

АВТОМАТИЗАЦИЯ ПРОЦЕССА ОБРАБОТКИ ДОКУМЕНТОВ С ПРИМЕНЕНИЕМ ТЕХНОЛОГИИ RPA

Пушкарева М.А.¹, Цапко И.В.²

*¹ Томский политехнический университет, ИШИТР, 8К03,
e-mail: map33@tpu.ru*

*² Томский политехнический университет, ИШИТР, к.т.н., доцент ОИТ,
e-mail: tsiv@tpu.ru*

Аннотация

Разрабатываемый программный робот предназначен для увеличения скорости обработки документов руководителем образовательной программы ТПУ. Предполагается частичная автоматизация процесса компоновки учебно-методических материалов по дисциплинам учебного плана направления подготовки.

Ключевые слова: RPA, программный робот, автоматизация процесса.

Введение

Программные роботы в наше время являются ключевыми помощниками больших корпораций, позволяющие усовершенствовать работу и сделать ее более быстрой и точной.

RPA (Robotic Process Automation) [1] – технология создания программных роботов, которые освобождают сотрудников от рутины, роботизируя бизнес-процессы компании. Роботы не ошибаются, легко масштабируются и работают круглосуточно, что повышает эффективность бизнеса. Компании, применяющие программных роботов, могут снизить затраты на заработную плату, а следовательно, повысить свою конкурентоспособность на рынке.

Целью данной работы является автоматизация процесса обработки документов руководителем основной образовательной программы (ООП) ТПУ при помощи программного робота. В работе будет отражена частичная автоматизация процесса компоновки учебно-методических материалов (УММ) по дисциплинам учебного плана направления подготовки, что позволит увеличить скорость и повысить качество их обработки.

Анализ автоматизируемого процесса

Для автоматизации процесса обработки документов выбрана среда разработки PIX Studio [2].

PIX Robotics – российский разработчик технологичных и инновационных решений, которые включают применение программных роботов для среднего и крупного бизнеса.

Платформа PIX Studio – это среда разработки и исполнения программных роботов, которая характеризуется следующими свойствами:

1. Создание логики процесса для робота, обработка возможных отклонений и исключений.
2. Наличие двух вариантов отображения алгоритма робота: в виде блок-схемы и в виде списка команд.
3. Доступность интерактивной среды обучения моделей искусственного интеллекта – AutoML Smart Activities.

Классические этапы автоматизации любого бизнес-процесса выглядят следующим образом:

1. Подробное изучение и анализ процесса, составление схем бизнес-процесса [3].
2. Разработка архитектуры и основных сценариев робота.
3. Проведение тестирования в тестовой среде, устранение возникших ошибок.

Руководитель ООП ТПУ – это ответственное лицо, в обязанности которого входит деятельность по проектированию, разработке, реализации и совершенствованию образовательной программы. Для достижения указанных целей руководитель ООП решает задачи по подготовке УМК (учебно-методических комплектов) ООП и последующего их внесения в фонд. Ежегодно для каждой специализации образовательной программы необходимо обработать и выложить в фонд порядка двухсот документов, в том числе и учебно-методических материалов (УММ) по дисциплинам учебного плана (рабочая программа дисциплины, аннотация рабочей программы, фонд оценочных средств). В общем виде, обработка УММ состоит из следующей последовательности действий: подготовка

документа, печать документа, подписание документа ответственными лицами, сканирование документа, занесение скана документа в фонд. Однако, для экономии средств, производится печать не всего документа, а только страниц с подписями ответственных лиц: руководителя образовательной программы, преподавателя дисциплины, руководителя подразделения, обеспечивающего преподавание дисциплины, директора инженерной школы.

В общем виде диаграмма процесса обработки документов в нотации BPMN выглядит следующим образом (рис. 1):

1. Печать страниц, на которых необходимо проставить подписи.
2. Подписание распечатанных страниц руководителем ООП.
3. Подписание распечатанных страниц в обеспечивающем подразделении.
4. Сканирование подписанных страниц.
5. Перевод УММ по дисциплинам в формат *.pdf.
6. Компоновка УММ (замена страниц в исходных документах на соответствующие подписанные страницы).

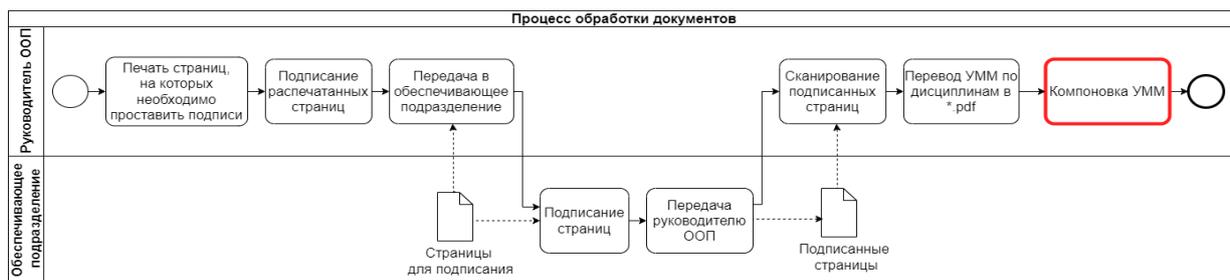


Рис. 1. Схема процесса обработки документов

Именно этап компоновки УММ является претендентом на автоматизацию, так как обладает следующими признаками: основан на повторяющейся последовательности действий для каждого документа, является рутинным, отнимает много времени при этом не требует специальной подготовки.

Автоматизация процесса

Рассмотрим этап компоновки УММ до применения автоматизации. Последовательность его шагов представлена на рисунке 2:

1. Открытие исходного файла документа.
2. Удаление в документах страниц, которые необходимо заменить (в рабочей программе дисциплины – это первая и последняя страницы, в фонде оценочных средств – первая).
3. Подстановка отсканированных изображений в исходный документ.

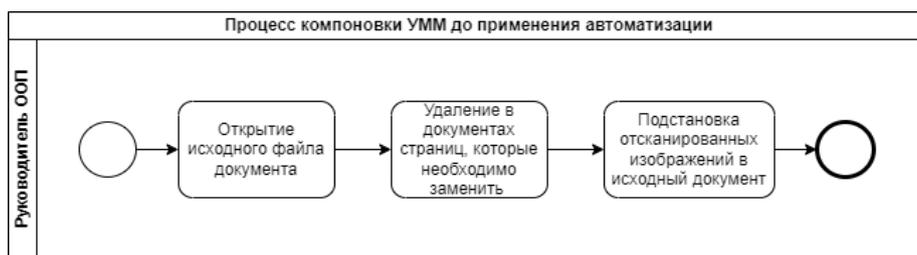


Рис. 2. Схема процесса компоновки УММ до применения автоматизации

Процесс подстановки подписанных страниц после автоматизации с использованием программного робота будет выглядеть следующим образом (рис. 3):

1. Предварительная подготовка данных (инициализация путей к папкам с исходными данными).
2. Считывание роботом очередного изображения подписанной страницы.
3. Распознавание текста на изображении при помощи Yandex Vision [4].
4. Проверка распознанного текста на содержание вида документа (рабочая программа, фонд оценочных средств) и названия дисциплины.
5. Замена страниц в исходных документах на соответствующие подписанные страницы.
6. Сохранение готовых документов в папке.



Рис. 3. Схема процесса компоновки УММ после применения автоматизации

Предварительная подготовка

При реализации данной задачи было необходимо ввести следующие ограничения:

1. Все документы, которые участвуют в обработке, должны иметь формат *.pdf.
2. Файлы в папках должны быть отсортированы по наименованию файла (по возрастанию).
3. Должна присутствовать таблица в формате *.xlsx, в которую необходимо внести названия дисциплин обрабатываемых документов.

Перед запуском программного робота нужно загрузить адрес местонахождения следующих элементов:

- папок с материалами по обрабатываемым дисциплинам;
- файла с отсканированными изображениями;
- таблицы в формате *.xlsx с перечислением наименований обрабатываемых дисциплин.

Процесс обработки после выполнения предварительных действий руководителем ООП по загрузке файлов будет следующий (рис. 4):

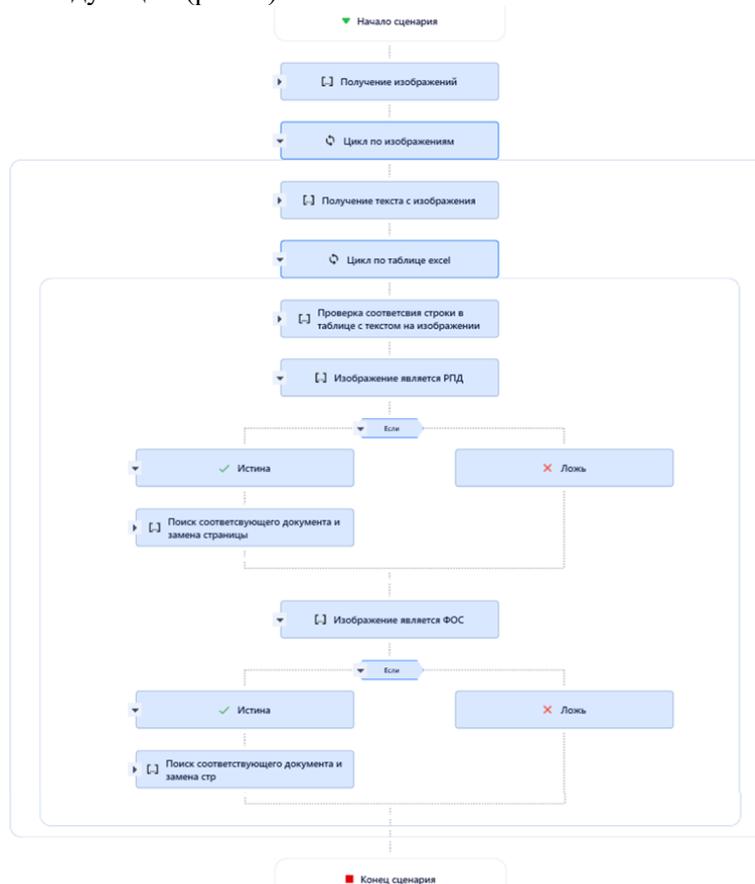
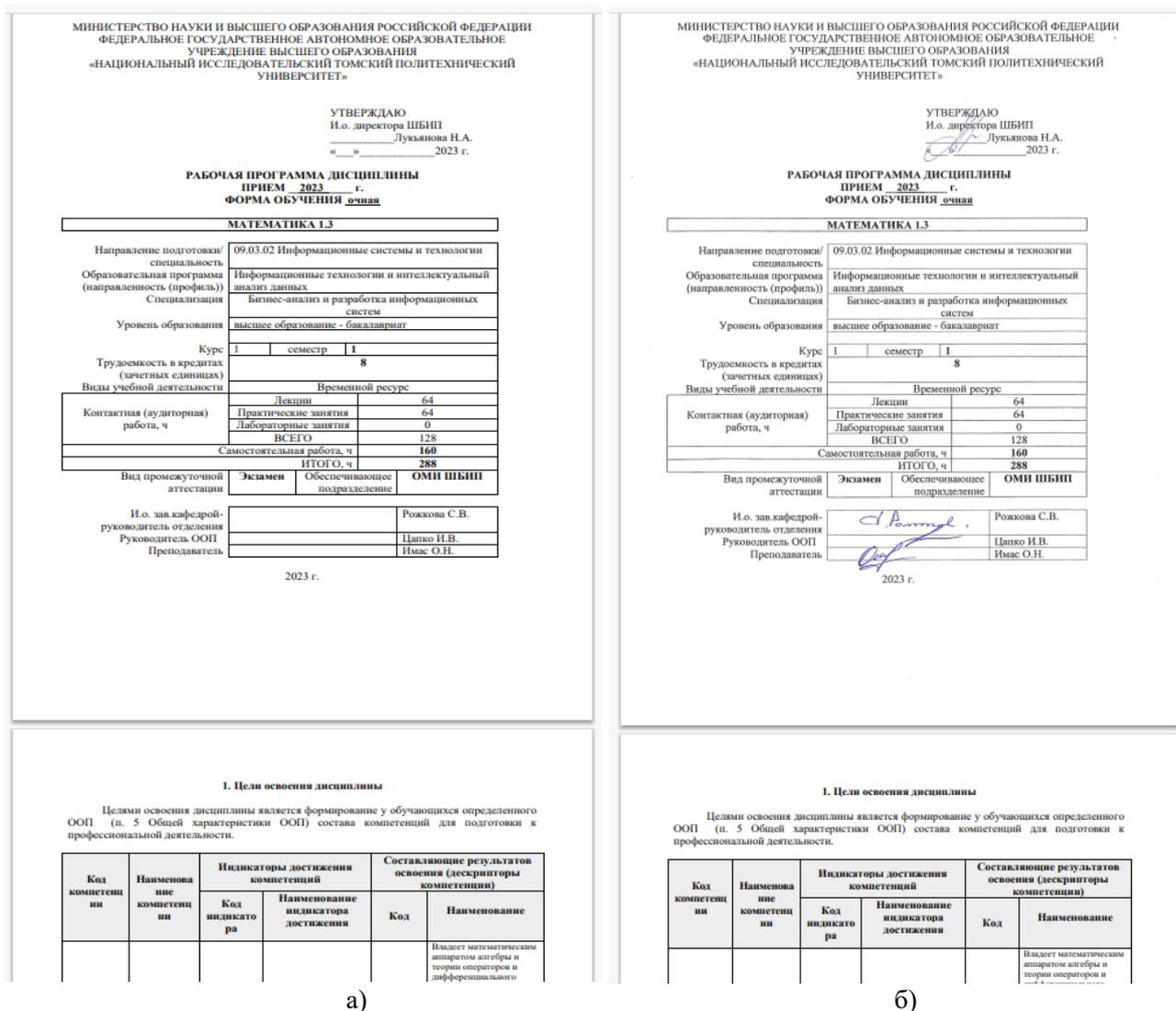


Рис. 4. Диаграмма процесса обработки

Таким образом, в исходных документах заменяются страницы, на которых было необходимо проставить подписи, на подписанные отсканированные страницы. Наглядно это можно увидеть на рисунке 5, где изображен исходный документ без подписей (а), а тот же документ, но уже с замененной отсканированной страницей, на которой были проставлены подписи (б).



а) исходный документ без подписей;
 б) документ с подписью

Заключение

Был написан скрипт, благодаря которому была реализована поставленная задача – автоматическая вставка отсканированных страниц с подписями в исходные соответствующие документы.

Написанный скрипт позволяет упростить процесс обработки документов и уменьшить затраченное на это время. При использовании скрипта время обработки одного документа занимает не более 10 секунд, таким образом, на обработку всех документов для одной специализации (рабочие программы и фонды оценочных средств в количестве 112 штук) – потратится 18 минут. В то время как обработка без использования скрипта для одного документа может занять от 3 до 5 минут.

Список использованных источников

1. PIX Studio – Студия разработки программных роботов // PIX Robotics: сайт. – 2024. – URL: <https://pix.ru/products/pix-rpa/pix-studio/>.
2. Вводный курс по RPA // Академия PIX: сайт. – 2024. – URL: <https://academy.pix.ru/course/view.php?id=53>.

3. Как начать работать с Vision // Yandex Vision OCR: сайт. – 2024. – URL:
<https://cloud.yandex.ru/docs/vision/quickstart>.

4. Основы анализа процесса в RPA // Академия PIX: сайт – 2024. – URL:
<https://academy.pix.ru/course/view.php?id=17>.