

РАЗРАБОТКА И ВНЕДРЕНИЕ МОДУЛЯ УНИВЕРСАЛЬНОГО ПАРСЕРА ОТЧЕТОВ О ПРОВЕДЕННОМ ТО ДЛЯ АВТОМАТИЗАЦИИ УЧЕТА ОТЧЕТНОСТИ В ТЕЛЕКОММУНИКАЦИОННОЙ КОМПАНИИ: АРХИТЕКТУРА И ФУНКЦИОНАЛЬНОСТЬ

Ильин Илья Игоревич

Аспирант, ФГБОУ ВО «Московский государственный технологический университет Станкин»
hitsukey@yandex.ru

INTEGRATION OF EXTERNAL DATA SOURCES INTO THE ACCOUNTING SYSTEM OF A TELECOMMUNICATION COMPANY: ARCHITECTURE AND IMPLEMENTATION

I. Ilyin

Summary. To improve the completeness and integrity of data, telecommunications companies need to solve the problem of integrating various external data sources into internal accounting systems.

The article describes the implemented architecture of the integration module with the accounting system of technical maintenance reports in PJSC VimpelCom.

The analysis demonstrates the need for a variety of interfaces (REST API, mobile application, telegram bot, web interface) for effective work with various categories of users (company employees, employees of various contractors, other systems within the company).

The implementation of the integration module has improved the transparency of the maintenance process, reduced the time for data processing, and increased the reliability of interaction between systems.

Keywords: JSON, Kafka, REST API, Telegram bot, data integration, telecommunications, maintenance, token, reporting.

Аннотация. В данной статье рассматривается разработка и внедрение модуля универсального парсера отчетов о проведенном техническом обслуживании (ТО) базовых станций в телекоммуникационной компании ПАО «Вымпелком».

Основной целью исследования является автоматизация учета отчетности за счет извлечения данных из архивных файлов в формате .zip, .rar, .7z, содержащих отчеты в формате Excel (.xlsx, .xls, .xlsm), а также сопутствующие фотоматериалы.

Для решения поставленных задач разработан парсер на языке программирования PHP, взаимодействующий с базой данных MySQL, обеспечивающий загрузку данных через веб-интерфейс и файловую систему.

В статье подробно описаны архитектура системы, схема базы данных, алгоритм обработки архивов и структуры хранения данных.

Результаты внедрения показали высокую надежность и эффективность работы системы: удалось автоматизировать обработку 95 % поступающих отчетов, снизить количество ошибок ручного ввода с 20 % до 2 % и сократить трудозатраты сотрудников на 70 %. Выявлены преимущества предлагаемого решения и обозначены направления для дальнейшего развития системы, такие как интеграция с OCR-технологиями.

Работа подчеркивает важность автоматизации учета отчетности для повышения операционной эффективности работы телекоммуникационных компаний и снижения вероятности ошибок при ручном вводе данных.

Ключевые слова: автоматизация, базы данных, обработка архивов, обработка данных, оптимизация бизнес-процессов, отчеты, парсер, телекоммуникации, техническое обслуживание.

Введение

Современные телекоммуникационные компании сталкиваются с необходимостью эффективного учета отчетности о проведенном техническом обслуживании (ТО). Традиционные методы обработки отчетов часто требуют значительных временных затрат и могут быть подвержены ошибкам. В связи с этим актуальной задачей является разработка автоматизированных решений для упрощения процесса обработки отчетов. Настоящая работа посвящена созданию модуля

универсального парсера, который позволяет извлекать данные из архивов с отчетами и сохранять их в базе данных.

Особенности задачи:

- Отчеты поставляются в архивах (.zip, .rar, .7z) с вложенными Excel-файлами (.xlsx, .xls, .xlsm) и фотографиями оборудования.
- Ручная обработка занимает до 3 часов на отчет из-за необходимости распаковки, проверки данных и загрузки в БД.

- Требуется интеграция с существующей веб-системой компании, написанной на языке программирования PHP и использующей БД MySQL.

Цель работы — создание модуля, который:

1. Автоматически распаковывает архивы.
2. Извлекает данные из Excel-файлов.
3. Сохраняет структурированную информацию в базе данных MySQL.

Литературный обзор

Развитие технологий автоматизации отчетности является одним из ключевых направлений повышения эффективности бизнес-процессов в современных компаниях, в том числе в телекоммуникационной отрасли.

Особенности различий между автоматизацией и цифровизацией учета рассматриваются в публикации Афанасьевой Е.Ю. [1], где показано, что автоматизация должна сопровождаться стандартизацией процедур для достижения максимальной эффективности.

В работе Жердевой О.В., Корчагиной А.А. и Плетневой Е.Г. [2] рассматриваются методы автоматизации составления отчетности на основе трансляции и трансформации данных, что актуально для телекоммуникационных компаний, имеющих дело с большим объемом отчетной информации.

Кундиус В.А. и Журавлев И.Д. [3] подробно анализируют вопросы моделирования и оптимизации бизнес-процессов в организациях, подчеркивая важность адаптации процессов обработки отчетности к изменяющимся условиям внешней среды.

Мучинский А.В. [4] подчеркивает значение стандартизации хранения учетных данных для обеспечения достоверности и оперативности получения отчетной информации.

Николян Г.Ю. [5] рассматривает применение цифровых технологий в учете и анализе налоговых отчетов, демонстрируя потенциал аналогичных подходов в технической отчетности.

Плис С.А. и Идигова Л.М. [6] делают акцент на эффективной оптимизации бизнес-процессов как ключевом факторе повышения устойчивости компаний в условиях цифровизации.

Рахимов Г.М. [7] исследует проблемы внедрения цифровых технологий в бухгалтерский учет и аудит, что перекликается с задачами цифровизации процессов учета технического обслуживания в телекоммуникационной сфере.

Санарикова Асель и Кожамметова А.К. [8] выделяют перспективные направления использования искусственного интеллекта для оптимизации бизнес-процессов, что может стать следующей ступенью развития автоматизированных систем учета отчетности.

Смоленцев В.М. и Заика И.Т. [9] анализируют построение сети бизнес-процессов в телекоммуникационных компаниях, что особенно важно для правильной интеграции автоматизированных модулей в существующие системы.

Суворова С.Д. и Мозговая А.П. [10] акцентируют внимание на современных тенденциях оптимизации бизнес-процессов, среди которых автоматизация учета занимает важное место как средство повышения производительности и снижения операционных затрат.

Проведенный анализ литературы показывает высокую степень проработки вопросов автоматизации учета и обработки отчетности в научной среде и подтверждает актуальность выбранного направления исследования.

Научная новизна:

- Комплексная обработка архивов и Excel-файлов в едином PHP-модуле.
- Гибкая настройка параметров парсинга отчетов через веб-интерфейс с помощью настроек, хранящихся в таблицах БД.

Материалы и методы

Разработка модуля парсера была осуществлена с использованием языка программирования PHP и системы управления базами данных MySQL.

Для масштабирования работы парсера на новый тип ТО, ответственные лица вносят настройки для парсера, чтобы указать параметры парсинга отчетов. В настройках указывают тип ТО, обязательные листы для парсинга, ячейки из которых нужно получить данные, а также указывают наименование для данных, которые будут распарсены (из конкретного листа и ячейки).

Архивы с отчетами загружаются на сервер в определенную директорию или через веб-интерфейс системы. Каждый новый архив помечается в системе, как ожидающий обработки. После работы парсера, отчет отмечается, как обработанный.

Алгоритм процесса парсинга отчетов представлен на рисунке 1.

Алгоритм работы:

1. Загрузка отчета:
 - Выбор способа загрузки (напрямую в файловую систему или через веб-интерфейс).

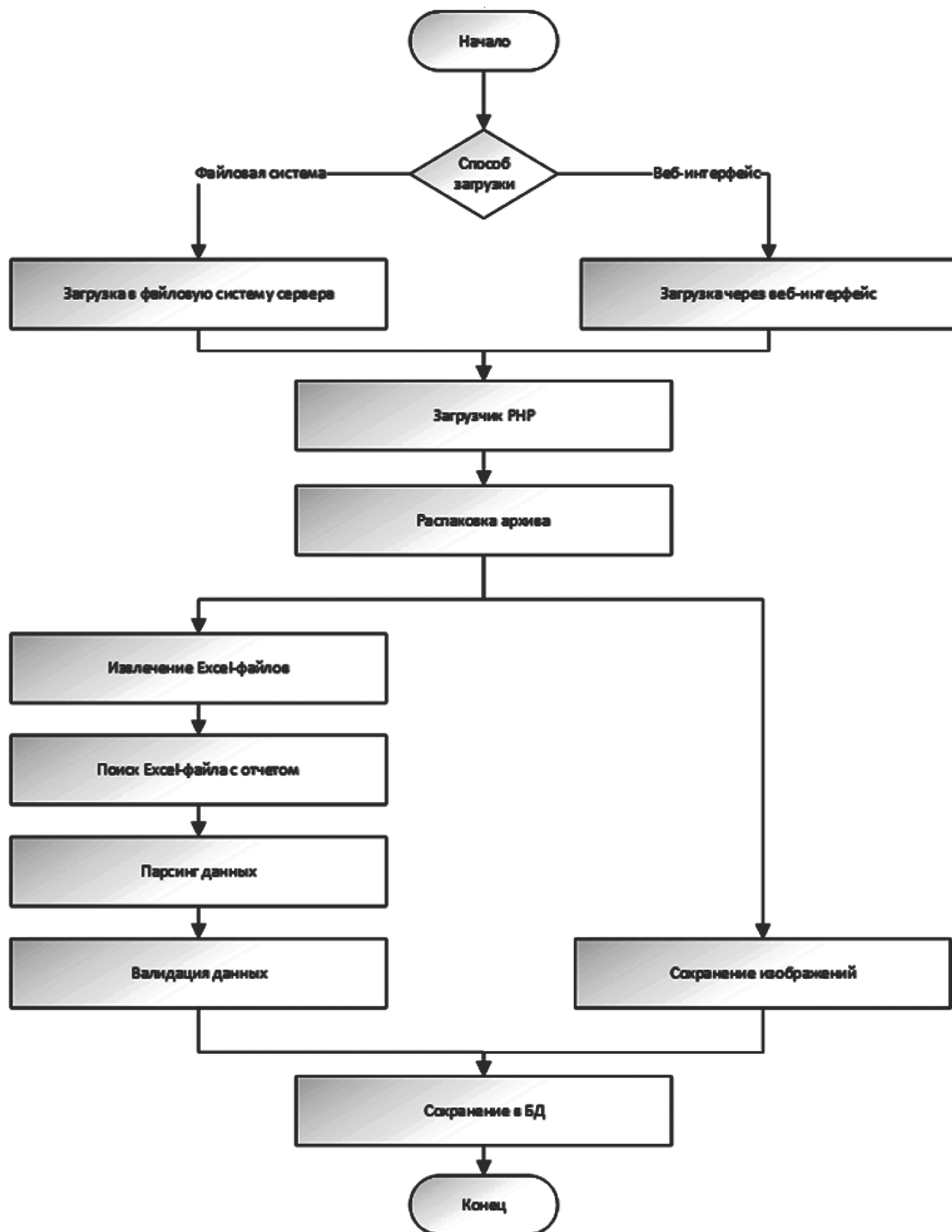


Рис. 1. Алгоритм процесса парсинга

- Загрузка отчета в систему, в зависимости от выбранного способа.
- 2. Выборка отчетов для парсинга:
 - Парсер получает из БД список всех загруженных отчетов, стоящих в очереди на обработку. Далее парсер поочередно обрабатывает каждый отчет.
- 3. Распаковка архива:
 - Проверка формата (.zip/.rar/.7z).
 - Рекурсивный поиск Excel-файлов (.xlsx/.xls/.xlsm) и изображений.
- 4. Обработка Excel:
 - Извлечение из архива всех Excel-файлов во временную директорию.
 - Получение из БД списка обязательных листов в отчете для конкретного типа ТО.
 - Поиск среди извлеченных Excel-файлов файла, в котором будут все обязательные листы. Этот файл будет считаться файлом с отчетом и именно из него будут распарсены данные.
 - Получение из БД списка данных с указанием ячеек, из которых их требуется извлечь.
 - Чтение данных с использованием библиотек.
 - Преобразование данных в ассоциативные массивы.
- 5. Сохранение данных в БД:
 - Транзакционная вставка в MySQL.

Для хранения настроек парсера была разработана схема таблиц БД и связи между ними (см. рис. 2).

Для хранения настроек парсера было решено использовать 2 таблицы БД:

1. UniversalParserSettingsMain — Таблица с настройками для каждого типа ТО, какие листы нужно парсить.
 - Поле ID (первичный ключ) — Уникальный идентификатор записи в таблице. Тип int, не может быть NULL.
 - Поле ToTypeID — ID типа ТО из соответствующей таблицы. Тип int, не может быть NULL.
 - Поле SheetName — Наименование листа из отчета. Тип varchar (31), не может быть NULL. Ограни-

чение в 31 символ обусловлено максимально возможной длиной наименования листа в Excel.

- Поле IsRequired — Флаг, является ли лист обязательным. Тип bit, не может быть NULL.
2. UniversalParserSettingsParams — Таблица с правилами парсинга ячеек.
 - Поле ID (первичный ключ) — Уникальный идентификатор записи в таблице. Тип int, не может быть NULL.
 - Поле UniversalParserSettingsMainID — Уникальный идентификатор (поле ID) связанной записи из таблицы UniversalParserSettingsMain. Тип int, не может быть NULL.
 - Поле Cell — Значение ячейки (например, «D7»), из которой необходимо извлечь данные. Тип varchar (20), не может быть NULL.
 - Поле ParamName — Наименование параметра, который будет извлечен из указанной ячейки (требуется для последующего анализа распарсенных данных). Тип varchar (2000), не может быть NULL.

Для хранения распарсенных данных из отчетов была разработана схема таблиц БД и связи между ними (см. рис. 3).

1. UniversalParserParsedDataMain — Основная таблица с распарсенными файлами отчетов.
 - Поле ID (первичный ключ) — Уникальный идентификатор записи в таблице. Тип int, не может быть NULL.
 - Поле InsertTimestamp — Дата и время добавления записи в таблицу. Тип timestamp, не может быть NULL.
 - Поле Year — Год проведенного ТО, по которому распарсились данные. Тип int, не может быть NULL.
 - Поле ERP — ERP проведенного ТО, по которому распарсились данные. Тип varchar (11), не может быть NULL. Количество символов обусловлено длиной ERP кода в компании.
 - Поле ToTypeID — ID типа ТО из соответствующей таблицы для проведенного ТО, по которому распарсились данные. Тип int, не может быть NULL.

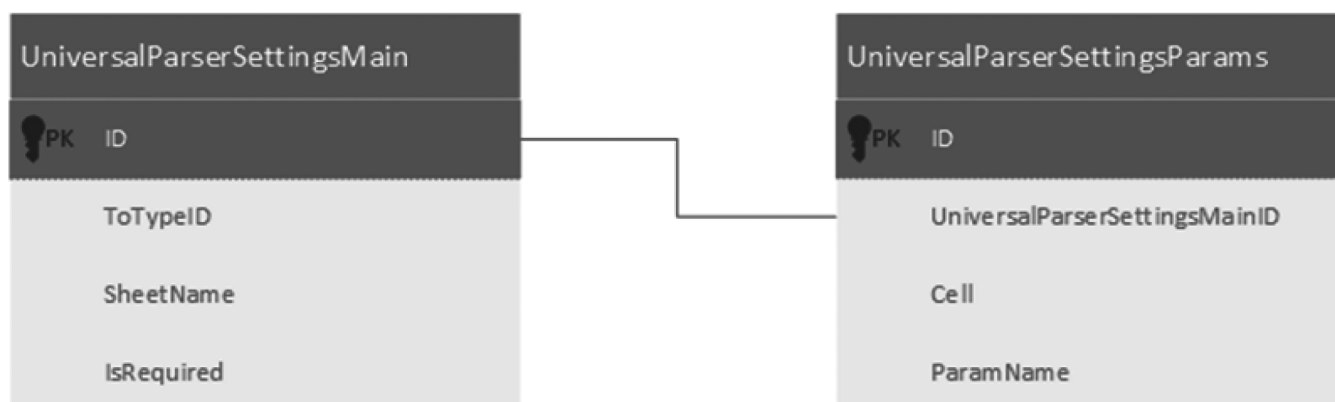


Рис. 2. Схема таблиц БД для настроек парсера

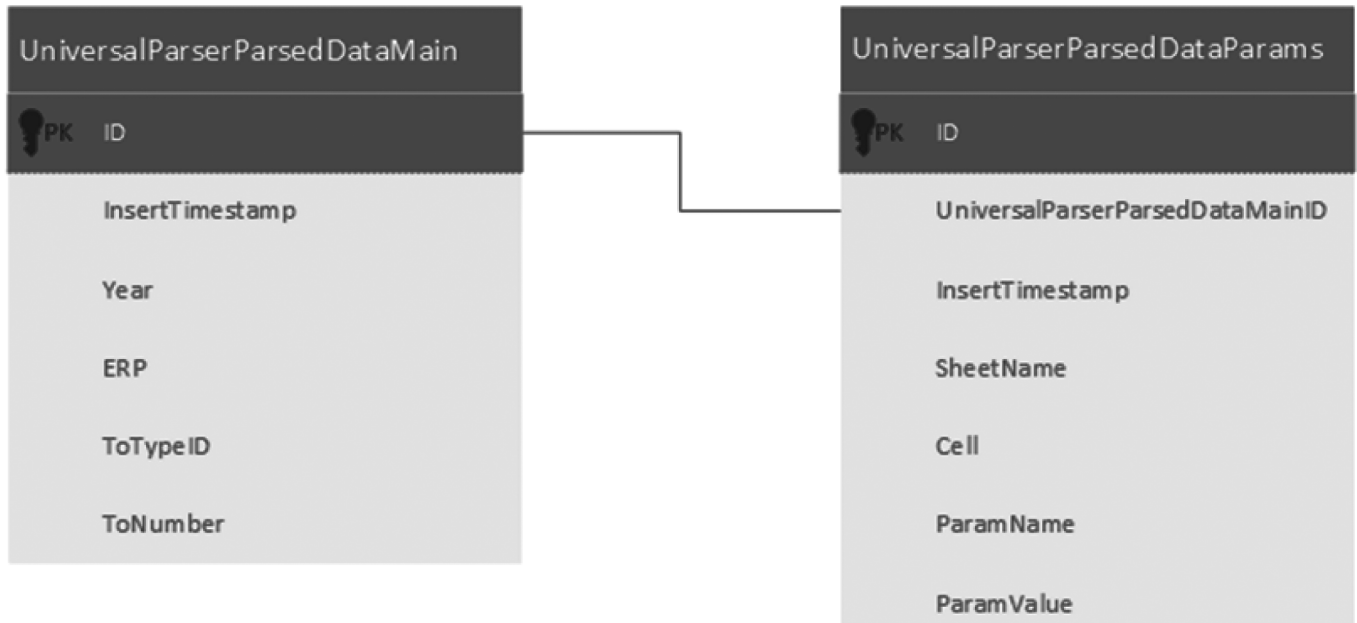


Рис. 3. Схема таблиц БД для распарсенных данных

- Поле ToNumber — Номер проведенного ТО, по которому распарсились данные. Тип int, не может быть NULL.
- 2. UniversalParserParsedDataParams — Таблица с самими распарсенными данными.
 - Поле ID (первичный ключ) — Уникальный идентификатор записи в таблице. Тип int, не может быть NULL.
 - Поле UniversalParserParsedDataMainID — Уникальный идентификатор (поле ID) связанной записи из таблицы UniversalParserParsedDataMain. Тип int, не может быть NULL.
 - Поле InsertTimestamp — Дата и время добавления записи в таблицу. Тип timestamp, не может быть NULL.
 - Поле SheetName — Наименование листа из отчета. Тип varchar (31), не может быть NULL. Ограничение в 31 символ обусловлено максимально возможной длиной наименования листа в Excel.
 - Поле Cell — Значение ячейки (например, «D7»), из которой необходимо извлечь данные. Тип varchar (20), не может быть NULL.
 - Поле ParamName — Наименование параметра, который был извлечен из указанной ячейки (требуется для последующего анализа распарсенных данных). Тип varchar (2000), не может быть NULL.
 - Поле ParamValue — Значение параметра, которое было извлечено из указанной ячейки. Тип varchar (2000), не может быть NULL.

Результаты

В ходе работы был разработан и внедрен модуль универсального парсера отчетов о проведенном техническом обслуживании базовых станций в ПАО «Вымпелком».

Результаты тестирования и внедрения показали:

- Автоматизация обработки 95 % отчетов: система автоматически извлекает данные из архивов различного формата (.zip, .rar, .7z) и загружает их в централизованную базу данных.
- Сокращение времени обработки: время на обработку одного отчета снизилось в среднем с 3 часов до 7–10 минут (включая парсинг и последующий анализ полученных данных ответственным лицом).
- Снижение ошибок ручного ввода: количество ошибок в отчетности, связанных с обработкой отчетов, уменьшилось с 20 % до 2 %.
- Оптимизация работы сотрудников: высвобождено около 70 % рабочего времени специалистов, ранее задействованных в ручной обработке отчетов.
- Экономия ресурсов компании: уменьшение затрат на ручной труд и исправление ошибок в документации.

Обсуждение

Результаты подтверждают, что автоматизация процессов обработки отчетности значительно повышает эффективность операционной деятельности. Разработанный модуль показал преимущество перед альтернативными решениями благодаря:

- Комплексной обработке архивов и вложенных файлов Excel в едином программном контуре.
- Простоте настройки новых шаблонов отчетности через веб-интерфейс без необходимости переписывать код.
- Высокой производительности и отказоустойчивости.

Ограничения текущей версии включают максимальный размер обрабатываемого архива (500 МБ) из-за ограничений PHP и отсутствие автоматического распознавания информации с вложенных изображений.

Проведенные оптимизации, такие как кеширование в Redis и пакетная вставка в БД, позволили ускорить работу модуля при высоких нагрузках.

Дополнительные исследования и разработки направлены на расширение функционала за счет интеграции OCR-технологий.

Заключение

Разработанный модуль универсального парсера отчетов успешно решает задачи автоматизации учета отчетности о проведенном техническом обслуживании базовых станций.

Реализация данного решения позволила добиться существенного повышения эффективности бизнес-процессов компании, снизить вероятность ошибок и значительно сократить трудозатраты сотрудников.

Внедрение модуля в промышленную эксплуатацию в ПАО «Вымпелком» подтвердило его практическую ценность и эффективность.

Перспективы развития проекта включают:

- Интеграцию OCR-модулей для обработки текстовой информации на фотографиях.
- Автоматическое определение структуры новых типов отчетов с применением технологий машинного обучения.

Таким образом, разработанное решение является важным шагом в направлении цифровизации операционных процессов телекоммуникационных компаний.

ЛИТЕРАТУРА

1. Афанасьева Е.Ю. Автоматизация и цифровизация бухгалтерского учета: сходства и различия // Вестник Полоцкого государственного университета. Серия D. Экономические и юридические науки. 2021. №6. С. 9–14. DOI: 10.52928/2070-1632-2021-57-6-9-14
2. Жердева О.В., Корчагина А.А., Плетнева Е.Г. Автоматизация составления отчетности по МСФО на основе методов трансляции и трансформации // Естественно-гуманитарные исследования. 2023. № 1 (45). С. 125–130.
3. Кундиус В.А., Журавлев И.Д. Моделирование и оптимизация бизнес-процессов в управлении организацией // Экономика и бизнес: теория и практика. 2020. №11–2. С. 77–81. DOI: 10.24411/2411-0450-2020-10928.
4. Мучинский А.В. Автоматизация и стандартизация учета и хранения // Форум молодых ученых. 2020. №2 (42). С. 232–237.
5. Николян Г.Ю. Использование цифровых технологий в учете и анализе налоговых отчетов // Вестник науки. 2024. №6 (75). С. 213–221.
6. Плис С.А., Идигова Л.М. Эффективная оптимизация бизнес-процессов — основа выбора стратегии компаний в условиях цифровизации // Международный журнал гуманитарных и естественных наук. 2023. №11–2 (86). С. 137–140. DOI: 10.24412/2500-1000-2023-11-2-137-140.
7. Рахимов Г.М. Проблемы использования цифровых технологий в бухгалтерском учете и аудите // The Scientific Heritage. 2022. №91. С. 35–39. DOI: 10.5281/zenodo.6695727
8. Санарикова Асель, Кожаметова А.К. Оптимизация бизнес-процессов с помощью искусственного интеллекта: перспективы и вызовы // ELS. 2025. №28 февраль ИН. С. 137–142.
9. Смоленцев В.М., Заика И.Т. Моделирование сети бизнес-процессов телекоммуникационной компании // Научный журнал КубГАУ. 2013. №89. URL: <https://cyberleninka.ru/article/n/modelirovanie-seti-biznes-protsessov-telekommunikatsionnoy-kompanii> (дата обращения: 27.04.2025).
10. Суворова С.Д., Мозговая А.П. Оптимизация бизнес-процессов: современное состояние исследований и проблемы практической реализации // Прогрессивная экономика. 2023. №7. С. 64–89. DOI: 10.54861/27131211_2023_7_64.

© Ильин Илья Игоревич (hitsukey@yandex.ru)

Журнал «Современная наука: актуальные проблемы теории и практики»