные или не соответствующие требованиям элементы в текстах, помогая адаптировать документацию под регуляторные стандарты. Данный алгоритм может быть встроен в систему реструктуризации документов, чтобы обеспечить контроль качества и отслеживание изменений. Как мы писали выше, полученные аннотации и выравнивания могут быть использованы для обучения моделей обработки естественного языка, направленных на автоматизацию реструктуризации и проверки текстов.

На завершающем этапе форматирования сертификационной документации БАС в стандартизированном формате возможно применение адаптированного алгоритма GOSTdown, программного комплекса с открытым программным кодом. Данный инструмент позволит управлять автоматической нумерацией, оглавлением и библиографией, написанием формул, форматированием рисунков и др., поддерживает принцип версионирования разработки через Git, исключает ручное форматирование, снижая

риск несоответствий в оформлении, что важно для сертификации.

Дополнительно могут быть предложены формальные методы для автоматизированного моделирования сценариев, анализа рисков и повышения прозрачности процесса доказательства соответствия требованиям, рассмотренные в статье «Formal Certification Methods for Automated Vehicle Safety Assessment» [30]. Интеграция формальных методов позволяет автоматизировать подготовку документации на разных этапах подготовки сертификационных документов, в том числе на этапе подготовки заявки и документации (определение спецификации, подготовка чертежей и проекта сертификационного базиса), подготовки макета (анализ выбранных методик оценки соответствия и перечня комплектующих изделий), проведения и документирования испытаний (систематизация доказательной документации на основе аналитики, расчетов и результатов стендовых, наземных и летных тестов), подтверждения соответствия изделия сертификационному базису.

#### Обсуждение перспектив и вызовов

Автоматизация подготовки сертификационной документации открывает новые возможности для повышения точности документов за счет минимизации человеческих ошибок при структурировании, форматировании и проверке документов на соответствие требованиям. Автоматизация поддерживает одновременную обработку нескольких документов, что обеспечивает эффективное решение для массовой сертификации БАС, которая становится все более актуальной в связи с быстрым развитием отрасли. Кроме того, технология может быть масштабирована на другие отрасли, для которых актуален вопрос сертификации.

Тем не менее нельзя не учитывать сложности, которые могут возникнуть при разработке интегрированного инструмента автоматизации подготовки технической документации. Проблемы могут быть вызваны отсутствием унификации исходных данных, динамичностью и разноплановостью сертификационного процесса, необходимостью последующего регулярного обслуживания автоматизированной системы, невозможностью полностью отказаться от экспертной корректировки документов.

Чтобы справиться с этими трудностями, необходимы дальнейшие исследования и разработки в направлении адаптивности систем, удобства интерфейса и операционной совместимости. Полезным будет взаимодействие с регуляторными органами и разработчиками в рам-

ках отраслевых форумов и выставок, а также совместных мероприятий, которые проводятся такими организациями, как Национальная технологическая инициатива (<https://nti2035.ru>) и Агентство стратегических инициатив (<https://asi.ru>), где уместным будет обсудить роль единообразия форматов данных и структуры документов для достижения эффективности автоматизированных решений.

#### Заключение

В статье рассмотрены теоретические основы и существующие технологии, которые могут способствовать автоматизации подготовки сертификационной документации БАС. Несмотря на перспективность этих методов, их эффективность в оптимизации процессов сертификации остается предметом дальнейших исследований и практической проверки.

Дальнейшие исследования должны быть направлены на практическое внедрение и изучение конкретных примеров, чтобы оценить влияние этих технологий в реальном мире. Сотрудничество с заинтересованными сторонами в отрасли и регулирующими органами также будет иметь важное значение для обеспечения соответствия автоматизированных решений существующим стандартам и удовлетворения конкретных потребностей сертификации БАС.

В заключение следует отметить, что автоматизация представляет собой перспективное направление для улучшения подготовки сертификационной документации, однако полностью реализовать ее потенциал можно только путем постоянного изучения и адаптации к динамичным требованиям нормативной среды.

#### Библиографические ссылки

1. Построение графов знаний нормативной документации на основе семантического моделирования и автоматического извлечения терминов / Д. В. Муромцев, И. А. Шилин, Д. А. Плюхин, И. Р. Баймуратов, Р. Р. Хайдарова, Ю. Ю. Дементьева, Д. А. Ожигин, Т. А. Малышева // Научно-технический вестник информационных технологий, механики и оптики. 2021. Т. 21, № 2. С. 256–266. DOI: 10.17586/2226-1494-2021-21-2-256-266.
2. Moen H., Hakala K., Peltonen L., Matinolli H., Suhonen H., Terho K., Danielsson-Ojala R., Valta M., Ginter F., Salakoski T., Salanterä S. Assisting nurses in care documentation: from automated sentence classification to coherent document structures with subject headings // J Biomed Semantics. 2020. Т. 11, № 1. С. 10. DOI: 10.1186/s13326-020-00229-7.
3. Wu Z., Ma G. NLP-based approach for automated safety requirements information retrieval from project documents // Expert Syst Appl. Pergamon, 2024. Vol. 239. P. 122401. DOI: 10.1016/J.ESWA.2023.122401.



4. Hassan F. ul, Le T. Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing // *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*. 2020. Т. 12, № 2. DOI: 10.1061/(ASCE)LA.1943-4170.0000379.
5. Бобров К. А., Шульман В. Д., Власов К. П. Анализ технологий распознавания текста из изображений // *Международный журнал гуманитарных и естественных наук*. 2022. № 3–2. С. 124–128.
6. Ponnuru M., Ponnmalar S., Likhitha A., SreeT., Chaitanya G. Image-Based Extraction of Prescription Information using OCR-Tesseract // *Procedia Comput Sci*. Elsevier, 2024. Vol. 235. P. 1077–1086. DOI: 10.1016/J.PROCS.2024.04.102.
7. Lee S.-Y., Park J., Yoon J., Lee J. A Validation Study of a Deep Learning-Based Doping Drug Text Recognition System to Ensure Safe Drug Use among Athletes // *Healthcare*. 2023. Vol. 11, № 12. P. 1769. DOI: 10.3390/healthcare11121769.
8. Koponen J., Haataja K., Toivanen P. Novel Deep Learning Application: Recognizing Inconsistent Characters on Pharmaceutical Packaging // *F1000Research* 2024 12:427. F1000 Research Limited, 2024. Vol. 12. P. 427. DOI: 10.3390/healthcare11121769.
9. Arslan H. End to End Invoice Processing Application Based on Key Fields Extraction // *IEEE Access*. 2022. Vol. 10. P. 78398–78413. DOI: 10.1109/ACCESS.2022.3192828.
10. Shi C., Funabiki N., Huo Y., Mentari M., Suga K., Toshida T. A Proposal of Printed Table Digitization Algorithm with Image Processing // *Algorithms*. 2022. Vol. 15, № 12. P. 471. DOI: 10.3390/a15120471.
11. Baek Y., Lee B., Han D., Yun S., Lee H. Character Region Awareness for Text Detection // *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society, 2019. Vol. 2019-June. P. 9357–9366. DOI: 10.1109/CVPR.2019.00959.
12. Shi B., Bai X., Yao C. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition // *IEEE Trans Pattern Anal Mach Intell*. IEEE Computer Society, 2015. Vol. 39, № 11. P. 2298–2304. DOI: 10.1109/TPAMI.2016.2646371.
13. Park Y., Shin Y. Gradual OCR: An Effective OCR Approach Based on Gradual Detection of Texts // *Mathematics*. 2023. Vol. 11, № 22. P. 4585. DOI: 10.3390/math11224585.
14. Biró A., Cuesta-Vargas A., Martín-Martín J., Szilágyi L., Szilágyi S. Synthesized Multilanguage OCR Using CRNN and SVTR Models for Realtime Collaborative Tools // *Applied Sciences*. 2023. Vol. 13, № 7. P. 4419. DOI: 10.3390/app13074419.
15. Chopra S. Leveraging YOLO Algorithm and PaddleOCR in Machine Learning Applications // *Int J Res Appl Sci Eng Technol*. 2024. Vol. 12, № 4. P. 4576–4584. DOI: 10.22214/ijraset.2024.60850.
16. Kleimenkin D. V., Dmitrienko N. A. Using OCR for Russian texts // *Тенденции развития науки и образования*. 2022. Vol. 92, № 10. P. 9–12. DOI: 10.18411/trnio-12-2022-456.
17. Sagade O. D., Sagade D. D. Restaurant Data Scraper: An Automated Tool for Extracting Restaurant Information Using Python Html, CSS and Selenium // *International journal of scientific research in engineering and management*. 2024. Vol. 08, № 11. P. 1–6. DOI: 10.55041/IJSREM39222.
18. Rizquina A. Z., Ratnasari C. I. Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce // *Jurnal Teknologi Dan Sistem Informasi Bisnis*. 2023. Vol. 5, № 4. P. 377–383. DOI: 10.47233/jteksis.v5i4.913.
19. Pichiyan V., Muthulingam S., Sathar G., Nalajala S., Akhil C., Das M. Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis // *Procedia Comput Sci*. 2023. Vol. 230. P. 193–202. DOI: 10.1016/j.procs.2023.12.074.
20. Dikilitaş Y., Dikilitaş Y., ÇakalÇ., Okumuş A., Yalçın H., Yıldırım E., Ulusoy Ö., Macit B., Kırkaya A., Yalçın Ö., Erdoğan E., Sayar A. Performance Analysis for Web Scraping Tools: Case Studies on BeautifulSoup, Scrapy, Htmlunit and Jsoup. 2024. P. 471–480.
21. Dong Z. et al. Transformation from human-readable documents and archives in arc welding domain to machine-interpretable data // *Comput Ind. Elsevier*, 2021. Т. 128. С. 103439. DOI: 10.1016/J.COMPIND.2021.103439.
22. Mani N., Harikumar S. A Knowledge Graph Approach towards Re-structuring of Scientific Articles // *2022 International Conference on Connected Systems & Intelligence (CSI)*. IEEE, 2022. С. 1–8. DOI: 10.1109/CSI54720.2022.9923954
23. Pérez-Suárez A., Martínez-Trinidad J., Carrasco-Ochoa J., Medina-Pagola J. OClustR: A new graph-based algorithm for overlapping clustering // *Neurocomputing*. Elsevier, 2013. Т. 121. С. 234–247. DOI: 10.1016/J.NEUCOM.2013.04.025
24. Ixebayeva Z.S. et al. Conceptual Model for Automatic Proofreading of Technical Documents // *Revue d'Intelligence Artificielle*. 2023. Т. 37, № 1. С. 165–170. DOI: 10.18280/ria.370120.
25. Schäfer P. OnToCode: Template-based code-generation from ontologies // *J Open Source Softw*. 2019. Т. 4, № 40. Т. 1513. DOI: 10.21105/joss.01513.
26. Paetzold G. H., Alva-Manchego F., Specia L. MASSAlign: Alignment and Annotation of Comparable Documents // *AFNLP*. 2017. Т. 1–4. URL: <https://aclanthology.org/I17-3001> (дата обращения 27.01.2025)
27. Павлов Д. А. Автоматическая вёрстка и оформление научной и программной документации // *Компьютерные инструменты в образовании*. 2018. № 6. Т. 39–46. DOI: 10.32603/2071-2340-2018-6-39-46
28. Timakov K. Research and development of methods for automating the design of scientific papers according to specified requirements // *International Conference on Digital Transformation: Informatics, Economics, and Education (DTIEEE2023)* / ed. Gibadullin A.,

Khalmatjanova G. SPIE, 2023. T. 15. DOI: 10.1117/12.2680718

29. Герус М. И., Миронов П. Н. Автоматизированная система генерации программной документации // Nanoindustry Russia. 2020. Т. 96, № 3. С. 106–110. DOI: 10.22184/1993-8578.2020.13.3s.106.110.

30. Zhao T., Yurtsever E., Paulson J., Rizzoni G. Formal Certification Methods for Automated Vehicle Safety Assessment // IEEE Transactions on Intelligent Vehicles. 2023. Т. 8, № 1. Т. 232–249. DOI: 10.1109/TIV.2022.3170517.

### References

1. Mouromtsev D.I., Shilin I.A., Pliukhin D.A., Baimuratov I.R., Khaydarova R.R., Dementyeva Yu.Yu., Ozhigin D.A., Malysheva T.A. [Building knowledge graphs of regulatory documentation based on semantic modeling and automatic term extraction]. *Nauchno-tekhnicheskij vestnik informacionnyh tekhnologij, mekhaniki i optiki*. 2021. Vol. 21, no. 2, pp. 256–266 (in Russ.). Doi: 10.17586/2226-1494-2021-21-2-256-266.

2. Moen H. et al. Assisting nurses in care documentation: from automated sentence classification to coherent document structures with subject headings. *J Biomed Semantics*. 2020. Vol. 11, no. 1. P. 10.

3. Wu Z., Ma G. NLP-based approach for automated safety requirements information retrieval from project documents. *Expert Syst Appl*. Pergamon. 2024. Vol. 239. P. 122401. DOI: 10.1016/J.ESWA.2023.122401.

4. Hassan F. ul, Le T. Automated Requirements Identification from Construction Contract Documents Using Natural Language Processing. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*. 2020 Vol. 12, no. 2.

5. Bobrov K.A., Shulman V.D., Vlasov K.P. [Analysis of technologies for text recognition from images]. *Mezhdunarodnyj zhurnal gumanitarnyh i estestvennyh nauk*. 2022. Iss. 3-2. P. 124–128 (in Russ.).

6. Ponnuru M., Ponnmalar S., Likhitha A., Sree T., Chaitanya G. (2024) Image-Based Extraction of Prescription Information using OCR-Tesseract. *Procedia Comput Sci*. Elsevier. Vol. 235. P. 1077–1086. DOI: 10.1016/J.PROCS.2024.04.102.

7. Lee S.-Y., Park J., Yoon J., Lee J. (2023) A Validation Study of a Deep Learning-Based Doping Drug Text Recognition System to Ensure Safe Drug Use among Athletes. *Healthcare*. Vol. 11, no. 12. P. 1769. DOI: 10.3390/healthcare11121769.

8. Koponen J., Haataja K., Toivanen P. (2024) Novel Deep Learning Application: Recognizing Inconsistent Characters on Pharmaceutical Packaging. *F1000Research* 2024 12:427. F1000 Research Limited. Vol. 12. P. 427. DOI: 10.3390/healthcare11121769.

9. Arslan H. (2022) End to End Invoice Processing Application Based on Key Fields Extraction. *IEEE Access*. Vol. 10. P. 78398–78413. DOI: 10.1109/ACCESS.2022.3192828

10. Shi C., Funabiki N., Huo Y., Mentari M., Suga K., Toshida T. (2022) A Proposal of Printed Table Digitization Algorithm with Image Processing. *Algorithms*. Vol. 15, № 12. P. 471. DOI: 10.3390/a15120471.

11. Baek Y., Lee B., Han D., Yun S., Lee H. (2019) Character Region Awareness for Text Detection. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE Computer Society. Vol. 2019-June. Pp. 9357–9366. DOI: 10.1109/CVPR.2019.00959.

12. Shi B., Bai X., Yao C. An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition. *IEEE Trans Pattern Anal Mach Intell*. IEEE Computer Society. 2015. Vol. 39, no. 11. Pp. 2298–2304. DOI: 10.1109/TPAMI.2016.2646371.

13. Park Y., Shin Y. (2023) Gradual OCR: An Effective OCR Approach Based on Gradual Detection of Texts. *Mathematics*. Vol. 11, no. 22. P. 4585. DOI: 10.3390/math11224585.

14. Biró A., Cuesta-Vargas A., Martín-Martín J., Szilágyi L., Szilágyi S. (2023) Synthesized Multilanguage OCR Using CRNN and SVTR Models for Real-time Collaborative Tools. *Applied Sciences*. Vol. 13, no. 7. P. 4419. DOI: 10.3390/app13074419.

15. Chopra S. (2024) Leveraging YOLO Algorithm and PaddleOCR in Machine Learning Applications. *Int J Res Appl Sci Eng Technol*. Vol. 12, no. 4. Pp. 4576–4584. DOI: 10.22214/ijraset.2024.60850.

16. Kleimenkin D.V., Dmitrienko N.A. Using OCR for Russian texts (2022) Trends in the development of science and education. Vol. 92, no. 10. Pp. 9–12. DOI: 10.18411/trnio-12-2022-456.

17. Sagade O.D., Sagade D.D. (2024) Restaurant Data Scraper: An Automated Tool for Extracting Restaurant Information Using Python Html, CSS and Selenium. *Interantional journal of scientific research in engineering and management*. Vol. 08, no. 11. Pp. 1–6. DOI: 10.55041/IJSREM39222.

18. Rizquina A.Z., Ratnasari C.I. (2023) Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*. Vol. 5, no. 4. Pp. 377–383. DOI: 10.47233/jteksis.v5i4.913.

19. Pichiyan V., Muthulingam S., Sathar G., Nalajala S., Akhil C., Das M. (2023) Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis. *Procedia Comput Sci*. Vol. 230. P. 193–202. DOI: 10.1016/j.procs.2023.12.074.

20. Dikilitaş Y., Dikilitaş Y., ÇakalÇ., Okumuş A., Yalçın H., Yıldırım E., Ulusoy Ö., Macit B., Kırkaya A., Yalçın Ö., Erdoğan E., Sayar A. (2024) Performance Analysis for Web Scraping Tools: Case Studies on BeautifulSoup, Scrapy, Htmlunit and Jsoup. Pp. 471–480.

21. Dong Z. et al. Transformation from human-readable documents and archives in arc welding domain to machine-interpretable data. *Comput Ind. Elsevier*, 2021. Vol. 128. P. 103439. Doi: 10.1016/J.COMPIND.2021.103439.

22. Mani N., Harikumar S. A Knowledge Graph Approach towards Re-structuring of Scientific Articles 2022 International Conference on Connected Systems &

Intelligence (CSI). 2022. IEEE. Pp. 1-8. Doi: 10.1109/CSI54720.2022.9923954.

23. Pérez-Suárez A., Martínez-Trinidad J., Carrasco-Ochoa J., Medina-Pagola J. OClustR: A new graph-based algorithm for overlapping clustering. *Neurocomputing*. Elsevier. 2013. Vol. 121. Pp. 234-247. Doi: 10.1016/J.NEUCOM.2013.04.025.

24. Ixebayeva Z.S. et al. Conceptual Model for Automatic Proofreading of Technical Documents. *Revue d'Intelligence Artificielle*. 2023. Vol. 37, no. 1. Pp. 165-170. Doi: 10.18280/ria.370120.

25. Schäfer P. OnToCode: Template-based code-generation from ontologies. *J Open Source Softw.* 2019. Vol. 4, no. 40. P. 1513. Doi: 10.21105/joss.01513.

26. Paetzold G.H., Alva-Manchego F., Specia L. MASSAlign: Alignment and Annotation of Comparable Documents. *AFNLP*. 2017. Pp. 1-4. URL: <https://aclanthology.org/I17-3001/> (accessed 27.01.2025).

27. Pavlov D.A. [Automatic Layout and Publishing of Scientific and Software Documentation]. *Komp'yuternye instrumenty v obrazovanii*. 2018. No. 6, pp. 39-46 (in Russ.). Doi: 10.32603/2071-2340-2018-6-39-46.

28. Timakov K. Research and development of methods for automating the design of scientific papers according to specified requirements // *International Conference on Digital Transformation: Informatics, Economics, and Education (DTIEEE2023)* / ed. Gibadullin A., Khalmatjanova G. SPIE, 2023. P. 15. doi: 10.1117/12.2680718.

29. Gerus M.I., Mironov P.N. [Automated program documentation generation system]. *Nanoindustry Russia*. 2020. Vol. 96, no. 3. Pp. 106-110. Doi: 10.22184/1993-8578.2020.13.3s.106.110.

30. Zhao T. et al. Formal Certification Methods for Automated Vehicle Safety Assessment. *IEEE Transactions on Intelligent Vehicles*. 2023. Vol. 8, № 1. P. 232-249. doi: 10.1109/TIV.2022.3170517.

\* \* \*

### Modern Approaches To Optimizing Technical Documentation: challenges and solutions for unmanned aircraft systems

E. V. Isaeva, PhD in Linguistics, Associate Professor, Perm State University, Perm, Russia

A. V. Khodasevich, Student, National University of Science and Technology MISIS, Moscow, Russia

Y. M. Isaeva, Student, Perm State University, Perm, Russia

*This article discusses the potential of automation in restructuring the technical documentation required for certification of unmanned aircraft systems (UAS). Certification is a critical regulatory process to ensure compliance with national and international safety standards. Preparing technical documentation for certification is time-consuming, error-prone, and costly process, with manual methods significantly reducing efficiency. These shortcomings pose challenges for mass UAS certification, which is becoming increasingly relevant due to the rapid growth of the industry. This article reviews current technologies that can be used to address these challenges, including machine learning, natural language processing (NLP), knowledge graphs, and semantic analysis. Tools such as EasyOCR, OnToCode, and advanced models such as BERT are reviewed in terms of their suitability for text recognition, data mining, and automatic restructuring of complex technical documents. In addition, the study provides examples of possible practical applications of these methods, focusing on their feasibility to standardise document preparation, reduce human errors and support mass certification of UAS. Challenges such as lack of harmonisation of document formats, integration of legacy systems, and regulatory dynamism are also discussed. The study concludes that despite the emerging opportunities for automation to optimise certification processes, its practical implementation requires further study and collaboration with regulators. The proposed review serves as a conceptual framework that highlights the potential benefits and limitations of integrating automation into the UAS certification workflows and paves the way for future advances in this critical area.*

**Keywords:** unmanned aircraft systems, UAS standardisation, automation, UAS certification, technical documentation, text restructuring, machine learning, knowledge graphs, natural language processing.

Получено: 14.01.25

### Образец цитирования

Исаева Е. В., Ходасевич А. В., Исаева Я. М. Современные подходы к оптимизации технической документации: вызовы и решения для беспилотных авиационных систем // *Интеллектуальные системы в производстве*. 2025. Т. 23, № 2. С. 93–103. DOI: 10.22213/2410-9304-2025-2-93-103.

### For Citation

Isaeva E.V., Khodasevich A.V., Isaeva Y.M. [Modern Approaches To Optimizing Technical Documentation: challenges and solutions for unmanned aircraft systems]. *Intellectual'nye sistemy v proizvodstve*. 2025, vol. 23, no. 2, pp. 93-103. DOI: 10.22213/2410-9304-2025-2-93-103.